



# A tutorial guide to geostatistics: Computing and modelling variograms and kriging



M.A. Oliver <sup>a</sup>, R. Webster <sup>b,\*</sup>

<sup>a</sup> Soil Research Centre, Department of Geography and Environmental Science, University of Reading, Reading RG6 6DW, Great Britain, United Kingdom

<sup>b</sup> Rothamsted Research, Harpenden AL5 2JQ, Great Britain, United Kingdom

## ARTICLE INFO

### Article history:

Received 4 June 2013

Received in revised form 18 September 2013

Accepted 22 September 2013

### Keywords:

Geostatistics

Sampling

Variogram

Model fitting

Trend

Kriging

## ABSTRACT

Many environmental scientists are analysing spatial data by geostatistical methods and interpolating from sparse sample data by kriging to make maps. They recognize its merits in providing unbiased estimates with minimum variance. Several statistical packages now have the facilities they require, as do some geographic information systems. In the latter kriging is an option for interpolation that can be done at the press of a few buttons. Unfortunately, the ease conferred by this allows one to krige without understanding and to produce unreliable and even misleading results. Crucial for sound kriging is a plausible function for the spatial covariances or, more widely, of the variogram. The variogram must be estimated reliably and then modelled with valid mathematical functions. This requires an understanding of the assumptions in the underlying theory of random processes on which geostatistics is based. Here we guide readers through computing the sample variogram and modelling it by weighted least-squares fitting. We explain how to choose the most suitable functions by a combination of graphics and statistical diagnostics. Ordinary kriging follows straightforwardly from the model, but small changes in the model function and its parameters can affect the kriging error variances. When kriging is automated these effects remain unknown. We explain the choices to be made when kriging, i.e. whether the support is at points or over blocks, and whether the predictions are global or within moving windows.

© 2013 Elsevier B.V. All rights reserved.

## Contents

1.	Introduction	57
2.	Random processes	57
2.1.	Stationarity	58
3.	The variogram	58
3.1.	Computing and modelling variograms	58
3.2.	Reliability of the empirical variogram	60
3.2.1.	Sample size	60
3.2.2.	Lag interval and bin width	61
3.2.3.	Marginal distribution	61
3.2.4.	Anisotropy	63
3.2.5.	Trend	63
3.3.	Inference	64
4.	Kriging	65
4.1.	Punctual or block kriging	65
4.2.	Kriging neighbourhood	65
4.3.	Cross-validation	66
4.4.	Smoothing	66
4.5.	Kriging in the presence of trend	66
5.	General conclusion	67
	Acknowledgements	67

\* Corresponding author at: Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, Great Britain, United Kingdom. Tel.: +44 1582 763133.  
E-mail address: [richard.webster@rothamsted.ac.uk](mailto:richard.webster@rothamsted.ac.uk) (R. Webster).

Appendix . . . . .	67
Models for variograms . . . . .	67
Robust variogram estimators . . . . .	67
The ordinary kriging system . . . . .	67
Cross-validation statistics . . . . .	68
References . . . . .	68

## 1. Introduction

Daniel Krige, the doyen of geostatistics, died earlier this year at the grand age of 93. Early in his career he developed empirically statistical methods to predict ore grades from spatially correlated sample data in the gold mines of South Africa (Krige, 1951, 1966). In the 1960s his approach was formalized by Matheron (1963, 1965), and the term 'kriging' was coined in his honour. In the two decades that followed environmental scientists – pedologists, hydrologists, geologists, and atmospheric scientists, to name a few – saw the merit of this technology in their own fields (e.g. Burgess and Webster, 1980; de Marsily and Ahmed, 1987; Gajem et al., 1981; McBratney et al., 1982; Vauclin et al., 1983; Russo, 1984; Oliver and Webster, 1987). Now kriging is applied widely and with increasing sophistication in petroleum engineering, mining and geology, meteorology, hydrology, soil science, precision agriculture, pollution control, public health, fishery, plant and animal ecology, and remote sensing. Kriging has become a generic term for several closely related least-squares methods that provide best linear unbiased predictions (BLUP) and also some non-linear types of prediction. It is a major advance over the mathematical methods of interpolation common in the first half of the 20th century.

Environmental surveys are almost always based on samples, but in general the measurements represent a continuum in space from which the sample has been drawn. Most analysts and their clients want to know what values are likely at intervening places. Kriging enables them to predict those values optimally, i.e. without bias and with minimum variance; hence its popularity.

Initially practitioners had to write their own code for geostatistical analysis; they had to have understanding of numerical analysis to program the methods. In the last 20 years the situation has changed dramatically with powerful software that has become widely and cheaply available in the public domain, such as GSLIB (Deutsch and Journel, 1998), gstat (Pebesma, 2004; Pebesma and Wesseling, 1998) and GenStat (Payne, 2013). Gstat in particular is now accessible through R free of charge (see <http://cran.r-project.org/web/packages/gstat/index.html>). Several geographic information system (GIS) packages also have facilities for geostatistical analysis, and kriging has become one of the favoured interpolation routines, if not *the* favoured one. The 'Spatial Analyst' component of ArcGIS (3D-Analyst and Geostatistical Analyst Tool, ArcGIS version 9.2) is especially congenial with attractive graphics. It has encouraged many environmental scientists to use geostatistics, and specifically ordinary kriging (see Section 4), for interpolation and mapping. With kriging in its various forms, environmental scientists can make spatial predictions at any location between their observation points without bias and take proper account of the errors, which are minimized and also estimated together with the predicted values. Unfortunately, the ease with which modern software can be used means that anyone can produce maps by kriging without understanding what happens between the data and the resulting maps. At the press of a few buttons on a computer one can interpolate from scattered data and display the result as a map. The software becomes a 'black box' in which, somehow, a variogram is computed and values from it are inserted into kriging equations without any intervention or assessment by the user.

There are several textbooks on geostatistics (e.g. Chilès and Delfiner, 2012; Goovaerts, 1997; Olea, 1999), including our own (Webster and

Oliver, 2007). Judging from the numerous scripts we are asked to read for this journal and others, however, we have the strong impression that these books do not provide the succinct guidance that authors seek to practice geostatistics wisely. Most authors seem to cull their knowledge from journal articles, many of which are sketchy or misleading and some that are actually wrong.

Our purpose here is deliberately educational; it is to guide investigators, in particular those intent on publishing records of their research in *Catena*, to use the basic geostatistical tools correctly and with understanding, and to avoid the pitfalls that lead to worthless results and misleading claims and to scientific papers that require major revision based on fresh analysis and often more data.

Many environmental scientists who use geostatistical packages have maps as their ultimate goals. But kriging for interpolation is only the penultimate step in a chain that begins with sampling and proceeds through the exploration and screening of data, perhaps transformation, crucially the estimation and modelling of one or more variograms, and ends with graphic display. Here we look at each of these steps and the assumptions required to implement them. We also tell intending authors what they should report so that readers know and could repeat what they have done. We introduce some algebraic notation for brevity, but we have placed most of the essential equations in Appendix A so as not to break the flow of the narrative. You can find them all with explanations in the textbooks cited above.

We are soil scientists, and we set the scene and illustrate the procedures with examples in soil survey. There are close analogies in other branches of land research, and scientists in those fields should find our guide equally apt.

## 2. Random processes

Features of the environment, such as soil, are the product of many interacting physical, chemical and biological processes. These processes are physically determined, but their interactions are so complex that the variation appears to be random. This complexity and incomplete understanding of the processes means that a deterministic or mathematical solution to quantify the variation is out of reach at present. The logical solution required a leap of imagination by Matheron (1965) in his seminal thesis to treat the variation as though it were random. Let us first translate this idea of a random property into a mathematical one, which we call a random process. We can formalize it in the notation that has become conventional as follows.

1. The value of a property, say  $z$ , at any place  $\mathbf{x}$ , equivalent to  $x_1, x_2$  in two dimensions, and denoted by  $z(\mathbf{x})$  is one of an infinity of values of a random variable  $Z(\mathbf{x})$  at that place. We call it a 'realization' of the process.
2. The set of random values at all such places, again infinite in number, in a region is a random process, and also denoted  $Z(\mathbf{x})$ .
3. The random variable is spatially correlated at some scale.

Variables, such as the heights of water tables, the concentrations of elements in soil, air temperatures and rainfall, for example, are regarded as spatial random variables. For each, however, we have only a single realization. Consequently, we cannot compute statistics for the realization or draw inferences from it. Inference requires many realizations, and so

to overcome this impasse we must make a further assumption, namely that the process is stationary.

### 2.1. Stationarity

The notion of stationarity underpins geostatistics and allows us to assume that there is the same degree of variation from place to place. We can represent the random process by the model

$$Z(\mathbf{x}) = \mu + \varepsilon(\mathbf{x}), \quad (1)$$

where  $\mu$  is the mean of the process and  $\varepsilon(\mathbf{x})$  is a random quantity with a mean of zero and a covariance,  $C(\mathbf{h})$ , given by

$$C(\mathbf{h}) = E[\varepsilon(\mathbf{x})\varepsilon(\mathbf{x} + \mathbf{h})], \quad (2)$$

which is equivalent to

$$C(\mathbf{h}) = E[\{Z(\mathbf{x}) - \mu\}\{Z(\mathbf{x} + \mathbf{h}) - \mu\}] = E[Z(\mathbf{x})Z(\mathbf{x} + \mathbf{h}) - \mu^2]. \quad (3)$$

In these equations  $\mathbf{h}$  is the separation between samples in both distance and direction;  $Z(\mathbf{x})$  and  $Z(\mathbf{x} + \mathbf{h})$  are the values of  $Z$  at places  $\mathbf{x}$  and  $\mathbf{x} + \mathbf{h}$ , and  $E$  denotes the expectation. If the mean is not constant then the covariance cannot exist, and we invoke Matheron's (1965) somewhat weaker assumption of intrinsic stationarity in which the expected differences are zero, i.e.  $E[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = 0$ , and the covariance is replaced by half the variance of the differences, the semivariance:

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = \frac{1}{2} E[\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2]. \quad (4)$$

Like the covariance, the semivariance depends on  $\mathbf{h}$  and only on  $\mathbf{h}$ , and as a function of  $\mathbf{h}$  it is the variogram,  $\gamma(\mathbf{h})$ . The variogram is more generally useful than the covariance function because of these weaker assumptions, and so it has become the central tool of geostatistics. For second-order stationary processes the covariance function and variogram are equivalent:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}), \quad (5)$$

where  $C(\mathbf{0})\sigma^2$  is the variance of the random process. We mention one more function of  $\mathbf{h}$ , namely the correlogram:  $\rho(\mathbf{h}) = C(\mathbf{h})/\sigma^2$ .

We close this section by emphasizing that randomness and stationarity are assumed attributes of our models of variation; they are not properties of either the real world or of data, and there is no formal test for them. Rather, they are useful in that they help us to understand the complexity of the real world and to predict its conditions at unvisited places.

### 3. The variogram

The variogram as defined above is that of the random process  $Z(\mathbf{x})$  which we assume to have given rise to the actual realization on the ground; it is a theoretical function.

There are two other variograms that must be recognized.

- The regional variogram is that of a particular realization of the random process in a finite region. You might compute it if you had complete information of the region and a computer with infinite capacity. It can differ from the theoretical variogram in that a region does not necessarily encompass all the variation in the assumed theoretical process. It is sometimes called the 'non-ergodic variogram' for this reason (see for example Brus and de Gruijter, 1994). We can get close to the regional variogram with dense data from satellite and proximal sensors when we compute their experimental variograms.
- The experimental variogram is one that we estimate from data,  $z(\mathbf{x}_i)$ ,

$i = 1, 2, \dots$ . It is usually computed by the method of moments and attributed to Matheron (1965):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{j=1}^{m(\mathbf{h})} \{z(\mathbf{x}_j) - z(\mathbf{x}_j + \mathbf{h})\}^2, \quad (6)$$

where  $m(\mathbf{h})$  is the number of paired comparisons at lag  $\mathbf{h}$ . By incrementing  $\mathbf{h}$  in steps we obtain an ordered set of values, as shown by the points plotted in each of the graphs in Fig. 1. This is the experimental variogram, also known as the sample variogram because it is based on a sample. It estimates points on the regional variogram.

#### 3.1. Computing and modelling variograms

By programming Equation (6) we can compute the experimental variogram from data. The result depends on the precise way we apply the program and the decisions we make, which we discuss after modelling. These choices affect the outcome and should be stated clearly in any paper for the Journal.

The experimental variogram consists of semivariances at a finite set of discrete lags, whereas the underlying function is continuous for all  $\mathbf{h}$ , Equation (4). Therefore, the next step is to fit a smooth curve or surface to the experimental values, one that describes the principal features of the sequence while ignoring the point-to-point erratic fluctuation. The curve must have a mathematical expression that can describe the variances of random processes with changing lag and guarantee non-negative variances in your predictions. The choice is limited to a few

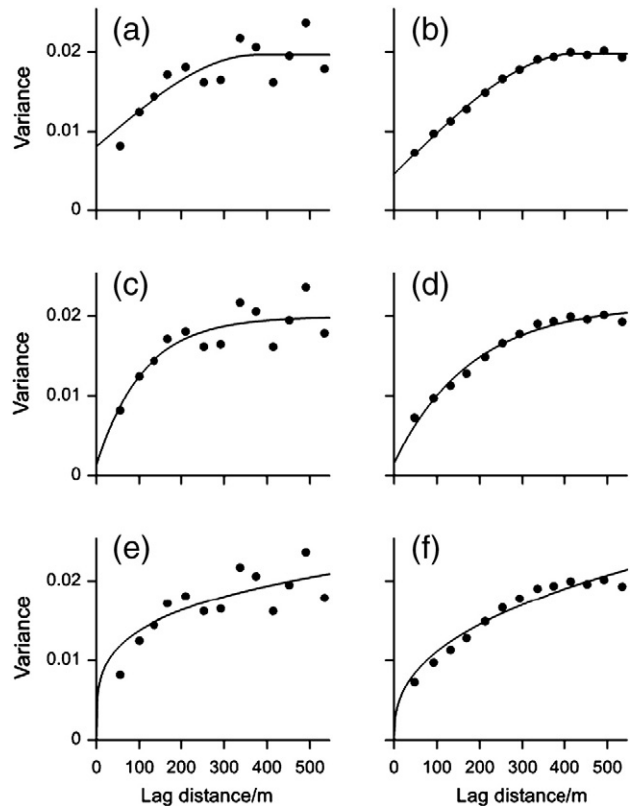


Fig. 1. Experimental variograms of  $\log_{10}K^+$  at Broom's Barn Farm. In the left-hand column are the ones computed from 87 data, and on the right are the ones computed from all 434 data. The lags have been binned over all directions and incremented in steps of 40 m, the sampling interval on the grid. The solid lines show the models fitted to them with (a) spherical model, (c) exponential model, (e) power function, and computed from 434 data and fitted with (b) spherical model, (d) exponential model and (f) power function.

simple functions that satisfy these. One of the most popular functions is the isotropic spherical-plus-nugget model. We give its equation here to introduce parameters mentioned later in the text:

$$\begin{aligned} \gamma(h) &= c_0 + c \left\{ \frac{3h}{2r} - \frac{1}{2} \left( \frac{h}{r} \right)^3 \right\} \text{ for } 0 < h \leq r \\ &= c_0 + c \text{ for } h > r \\ &= 0 \text{ for } h = 0, \end{aligned} \tag{7}$$

in which  $h = |\mathbf{h}|$  is the lag distance, and the parameters are  $c_0$ , the nugget variance,  $c$  the spatially correlated variance, and  $r$  the range, which is the limit of spatial correlation. The quantity  $c_0 + c$  estimates the variance of the random process and is known as the ‘sill’. The nugget variance,  $c_0$ , represents the uncorrelated variation at the scale of sampling; it is the variation that remains unresolved including any measurement error. The quantity  $c$  is the correlated component of the variation that represents continuity. Three other popular functions are the exponential, the Gaussian and the power models; their equations are given in [Appendix A](#). Their definitions are in all the standard textbooks, and they are available in most geostatistical software. We caution against the ill-considered use of the Gaussian model, which is at the limit of acceptability for a random process and can lead to bizarre predictions ([Chilès and Delfiner, 2012](#)). [Wackernagel \(2003\)](#) also warns against it and shows implausible results from its use.

Fitting models is perhaps the most controversial aspect of geostatistics. Some practitioners still fit models by eye; the procedure is unreliable for several reasons including fluctuations in the experimental variogram, the varying accuracy of the computed semivariances

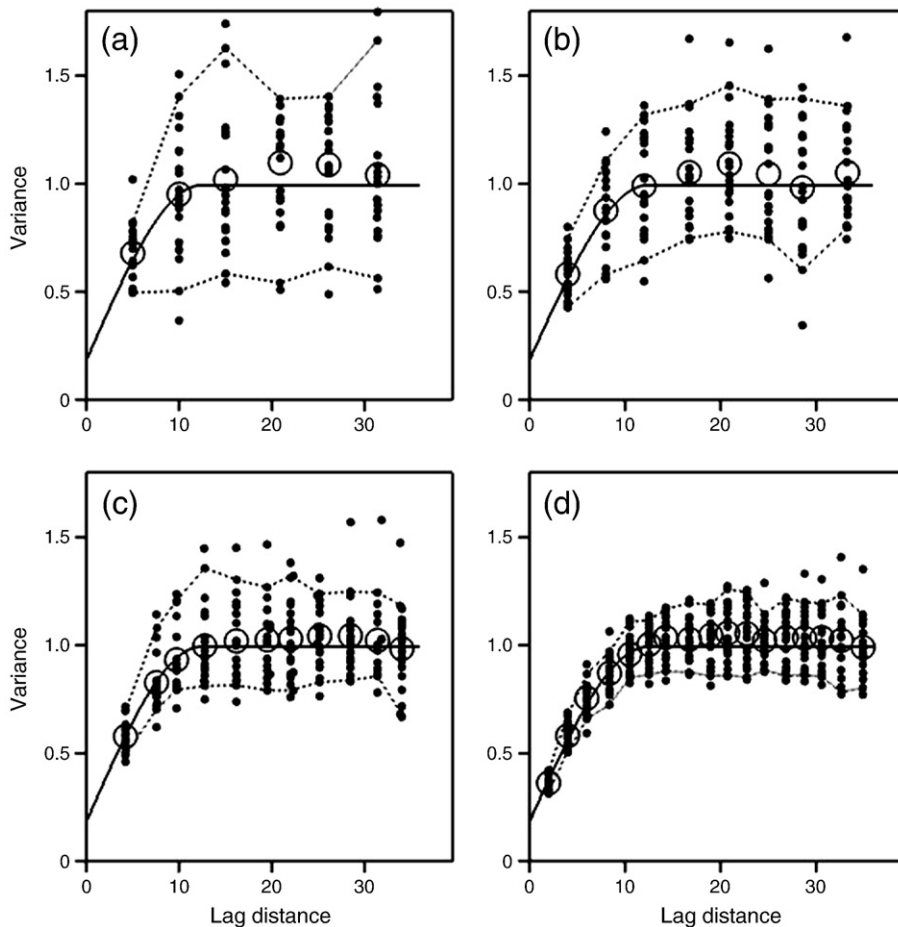
and anisotropy, i.e. variation in the underlying function with change in direction. We see no need for this approach now that we have excellent statistical software. [Pannatier \(1995\)](#), however, combined fitting by eye with statistical evaluation in his program *Variowin*. It seems to work well, but in our experience not quite as well as the next approach ([Webster and Oliver, 1997](#)).

The approach we take is the one taken in most modern software (though not in GSLIB) and is the one we strongly recommend.

1. Plot the experimental variogram.
2. Choose several models that appear to have the right shape and fit each in turn by weighted least squares ([Cressie, 1985](#); [McBratney and Webster, 1986](#)) in an accredited program.
3. Plot the fitted models on the graph of the experimental variogram and assess whether the fit looks reasonable.
4. If all plausible models seem to fit well, choose the one with the smallest residual sum of squares (RSS) or smallest mean square.
5. One might be able to improve a fit in the above sense by elaborating the model. Any combination of the simple valid models is itself valid. But is the added complexity worth it? The Akaike information criterion, the AIC ([Akaike, 1973](#)), may help to answer. It is defined as

$$AIC = -2 \times \ln L + 2 \times p, \tag{8}$$

in which  $L$  is the maximized likelihood and  $p$  is the number of parameters in the model; the aim is to minimize it ([Webster and McBratney, 1989](#); [Webster and Oliver, 2007](#)). For any one experimental variogram



**Fig. 2.** Experimental variograms obtained from a large simulated field by repeated sampling with four sizes of sample, namely 49, 81, 144 and 324 points. The black discs are the experimental values, the dashed lines join the 5% and 95% quantiles. The circles are the mean values, and the solid lines are the spherical model fitted to the exhaustive variogram with parameters  $c_0 = 0.182$ ,  $c = 0.811$  and  $r = 12.4$ .

it has a variable part which we can calculate as

$$A = n \ln R + 2p, \quad (9)$$

where  $n$  is the number of estimated semivariances and  $R$  is the mean of the squared residuals. We can usually increase  $p$  to improve the fit, i.e. to diminish  $R$ , but if in doing so we do not diminish  $A$  then the elaboration is of little worth.

- If you are still unsure which model to choose specifically for kriging then do a cross-validation and choose the model that produces a mean squared error closest to the mean kriging variance; we describe it below after we have introduced kriging.

Likelihood methods of choosing and fitting models are also gaining ground among geostatisticians, especially to incorporate trend and external drift (Kerry and Oliver, 2007a; Lark, 2012; Lark and Webster, 2006; Lark et al., 2006). However, in 90% of investigations the above approach should be satisfactory if applied with understanding.

### 3.2. Reliability of the empirical variogram

Several factors affect the reliability of the experimental variogram; they include the following:

- Size of sample.
- Lag interval and bin width.
- Marginal distribution of the data.
- Anisotropy.
- Trend.

We consider these in turn.

#### 3.2.1. Sample size

The most important factor determining the reliability, or accuracy, of the empirical variogram and over which we have control is the size of the sample on which it is based. In general the more data you have the greater is the accuracy. Fig. 2, computed by repeated grid sampling from a large two-dimensional simulated field, shows how the confidence intervals narrow as the number of data increases and the sampling interval decreases. Note also

- that the same data are used many times over in the calculation of the semivariances with the result that estimates are correlated with one another, and therefore one cannot use the classical formula for confidence intervals based on  $\chi^2$ ;
- that the confidence intervals widen as the lag distance increases; and
- that the confidence intervals for the smaller samples are wide at all lags; clearly variograms computed from fewer than 100 data (see Fig. 2) are unreliable, a claim we have made forcibly before (Webster and Oliver, 1992).

In the above illustration, the sampling interval decreases as the size of sample increases. As a result the spread of values of  $\hat{\gamma}(\mathbf{h})$  narrows at short lags. But the sampling interval itself is important for another reason; it determines the utility of the empirical variogram. If the interval is larger than the correlation range of the process or in the realization then the empirical variogram will be flat: 'pure nugget' in the jargon. It is useless for prediction and tells us only that all variation occurs within a shorter distance. Fig. 3 derived from Oliver's first survey of the soil in the Wyre Forest of England is an example (see Oliver and Webster, 1987). The sampling interval was approximately 165 m; later after more intense sampling the spatial correlation was found to extend to no more than about 70 m.

We can further illustrate the effect of sample size and intensity with actual data from the survey of the 80-ha Broom's Barn Farm in Suffolk (Webster and McBratney, 1987). The topsoil was sampled at 40-m intervals on a square grid to give 434 data. The exchangeable potassium

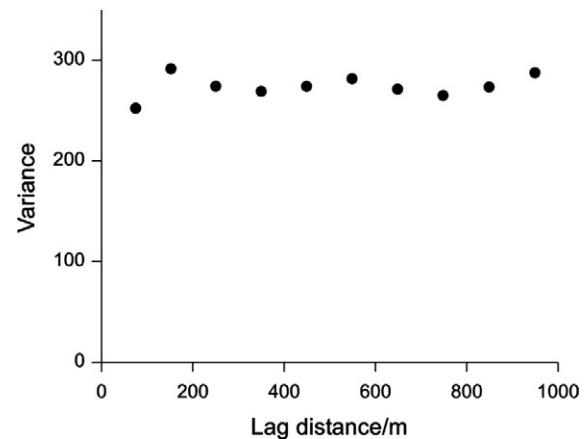


Fig. 3. Experimental variogram of topsoil sand from a survey in the Wyre Forest, England.

( $K^+$ ) concentration was measured, and Table 1 summarizes the statistics. The data were transformed to common logarithms ( $\log_{10}$ ) to stabilize the variance because the skewness coefficient is 2.04 (see Section 3.2.3 below). Fig. 1(b) shows the experimental variogram computed from all 434 data plotted as a series of points; they follow a smooth progression to which the spherical model fits well. In Fig. 1(a) the variogram has been computed from only 87 points. The progression is now erratic because the estimates are less reliable, and the fitted spherical model appears poor and we cannot be confident that we have chosen wisely. Fig. 1(c,e) shows that the exponential and power functions, respectively, appear to fit this experimental variogram equally well.

Of the three graphs on the left-hand side of Fig. 1, the exponential model, Fig. 1(c), appears to provide the best fit visually to the experimental values, whereas the power function, Fig. 1(e), appears to fit the least well. For comparison, we fitted these functions to the variogram of the full set of data, Fig. 1(d,f). The power function, Fig. 1(f), deviates from the experimental values at both the short and long lag distances. The lack of fit of the exponential function, Fig. 1(d), is less obvious but perceptible. The choice of function for the full set of data could be done visually, but for the sub-sample diagnostic information from the model fitting is needed. Table 2 gives the model type, the parameters and the diagnostics. For the full set of data, the spherical function clearly fits the smallest residual sum of squares (RSS) and the largest percentage variance explained. For the subset of 87 there is little to choose between the three, but the diagnostics provide a weak indication in favour of the exponential function.

Most introductory statistical texts instruct their readers to sample at random so as to obtain unbiased estimates, and they describe ways of designing schemes for the purpose. In geostatistics the randomness is assumed to be in the underlying process – it is part of the model – and we can take a more relaxed attitude to sampling provided we avoid bias. Grids are convenient in the field and provide even cover for kriging. As above, however, if they are coarse they might miss the short-range variation crucial for estimating the most important part of the variogram. They are best supplemented with extra sampling points within the grid,

Table 1  
Summary statistics of exchangeable potassium ( $K^+$ ) at Broom's Barn.

	$K^+/\text{mg l}^{-1}$	$\log_{10}(K^+)$	$K^+/\text{mg l}^{-1}$	$\log_{10}(K^+)$
Number of data	434	434	87	87
Minimum	12.0	1.079	14.0	1.146
Maximum	96.0	1.982	70.0	1.845
Mean	26.3	1.398	26.7	1.404
Median	25.0	1.398	26.0	1.415
Std dev.	9.039	0.134	9.403	0.138
Variance	87.71	0.0180	88.42	0.019
Skewness	2.04	0.39	1.76	0.39

**Table 2**  
Parameters of models fitted to experimental variograms of  $\log_{10}K^+$  at Broom's Barn.

Data set and model	Estimates of parameters						Diagnostics	
	$c_0$	$c$	$r/m$	$a/m$	$\beta$	$\alpha$	RSS <sup>a</sup>	Variance explained/%
<i>Full set (434)</i>								
Spherical	0.00453	0.01524	397				0.00368	99.4
Exponential	0.00158	0.01981		180			0.01943	96.7
Power	0				0.00778	0.388	0.05727	91.2
<i>Subset (87)</i>								
Spherical	0.00811	0.01151	376				0.01184	40.9
Exponential	0.00135	0.01601		110			0.01127	43.8
Power	0				0.01092	0.248	0.01254	43.1

The symbols for the parameters are defined in the text.

<sup>a</sup> RSS is the sum of squares of the residuals from the fitted function.

either at shorter intervals along the rows and columns of the grid or in a nested arrangement. Webster and Lark (2013) illustrate several options.

3.2.2. Lag interval and bin width

For data on a regular grid or at equal intervals on transects the natural increment in the experimental variogram is one interval. Where data are irregularly scattered, the comparisons must be grouped by distance and perhaps by direction also; Fig. 4 shows the geometry of the grouping. The practitioner must choose both the length of the step,  $h$ , and the width of the bin,  $w$ , within which the squared differences are averaged for each step. Usually the two are coordinated such that each comparison is placed in one and only one bin. Choosing  $h$  and  $w$  requires judgement. If  $h$  is short and  $w$  is narrow then there will be many estimates of  $\gamma(\mathbf{h})$ , each based on few comparisons and subject to large error, and the variogram will appear 'noisy'. If in contrast  $h$  is large and  $w$  is wide there might be too few estimates of  $\gamma(\mathbf{h})$  to reveal the form of the variogram. You should graph the experimental values so that you can select sensibly.

3.2.3. Marginal distribution

Another attribute of data that can affect the reliability of variograms is the marginal distribution. Long upper tails in the distributions of positively skewed data inflate variances and distort the variogram. We recommend that you compute histograms and box-plots and calculate the skewness coefficient to assess the distribution.

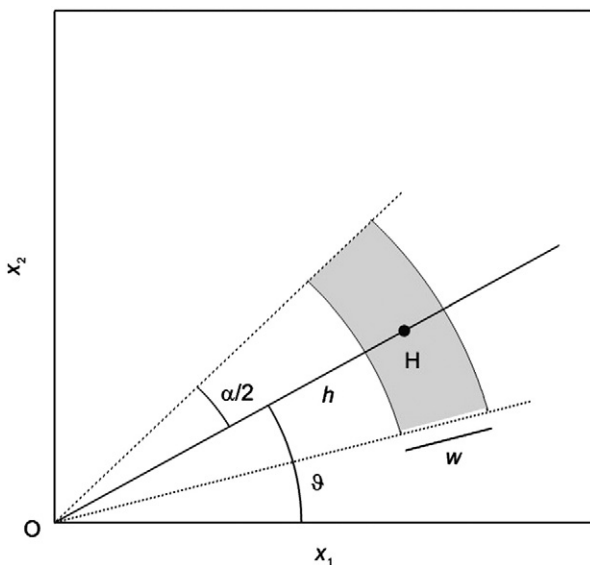


Fig. 4. The geometry in two dimensions for discretizing the lag into bins by distance and direction. The shaded area is one bin.

The distributions of many environmental variables are positively skewed, some strongly so. In many instances they are approximately log-normal, as Ahrens (1965) noted for the chemical elements in the earth. If the skewness coefficient of your data exceeds 1 then try fitting a three-parameter log-normal curve to them, and if that fits well then you would do well to transform the data to logarithms and do all further analyses on the logarithms, as in the example of exchangeable  $K^+$  in the soil at Broom's Barn Farm. After kriging (see below) you may wish to transform your predictions back to the original scale of measurement, and there are standard formulae for that (see Webster and Oliver, 2007, page 185). If the skewness is less pronounced, with a coefficient between 0.5 and 1, transformation to square roots might normalize the distribution approximately. Other variables, such as proportions of particle-size fractions in the soil, are constrained between 0 and 1, and for these you might find it efficient to transform them to angles by  $\varphi = \arcsin\sqrt{q}$  where  $q$  is the proportion.

Outliers cause more serious distortions in geostatistics. They are not simply extremes or near extremes in a frequency distribution, but are unexpectedly large or small values (Barnett and Lewis, 1994). They are values that seem highly unlikely to belong to the populations of interest. Examples include the phosphorus concentration in the soil under

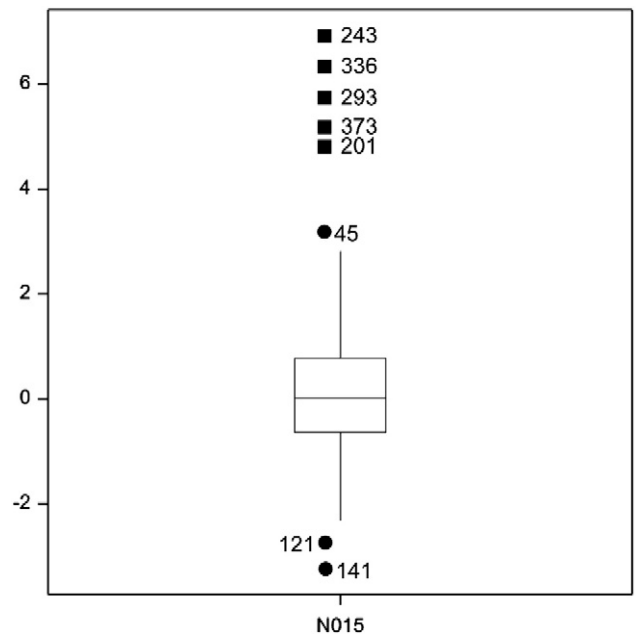
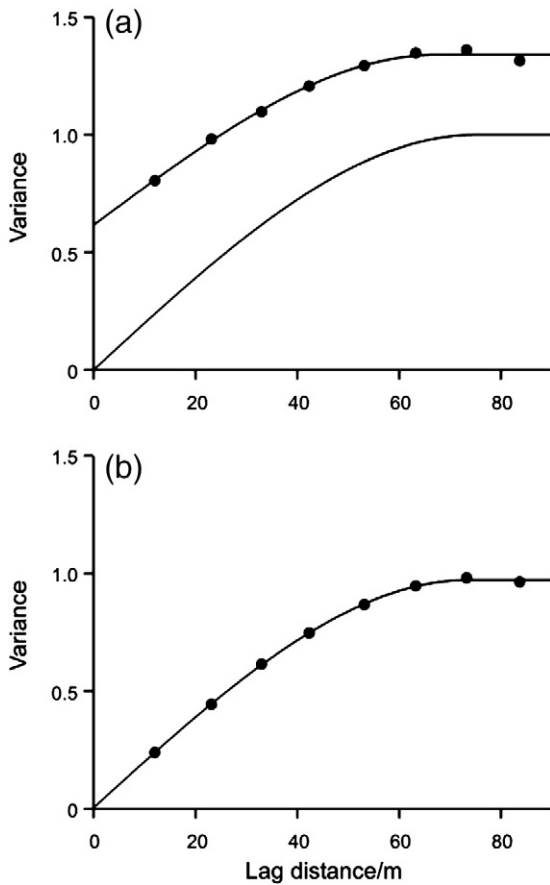


Fig. 5. Box-plot computed from a field of 400 values simulated with a spherical variogram function with  $c_0 = 0$ ,  $c = 1$  and  $r = 75$  and contaminated with five outliers, ■, which are three times beyond the interquartile range, and ● are the near outliers.



**Fig. 6.** Experimental variograms (symbols) and fitted models (solid lines) computed from a field of 400 values simulated with a spherical variogram function with zero nugget (lower line in (a)): (a) contaminated with five outliers and (b) with the outliers removed.

former dung heaps in arable fields and hydrocarbons left by spilled oil on derelict industrial land (brown-field). They can also be mistakes in a chemical analysis or transcription of data. Like data in long tails of a distribution they inflate variograms; but in these instances they can falsify descriptions of the processes of interest, and they certainly do if they are mistakes.

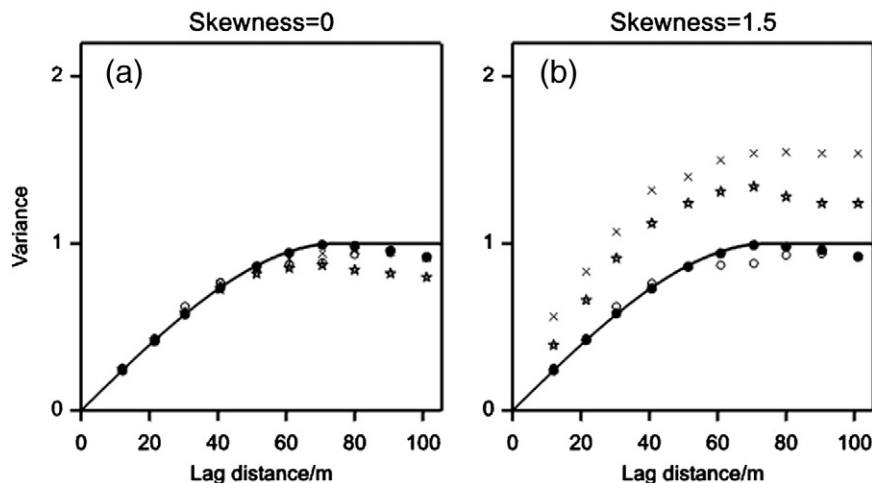
A histogram or box-plot should help you to identify outliers beyond the limits of the main distribution. Having identified an outlier you must then decide what to do with it. If it seems a mistake then remove it from the data or replace it with a correct value. If it seems genuine but seems

to belong to some population other than the one you are investigating then again remove it. These instructions might seem obvious when stated, but many authors do not notice such anomalies until referees and editors point them out.

To show the effect of outliers [Kerry and Oliver \(2007b\)](#) created a normally distributed random field,  $\mathcal{N}(0,1)$ , of 400 values on a 10-m grid generated by a spherical function with  $c_0 = 0$ ,  $c = 1$  and  $r = 75$  m, Equation (7). Five grid nodes were contaminated by a secondary process drawn from a normally distributed random population with unit variance and mean of 1.5. These values were added to the original values of the primary process to give values greater than 4.0 and a skewness coefficient of 1.5. [Fig. 5](#), a box-plot of the field, shows these clearly. We computed and modelled an experimental variogram from all the values, [Fig. 6\(a\)](#). The nugget variance has increased dramatically to 0.617 and the sill variance to 1.341, showing the effect of adjacent disparate values, and both much larger than those of the generator, shown by the lower line. [Fig. 6\(b\)](#) shows the experimental variogram and model for the same field, but with the outliers removed. The parameters are almost identical to those of the generator, namely  $c_0 = 0$ ,  $c \approx 1.0$  and  $r = 73.6$  units. Clearly, a variogram containing outliers would mislead us if those outliers belong to a process other than the one in which we are interested; we should remove them if we suspect them as contaminants.

There are situations, however, in which it is difficult to decide whether what seem to be outliers belong to a different process or where the locations of areas with very large values of some pollutant need to be known. In such situations you may wish to retain these values when you subsequently kriged, and if you do then you should use one of the robust variogram estimators, such as those of [Cressie and Hawkins \(1980\)](#), [Dowd \(1984\)](#) and [Genton \(1998\)](#). All of these downplay the effects of outliers. Their formulae are listed in [Appendix A](#). We illustrate their merits below.

Experimental variograms were computed with the three robust estimators above and Matheron's method of moments for the simulated fields with skewness coefficients of zero and 1.5. [Fig. 7\(a\)](#) shows that the experimental values for the four estimators give similar results and follow the spherical generating function of the simulated field closely. For the field contaminated with five outliers, both variograms computed by the Matheron and Genton estimators, [Fig. 7\(b\)](#), depart considerably from the generating function, but the sill of the former is the more inflated. The variograms estimated by Cressie and Hawkins's and Dowd's methods, [Fig. 7\(b\)](#), are similar to each other and to the generating function. These variograms could be modelled and used for kriging with the outliers remaining in the data. Although these results illustrate the success of the Cressie and Hawkins and Dowd estimators, you should try them all to see which performs best for your data.



**Fig. 7.** Experimental variograms computed by Matheron's method of moments ( $\times$ ) and the robust estimators of Cressie and Hawkins ( $\bullet$ ), Dowd ( $\circ$ ) and Genton ( $\ast$ ).

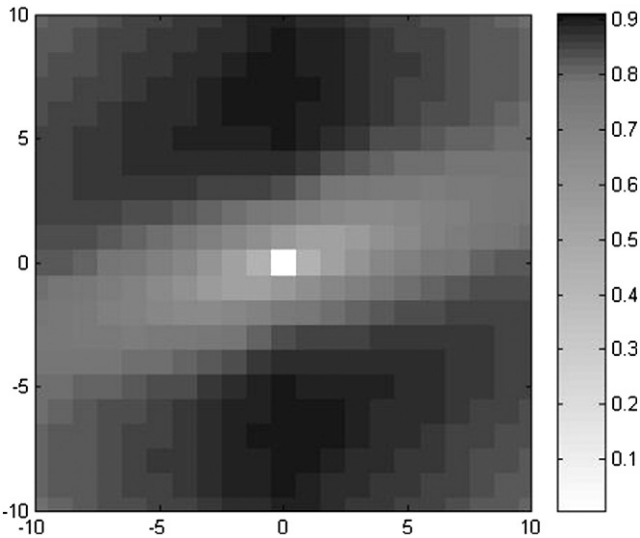


Fig. 8. Two-dimensional anisotropic experimental variogram of a simulated field of 100 000 values computed to 11 intervals on the principal axes.

3.2.4. Anisotropy

In many instances variation is anisotropic, yet investigators either fail to detect it, disregard it when they have detected it or model it improperly. The result is that they present misleading views of the true situations.

If the anisotropy is geometric, which is often approximately the case, then a simple linear transformation of the spatial coordinates will make the variation isotropic. The equation for the transformation is

$$\Omega(\vartheta) = \{A^2 \cos^2(\vartheta - \varphi) + B^2 \sin^2(\vartheta - \varphi)\}^{1/2}, \tag{10}$$

where  $\Omega(\vartheta)$  defines the anisotropy,  $\varphi$  is the direction of maximum continuity and  $\vartheta$  is the direction of the lag. For a spherical or exponential variogram,  $A$  is the distance parameter in the direction of greatest continuity, i.e. the maximum value, and  $B$  is the distance parameter in the perpendicular direction, the direction of least continuity or greatest variation, and is the minimum. For a power function, the roles of  $A$  and  $B$  are

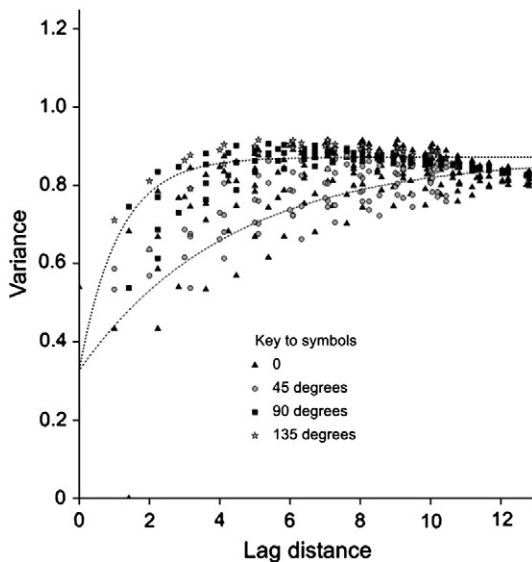


Fig. 9. The same experimental variogram as in Fig. 8 but with values projected into one dimension and the envelope of the fitted anisotropic exponential model shown by the dashed lines. The directions are in 45° segments anticlockwise from  $\varphi = 23.8^\circ$ .

Table 3

Diagnostics from cross-validation of variogram models fitted to experimental variograms of  $\log_{10}K^+$  at Broom's Barn and typical punctual kriging variances.

Data set and model	ME	MSE	MSDR	Kriging variance <sup>a</sup>
<i>Full set (434)</i>				
Spherical	0.000363	0.00761	1.031	
Exponential	0.002321	0.00744	0.590	
Power	0.000827	0.00753	0.634	
<i>Subset (87)</i>				
Spherical	0.001394	0.01314	0.977	0.00866
Exponential	0.004374	0.01366	0.703	0.00159
Power	0.001934	0.01341	0.903	0.01100

<sup>a</sup> At the centres of grid cells surrounded by data at 16 grid nodes.

reversed:  $A$  has the larger gradient in the direction of the greatest rate of change and  $B$  has the smaller gradient in the direction of least change.

Fig. 8 displays a two-dimensional experimental variogram in which the effective range varies with changes in direction. It shows semivariances, shaded from light (small) to dark (large), that have been computed from a field of 100 000 simulated values on a grid of  $100 \times 100$  with unit interval generated with an anisotropic exponential function. The anisotropy is evident, with continuity most pronounced in the direction approximately  $-20^\circ$  ( $\approx 70^\circ$  clockwise from north).

Fig. 9 shows the same semivariances projected into the familiar one-dimensional form with lag distance on the abscissa. It also shows by dashed lines the envelope of the fitted anisotropic exponential model:

$$\gamma(h, \vartheta) = c_0 + c \left[ 1 - \exp \left\{ - \frac{|\mathbf{h}|}{\Omega(\vartheta)} \right\} \right], \tag{11}$$

where  $|\mathbf{h}| = h$  is the lag distance,  $c$  is the maximum of the correlated variance and  $\Omega(\vartheta)$  is as defined in Equation (Deutsch & Journel, 1998) above. Table 4 lists the values of the parameters of the model.

For irregularly scattered data, one must group the separations by both direction and distance, as in Fig. 4. The angle,  $\alpha$ , within which data are included in estimating the semivariance should initially allow complete cover, e.g.  $\alpha = \pi/4$  for four principal directions; all comparisons will then fall within one and only one of those directions. The procedure loses some directional information. If it reveals anisotropy then you should narrow  $\alpha$  to emphasize its expression and hope to obtain a better idea of the direction of maximum continuity,  $\varphi$  in Equation (10). As  $\alpha$  becomes smaller, however, the number of paired comparisons in each estimate of  $\gamma(\mathbf{h})$  becomes fewer and the error in it increases, unless you increase the size of the sample. Binning is a compromise, but it always leads to an underestimate of the anisotropy ratio,  $A/B$ , even with very large samples.

3.2.5. Trend

Trend in geostatistics is gradual variation in space. It adds to the random variation that we have already encountered. Equation (6) estimates the theoretical variogram,  $\gamma(\mathbf{h})$ , only where the underlying process is wholly random. We can represent the two together theoretically by the model:

$$Z(\mathbf{x}) = u(\mathbf{x}) + \varepsilon(\mathbf{x}). \tag{12}$$

Table 4

Parameters of the anisotropic exponential function fitted to the experimental variogram of the simulated anisotropic field, Figs. 9 and 10, accounting for 76.6% of the variance.

$c_0$	$c$	$\varphi$	$A$	$B$	$A/B$
0.327	0.545	23.8°	4.261	1.246	3.82

The symbols for the parameters are defined in the text.



This model contains the term  $u(\mathbf{x})$  for the trend to replace the constant mean  $\mu$  in Equation (1) because with trend the mean varies according to geographical position in a predictable way. In practice it presents us with a problem: how do we decide whether there is trend? There is not necessarily an easy answer; what may appear as smooth variation at one scale can appear as random at another. Only if the trend is global is the answer easy. You should be able to detect a global trend by mapping the data with a suitable graphics program. If the map shows gradual continuous change rather than a patchy distribution then you have identified trend. You should suspect trend if an experimental variogram increases without bound ever more steeply as the lag distance increases. You must take it into account, otherwise it will affect the reliability of any kriged estimates.

You should be able to confirm the presence of trends by fitting a simple trend surface, say linear or quadratic, and mapping the residuals. Such a surface is not optimal in the sense we have used it, however; the variance of the residuals is not minimized, and the residuals themselves are not independent of one another. What is more, these residuals are not those of the true residuals  $\varepsilon(\mathbf{x})$  above, for which the variogram is

$$\gamma(\mathbf{h}) = \frac{1}{2} E \left[ \{ \varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x} + \mathbf{h}) \}^2 \right]. \quad (13)$$

Where there is a dominantly linear global trend (by 'global' we mean a trend that extends throughout the region of interest) a statistically sound procedure is to identify the principal direction of the trend and compute the experimental variogram in the direction perpendicular to it. This variogram is free from trend. Its weakness is that the computation can use only a small proportion of the comparisons, and so there are likely to be large errors in the estimates from small sets of data.

A more general solution to the problem is to estimate the trend and the variogram of the random residuals from it simultaneously. The current best practice is to do this by residual maximum likelihood (REML) (Lark, 2012; Lark et al. 2006; Webster and Oliver 2007). In this short guide we cannot describe this method in detail. Instead we illustrate the problem and its solution with data from a case study by Oliver and Carroll (2004) of a field on the Yattendon Estate, Berkshire, England.

The soil's apparent electrical conductivity ( $EC_a$ ) was recorded at the nodes of a 30-m grid with additional samples along transects at 15-m intervals. The experimental variogram of the data, shown by the grey triangles in Fig. 10, increases without bound, and on the right-hand

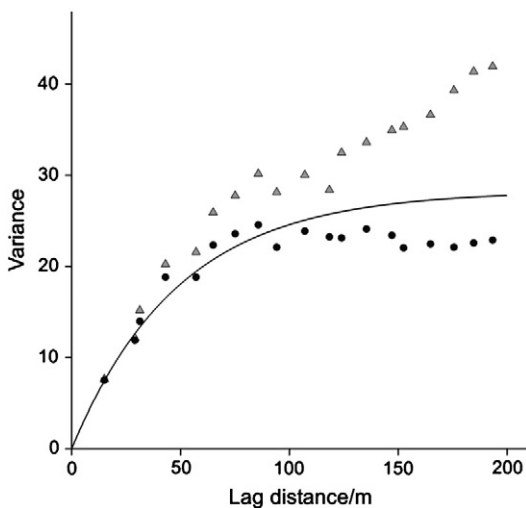


Fig. 10. Variograms computed from the soil's apparent electrical conductivity ( $EC_a$ ) measurements in a field on the Yattendon Estate, Berkshire, England:  $\Delta$  experimental variogram of the original values,  $\bullet$  experimental variogram of OLS residuals and the variogram estimated by REML, the line.

side of the figure the semivariances appear to increase at an ever increasing rate. If we fit a global trend surface, in this case quadratic, to the data by ordinary least squares analysis (OLS) then we can obtain an experimental variogram of the residuals from it. This is shown as the black discs in Fig. 10. A mixed model of quadratic trend and exponential variogram of the random residuals optimized by REML gives the exponential function shown by the solid line; the parameters of this function are a sill  $c = 28.26$  and distance parameter  $a = 69.4$  m.

As it happens in this instance the variogram of the REML residuals is little different from that of the residuals from an ordinary quadratic trend surface at the short lags. As the lag distance increases, however, the experimental semivariances of the OLS residuals increasingly underestimate the true values; the variogram becomes increasingly biased, as explained by Cressie (1993). If you were to krig the residuals, the estimates would be unbiased but the kriging variances would be underestimated.

Lark and Webster (2006) found larger discrepancies between the variogram of OLS residuals and one estimated by REML when analysing data on the depths of the Lower Chalk beneath the Chiltern Hills in south east England and that the kriging variances were seriously underestimated by the simple OLS technique. In Fig. 10, the underestimation by OLS is not so severe. Nevertheless, we now know that the OLS approach is not best practice and should be replaced by REML. Unfortunately, we know of no GISs that have the facility for doing so, and even the professional statistical programs GenStat, SAS and R have limited pre-programmed options.

### 3.3. Inference

Numerous authors of the papers we see want to draw inferences from the variograms they have computed. They should do so with caution, however; they will usually need additional information from the field or general back-ground knowledge to gain insight into the physical, geological or geomorphic processes.

For example, Webster and Cuanalo (1975), studying the soil over the Jurassic sediments in the English midlands, obtained experimental correlograms of several soil properties that decreased to minima at approximately the same lag distance, beyond which the correlations remained nearly constant. They interpreted that lag, which we now recognize as the correlation range, as an expression of the average width of outcrops of the strata in the region. Their interpretation proved to be correct when matched to the geological map of the region. But as Lark (2010) pointed out, such forms of correlograms could equally well arise from a moving average process (Webster and Oliver, 2007, pages 85 and 86); there is no way of distinguishing from the variograms alone without other information.

A second example arises when a variogram increases without bound. Some authors see this as a sign of trend. If the rate of that increase itself increases as a power of  $|\mathbf{h}|$ , as in the function  $\gamma(|\mathbf{h}|) = b|\mathbf{h}|^\eta$  with  $\eta \geq 2$ , then there is almost certainly a trend. If  $\eta < 2$  then greater caution is needed. The outcome of simple Brownian motion, a purely random physical process, has an unbounded linear variogram: i.e. a power function with exponent  $\eta = 1$ . With stronger correlation in the process  $\eta$  will lie between 1 and 2. So any inference to be drawn on trend must be matched by other evidence, and the best way to see that is by mapping the variable of interest.

We mention one other matter under this heading. It is the use, or rather abuse ad nauseam, of the nugget-to-sill ratio of the empirical variogram. The nugget of this variogram is the intercept of the fitted model on the ordinate, and the sill, if it exists, is the fitted upper bound. Numerous authors have been deluded and continue to be deluded into using this ratio as a measure of spatial dependence. Its source has three flaws.

1. The empirical nugget variance embodies measurement error plus error arising from the uncontrolled fitting over distances shorter

than the shortest lag for which there are estimates of  $\gamma(\mathbf{h})$ . It also depends on the choice of model, as is evident in Fig. 1, in particular on the left-hand side in which the three equally plausible models have disparate nugget variances. For the spherical model the ratio is 0.41, whereas for the exponential model it is 0.07; for the power model it cannot be calculated at all, of course.

2. The sampling fluctuation at long lags (Fig. 2) means that the fitted sill is an uncertain estimate of the sill of the underlying process.
3. The ratio takes no account of the correlation range, which can lead to sensible inference, as mentioned above.

This ratio of the empirical variogram undoubtedly has important practical implications for prediction, and we describe them below in the section on kriging. It also has merit in mining where gold nuggets, which are typically isolated and unpredictable discontinuities in ore bodies, give rise to real nugget variances. A large ratio can warn us of large measurement errors or of the need to sample more densely, or both. But the ratio tells us nothing about the underlying variation of properties of the soil or water that vary continuously in space. Variables such as the potassium concentration and water potential of the soil do not have 'nuggets'; they do not contain what acoustic engineers call 'white noise'. The nugget-to-sill ratio of the empirical variogram is simply that, and you should not use it for inferences about the underlying processes.

#### 4. Kriging

As mentioned above, kriging is a generic term for a range of least-squares methods to provide the best linear unbiased predictions (BLUP), best in the sense of minimum variance. Ordinary kriging in Matheron's (1965) original formulation is the most popular, and with good reason; it serves well in most situations with its assumptions easily satisfied. That is why it is often regarded as the 'work-horse' of geostatistics. It requires only knowledge of the variogram function and data for its implementation. It is also robust with regard to moderate departures from those assumptions and a less than optimal choice of model for the variogram.

Kriging solves a set of linear equations, known as the kriging system, which contain semivariances drawn from a fitted variogram function. The equations for ordinary kriging are set out in Appendix A. If that were all then kriging could be completely automatic, but there are essential choices that must be made and we focus on these below.

##### 4.1. Punctual or block kriging

You must first choose whether to predict values at points (punctual kriging) or over larger blocks (block kriging). In punctual kriging the points, denoted  $\mathbf{x}_0$ , have the same supports as the data,  $\mathbf{x}_i$   $i = 1, 2, \dots$ . Geostatisticians' clients more often want predictions for larger bodies; they want averages for areas or volumes larger than the punctual supports on which the measurements are made. In Appendix A we denote these larger bodies by  $B$ , which may be of any reasonable size and shape.

In mining, for example, a company is likely to want to know how much metal there is in blocks of rock the size of a truck-load so as to decide whether to extract them, and, if it has to extract them, whether to send them to a processing plant or to waste. In environmental protection, agencies want estimates of pollutants in blocks that are of a size convenient for remediation. In soil science the principal reason for block kriging is to help farmers with their applications of agricultural chemicals such as lime, fertilizers and pesticides, a practice that is becoming increasingly important in precision agriculture (Oliver, 2010). The size of the block is then determined by the width of farmers' machinery. In Europe this is typically 24 m wide, so blocks should be of that width also.

The choices here depend largely on the purpose for which predictions are required, and for practical purposes block kriging will usually be more appropriate than punctual kriging.

We draw attention to another distinction. In punctual kriging the kriging variance,  $\sigma^2(\mathbf{x}_0)$ , includes the nugget variance of the empirical variogram (except at data points themselves), and that limits the precision of prediction; the kriging variance cannot be less than the nugget variance.

The best fitting model may have a larger nugget variance than other reasonable models, and consequently when used for kriging it can lead to larger estimates of the true absolute errors of predictions than others would. Fig. 1b illustrates the situation; the spherical model fits best overall, but its nugget variance is the largest.

In block kriging the nugget variance is subsumed into the within-block variance and so does not contribute to the kriging variance. Block kriging variances are therefore typically less than the corresponding punctual kriging variances from otherwise similar kriging systems. How much less depends on the sizes of the nugget variances, and they depend on the models you chose for the empirical variograms. Now, however, the model of Fig. 1(b) is likely to lead to underestimates of the true absolute errors.

##### 4.2. Kriging neighbourhood

The kriging systems, Equations (22) and (26) in Appendix A contain the semivariances for all pairs of  $N$  data points. In other words they imply the use of all the available data. Practically, however, this is hardly ever the case for at least three reasons.

1. Kriging is a local predictor, and only the nearest few points to the target point or block carry significant weight, especially if the nugget variance is a small proportion of the total variance. This means that we can replace  $N$  in kriging equations by a much smaller number of data,  $n \ll N$ , within the kriging neighbourhood. This has the advantages of reducing computing time with large sets of data and of the inversion of much smaller matrices, thereby avoiding the instability that can arise with large matrices.
2. The local nature of ordinary kriging means that we can restrict the assumption of stationarity of the mean to that within neighbourhoods.
3. The local nature of kriging means that only semivariances close to the ordinate of the variogram are used, and so one needs to estimate and model the variogram well over the first few lags only.

Mapping programs tend to take advantage of the local nature of kriging and move a window based on a chosen size of neighbourhood over each target point or block in a region. Automation of the procedure carries a risk, however. In the extreme with a variogram that is all nugget the correct prediction would be the same everywhere; it would be the mean of the data. If you were to krig with a moving window the result would be a surface with fluctuations that depended solely on the local values caught in the window as it moved. With a pure nugget variogram it is unwise to attempt any type of interpolation.

There are no rules for defining the kriging neighbourhood, but we provide some guidance below.

1. If the data are dense and the variogram is bounded and has a small nugget variance then the radius of the neighbourhood can be set close to the range or effective range of the variogram because data beyond the range will have negligible weight. The predicted surface should be smooth (except at the data points in punctual kriging).
2. For a variogram with a large nugget variance, the radius of the neighbourhood should be greater than the range because distant points are likely to carry some weight. The same applies if the data are sparse and points beyond the range may carry sufficient weight to be important.

3. You may choose to set the neighbourhood in terms of a minimum and maximum number of nearest data to the target; we usually recommend a minimum  $n \approx 7$  and a maximum  $n \approx 25$ .
4. If the data are very unevenly scattered it is good practice to divide the neighbourhood into octants so that there are at least two points in each.
5. When you start to analyse new data it is instructive to examine what happens to the kriging weights as you change the neighbourhood, especially where the data are irregularly scattered and the neighbourhood moves as it does to estimate a field of values for mapping. We have illustrated a range of situations in Webster and Oliver (2007), pages 160–173.

#### 4.3. Cross-validation

We mentioned above that cross-validation provides a means of choosing among plausible models for variograms, but we delayed its explanation until we had dealt with kriging. The technique is simple: each and every one of the  $N$  data points is omitted in turn from the set of data and its value there is predicted by ordinary punctual kriging with the proposed model. Three statistics are then calculated; they are the mean error (ME), the mean squared error (MSE) and the mean squared deviation ratio (MSDR), which is the mean of the squared errors divided by the corresponding kriging variances. We list the equations in Appendix A.

Ideally the mean error is zero: kriging is unbiased, even with a poorly chosen model. So the ME is of little help. Kriging should minimize the MSE, and so the MSE is a better diagnostic; but it will not identify a ‘correct’ model. The most telling of the three criteria is the MSDR. It should equal 1, and so for kriging one might choose the model for which the MSDR is closest to 1.

By applying this technique to the data on  $\log_{10}K^+$  for Broom’s Barn we obtain the results listed in Table 3. The mean errors are all close to 0, as we should expect. The mean squared errors are small and approximately the same for the three models within each set of data; they are clearly larger for the subset, however, as we should expect. The big differences are in the MSDR, and judged on that criterion the spherical function is the one we should use for kriging in both cases.

To complete the picture, in Fig. 11 we graph the observed values for  $\log_{10}K^+$  at Broom’s Barn against the cross-validation predictions obtained with the spherical variogram.

#### 4.4. Smoothing

Kriging smooths; it loses variance. This often surprises investigators, though it should not. It is no different from ordinary least-squares regression in this respect; it tends to under-estimate large values and over-estimate small ones. The effect is evident in Fig. 11. In it the spread of predictions on the abscissa is clearly less than that of the observations on the ordinate. The variance of the observations is 0.01800, whereas that of the predictions is only 0.00990; so kriging has lost approximately half of the variance in the data. In fact, the degree of smoothing depends on the nugget-to-sill ratio; the larger is that ratio the greater is the smoothing, and in the limit when the ratio is 1 the kriged surface is flat, as mentioned above.

Though a kriged surface shows our best estimates it can give a misleading picture of the variation in a region. If you want a realistic picture of the variation then you must switch to simulation. There are several techniques from which to choose; they include the Cholesky decomposition of the covariance matrix, sequential Gaussian simulation and simulated annealing. There is also the method of turning bands, the earliest but little used nowadays. Each run of any of these methods will give its unique outcome and retain the variance in the data, or at least in the variogram model, and a map will show the fluctuation to expect. The simulation can be conditioned on the data so that the true values are

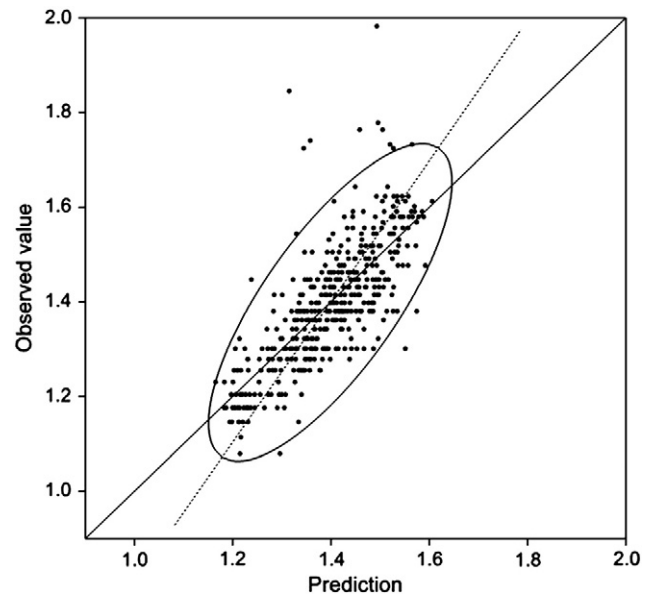


Fig. 11. Scatter diagram of observed values of  $\log_{10}K^+$  at Broom’s Barn plotted against values predicted by ordinary punctual kriging during cross-validation. The solid line (approximately 1:1) is the regression of observed on predicted values, the dashed line is the principal axis. The variances are 0.01800 and 0.00990 respectively. From Webster and Oliver (2007).

retained in the outcome and the fluctuations never stray far from actuality.

Some practitioners run numerous conditional simulations and from them compute averages. You should realize that such averages converge on kriged estimates as the number of simulations increases. If you want averages then krige directly; you cannot do better by simulation.

#### 4.5. Kriging in the presence of trend

Matheron (1969) recognized the problem of dealing with trends and devised universal kriging to take it into account. The method involves an elaboration of the kriging system with semivariances drawn from the variogram of the residuals, i.e. from estimates of Equation (13). What he did not specify was how to find those estimates. A practical solution recommended by Stein (1999) is to compute the empirical best linear unbiased predictor (E-BLUP) with a variogram estimated by maximum likelihood.

Closely related to kriging in the presence of trend is kriging with external drift. Universal kriging effectively treats the geographic coordinates as auxiliary variables. There can be other variables with which the target variable  $Z$  is simply related and which are known or can be calculated everywhere. These too, known as external drift variables, can be treated as auxiliary variables. Hudson and Wackernagel (1994) presented an early example in which they took into account the height of the land as an auxiliary variable when mapping the annual rainfall in Scotland. Nowadays we can estimate simultaneously the regression relation between the target variable and the auxiliary variable(s) and the variogram of the residuals from that regression by residual maximum likelihood (REML). Lark et al. (2006) described and illustrated the method, and more recently Lark (2012) has summarized it. You can find further details and another example in Webster and Oliver (2007). A stumbling block to its widespread use in the past was the lack of software with the desired flexibility in the public domain. The ‘proc (MIXED)’ procedure in SAS has included some of the standard variogram models for several years (<http://support.sas.com/documentation>). Now the most recent release of GenStat (Payne, 2013) contains the facilities,

as does R (Pinheiro et al., 2013; R Development Core Team, 2010). There is no longer an excuse for inferior practice.

**5. General conclusion**

We close with a list of steps and of what you should report in an investigation that requires only straightforward least-squares geostatistical analysis.

1. Sample sufficiently without bias. For the variogram aim for a minimum of 100–150 points to provide six to ten estimates within the expected effective range. For mapping by kriging sample evenly to give even coverage at intervals of less than half the effective range.
2. Compute the marginal distribution of each variable, identify outliers and decide what to do about them, and transform the data to stabilize variances if necessary. Summarize the statistics and report your treatment of outliers and any transformation.
3. Compute the variogram on the raw or transformed data by the method of moments (or a more robust method), fit plausible models by weighted least squares approximation. State the steps in which lag was incremented, your binning and the model you finally choose, and display that model on a graph of the experimental variogram.

If you identify a trend then estimate the trend and variogram of the residuals from the trend by REML. State the equation for the trend and display the variogram of the residuals.

4. Krige at points or over blocks of a size suitable for the application. In the absence of trend use ordinary kriging; if there is a trend then use universal kriging. Krige from a global kriging system or local ones in a moving window. State your choices.
5. Map the kriged estimates and their associated kriging variances.

Finally, bear in mind that in writing for *Catena* you should discuss the relevance of your results. What have you learned about the soil, geomorphology or hydrology that was not known before? Or perhaps, how might your results be used to manage land or water resources? Telling us no more than what you did is not enough. Geostatistics is a technological means to an end; it is not an end in itself.

**Acknowledgements**

We thank Dr R. Kerry for the simulated anisotropic field, Dr B.P. Marchant for fitting the REML variogram (Fig. 10), Dr A.E. Milne for Fig. 8 and the Home-Grown Cereals Authority for its support in obtaining the data on the Yattendon Estate. All the other analyses have been programmed by us in GenStat (Payne, 2013) with specifically its FITNONLINEAR directive to fit the models to the variograms.

**Appendix**

*Models for variograms*

The equations of the popular models for variograms are given below in their isotropic form with lag in distance only, i.e. for  $h = |\mathbf{h}|$ , and with an uncorrelated nugget component,  $c_0$ .

*Spherical*

$$\begin{aligned} \gamma(h) &= c_0 + c \left\{ \frac{3h}{2r} - \frac{1}{2} \left( \frac{h}{r} \right)^3 \right\} \text{ for } 0 < h \leq r \\ &= c_0 + c \text{ for } h > r \\ &= 0 \text{ for } h = 0. \end{aligned} \tag{14}$$

Here  $c$  is the spatially correlated variance, and  $r$  is the range. The quantity  $c_0 + c$  is known as the ‘sill’.

*Exponential*

$$\begin{aligned} \gamma(h) &= c_0 + c \left\{ 1 - \exp\left(-\frac{h}{a}\right) \right\} \text{ for } 0 < h \\ &= 0 \text{ for } h = 0. \end{aligned} \tag{15}$$

The parameter  $c$  is as for Equation (Cressie, 1985), and  $a$  is a distance parameter. The function approaches its sill asymptotically and so has no finite range. Its effective range is usually taken as  $r' = 3a$ .

*Gaussian*

$$\begin{aligned} \gamma(h) &= c_0 + c \left\{ 1 - \exp\left(-\frac{h^2}{a^2}\right) \right\} \text{ for } 0 < h \\ &= 0 \text{ for } h = 0. \end{aligned} \tag{16}$$

The parameters are the same as for the exponential model. The effective range, however, is less:  $r' = \sqrt{3}a$ .

*Power*

$$\begin{aligned} \gamma(h) &= c_0 + bh^\eta \text{ for } 0 < h \\ &= 0 \text{ for } h = 0, \end{aligned} \tag{17}$$

with the strict constraint  $0 < \eta < 2$ .

*Robust variogram estimators*

*Cressie–Hawkins (Cressie and Hawkins, 1980)*

$$2\hat{\gamma}_{CH}(\mathbf{h}) = \frac{\left\{ \frac{1}{m(\mathbf{h})} \sum_{j=1}^{m(\mathbf{h})} |z(\mathbf{x}_j) - z(\mathbf{x}_j + \mathbf{h})|^{1/2} \right\}^4}{0.457 + \frac{0.494}{m(\mathbf{h})} + \frac{0.045}{m^2(\mathbf{h})}}. \tag{18}$$

*Dowd (Dowd, 1984)*

$$2\hat{\gamma}_D(\mathbf{h}) = 2.198 \left\{ \text{median} |y_j(\mathbf{h})| \right\}^2, \tag{19}$$

where  $y_j(\mathbf{h}) = z(\mathbf{x}_j) - z(\mathbf{x}_j + \mathbf{h})$ ,  $j = 1, 2, \dots, m(\mathbf{h})$ .  
*Genton (Genton, 1998)*

$$2\hat{\gamma}_G(\mathbf{h}) = \left[ 2.219 \left\{ |y_j(\mathbf{h}) - y_k(\mathbf{h})|; j < k \right\} \left( \frac{H}{2} \right) \right]^2, \tag{20}$$

where  $H$  is the integral part of  $1 + m(\mathbf{h})/2$ .

*The ordinary kriging system*

Let  $z(\mathbf{x}_i), i = 1, 2, \dots, N$ , be observed values of variable  $z$  at points  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ , where in two dimensions  $\mathbf{x}_i \equiv \{x_{i1}, x_{i2}\}^T$ . For any new point  $\mathbf{x}_0$  we predict  $Z$  by

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^N \lambda_i z(\mathbf{x}_i). \tag{21}$$

The  $\lambda_i, i = 1, 2, \dots, N$ , are the weights chosen to minimize the prediction error variance by solution of the following set of equations:

$$\sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_j) + \psi(\mathbf{x}_0) = \gamma(\mathbf{x}_j - \mathbf{x}_0) \text{ for all } j \quad (22)$$

$$\sum_{i=1}^N \lambda_i = 1.$$

Here  $\gamma(\mathbf{x}_i - \mathbf{x}_j)$  is the semivariance between data points  $i$  and  $j$ ,  $\gamma(\mathbf{x}_j - \mathbf{x}_0)$  is the semivariance between data point  $j$  and the target point  $\mathbf{x}_0$ , and  $\psi(\mathbf{x}_0)$  is a Lagrange multiplier introduced for the minimization of the error variance. We can write this in matrix terms thus

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{b}. \quad (23)$$

We invert matrix  $\mathbf{A}$ , and we post-multiply the result by  $\mathbf{b}$  to obtain the optimal weights  $\boldsymbol{\lambda} \equiv \{\lambda_1, \lambda_2, \dots, \lambda_N\}^T$  which we then insert into Equation (21) to obtain the predictions. The prediction error variance is given by

$$\sigma^2(\mathbf{x}_0) = \mathbf{b}^T \boldsymbol{\lambda}. \quad (24)$$

The predictions for blocks,  $B$ , are still weighted sums of the data:

$$\hat{Z}(B) = \sum_{i=1}^N \lambda_i z(\mathbf{x}_i), \quad (25)$$

but now the kriging system to be solved is

$$\sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i - \mathbf{x}_j) + \psi(B) = \bar{\gamma}(\mathbf{x}_j, B) \text{ for all } j \quad (26)$$

$$\sum_{i=1}^N \lambda_i = 1,$$

where  $\bar{\gamma}(\mathbf{x}_j, B)$  is the average semivariance between the data and the target block and  $\mathbf{b}$  now comprises the right-hand sides of Equation (Matheron, 1969). The error variance is then

$$\sigma^2(B) = \mathbf{b}^T \boldsymbol{\lambda} - \bar{\gamma}(B, B), \quad (27)$$

in which  $\bar{\gamma}(B, B)$  is the within-block variance.

#### Cross-validation statistics

Mean error: ME

$$\text{ME} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i). \quad (28)$$

Mean squared error: MSE

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)\}^2. \quad (29)$$

Mean squared deviation ratio: MSDR

$$\text{MSDR} = \frac{1}{N} \sum_{i=1}^N \frac{\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)\}^2}{\hat{\sigma}_K^2(\mathbf{x}_i)}. \quad (30)$$

In these equations  $z(\mathbf{x}_i)$  is the  $i$ th datum at  $\mathbf{x}_i$ ,  $\hat{Z}(\mathbf{x}_i)$  is the kriged prediction there, and  $\hat{\sigma}_K^2(\mathbf{x}_i)$  is the kriging variance.

#### References

Ahrens, L.H., 1965. *Distribution of Elements in Our Planet*. McGraw-Hill, New York.

- Akaike, H., 1973. Information theory and an extension of maximum likelihood principle. In: Petrov, B.N., Csáki, F. (Eds.), *Second International Symposium on Information Theory*. Akadémiai Kiadó, Budapest, pp. 267–281.
- Barnett, V., Lewis, T., 1994. *Outliers in Statistical Data*, third edition. John Wiley & Sons, Chichester.
- Brus, D.J., de Groot, J.J., 1994. Estimation of non-ergodic variograms and their sampling variance by design-based sampling strategies. *Math. Geol.* 26, 437–454.
- Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *J. Soil Sci.* 31, 315–331.
- Chilès, J.-P., Delfiner, P., 2012. *Geostatistics: Modeling Spatial Uncertainty*, second edition. John Wiley & Sons, New York.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data*, revised edition. John Wiley & Sons, New York.
- Cressie, N., 1985. Fitting variogram models by weighted least squares. *J. Int. Assoc. Math. Geol.* 17, 563–586.
- Cressie, N., Hawkins, D.M., 1980. Robust estimation of the variogram. *J. Int. Assoc. Math. Geol.* 12, 115–125.
- De Marsily, G., Ahmed, S., 1987. Application of kriging techniques in groundwater hydrology. *J. Geol. Soc. India* 29, 57–82.
- Deutsch, C.V., Journé, A.G., 1998. *GSLIB: Geostatistical Software and User's Guide*, second edition. Oxford University Press, New York.
- Dowd, P.A., 1984. The variogram and kriging: robust and resistant estimators. In: Verly, G., David, M., Journé, A.G., Marechal, A. (Eds.), *Geostatistics for Natural Resources Characterization*. D. Reidel, Dordrecht, pp. 91–106.
- Gajem, Y.M., Warrick, A.W., Myers, D.E., 1981. Spatial dependence of physical properties of a Typic Torrifluent soil. *Soil Sci. Soc. Am. J.* 45, 709–715.
- Geneton, M.G., 1998. Highly robust variogram estimation. *Math. Geol.* 30, 213–221.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Hudson, G., Wackernagel, H., 1994. Mapping temperature using kriging with external drift: theory and example from Scotland. *Int. J. Climatol.* 17, 77–91.
- Kerry, R., Oliver, M.A., 2007a. 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., 2007b. Determining the effect of asymmetric data on the variogram. II. Outliers. *Comput. Geosci.* 33, 1233–1260.
- Krige, D.G., 1951. A statistical approach to some basic mine problems on the Witwatersrand. *J. Chem. Metall. Min. Soc. S. Afr.* 52, 119–139.
- Krige, D.G., 1966. Two-dimensional weighted moving average trend surfaces for ore evaluation. *J. South. Afr. Inst. Min. Metall.* 66, 13–38.
- Lark, R.M., 2010. Two contrasting spatial processes with a common variogram: inference about spatial models from high-order statistics. *Eur. J. Soil Sci.* 61, 479–492.
- Lark, R.M., 2012. Towards soil geostatistics. *Spat. Stat.* 1, 92–99.
- Lark, R.M., Webster, R., 2006. Geostatistical mapping of geomorphic surfaces in the presence of trend. *Earth Surf. Process. Landf.* 31, 862–874.
- Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On optimal prediction of soil properties in the presence of trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *Eur. J. Soil Sci.* 57, 787–799.
- Matheron, G., 1963. Principles of geostatistics. *Econ. Geol.* 58, 1246–1266.
- Matheron, G., 1965. *Les variables régionalisées et leur estimation*. Masson, Paris.
- Matheron, G., 1969. *Le krigeage universel*. Cahiers du Centre de Morphologie Mathématique, No 1. Ecole des Mines, Fontainebleau.
- McBratney, A.B., Webster, R., McLaren, R.G., Spiers, R.B., 1982. Regional variation in extractable copper and cobalt in the topsoil of south-east Scotland. *Agronomie* 2, 969–982.
- McBratney, A.B., Webster, R., 1986. Choosing functions for semivariograms of soil properties and fitting them to sampling estimates. *J. Soil Sci.* 37, 617–639.
- Olea, R.A., 1999. *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers, Boston.
- Oliver, M.A. (Ed.), 2010. *Geostatistical Applications in Precision Agriculture*. Springer, Dordrecht.
- Oliver, M.A., Carroll, Z.L., 2004. *Description of Spatial Variation in Soil to Optimize Cereal Management*. Home-Grown Cereals Authority, London (Project Report 330).
- Oliver, M.A., Webster, R., 1987. The elucidation of soil pattern in the Wyre Forest of the West Midlands, England. II. Spatial distribution. *J. Soil Sci.* 38, 293–307.
- Payne, R.W. (Ed.), 2013. *The Guide to GenStat Release 16 – Part 2: Statistics*. VSN International, Hemel Hempstead.
- Pannatier, Y., 1995. *Variowin. Software for Spatial Analysis in 2D*. Springer-Verlag, New York.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30, 683–691.
- Pebesma, E.J., Wesseling, C.G., 1998. Gstat: a program for geostatistical modelling, prediction and simulation. *Comput. Geosci.* 24, 17–31.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., The R Development Core Team, 2013. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-110.
- R Development Core Team, 2010. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien. ISBN 3-900051-07-0 (URL <http://www.R-project.org/>).
- Russo, D., 1984. A geostatistical approach to solute transport in heterogeneous fields and its applications to salinity management. *Water Resour. Res.* 20, 1260–1270.
- Stein, M.L., 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York.
- Vauclin, M., Vieira, S.R., Vachaud, G., Nielsen, D.R., 1983. The use of cokriging with limited field soil observations. *Soil Sci. Soc. Am. J.* 47, 175–184.

- Wackernagel, H., 2003. *Multivariate Geostatistics*, third edition. Springer-Verlag, Berlin.
- Webster, R., Cuanalo de la C., H.E., 1975. Soil transect correlograms of north Oxfordshire and their interpretation. *J. Soil Sci.* 26, 176–194.
- Webster, R., Lark, R.M., 2013. *Field Sampling for Environmental Science and Management*. Routledge, London.
- Webster, R., McBratney, A.B., 1987. Mapping soil fertility at Broom's Barn by simple kriging. *J. Sci. Food Agric.* 38, 97–115.
- Webster, R., McBratney, A.B., 1989. On the Akaike information criterion for choosing models for variograms of soil properties. *J. Soil Sci.* 40, 493–496.
- Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* 43, 493–496.
- Webster, R., Oliver, M.A., 1997. Software review. *Eur. J. Soil Sci.* 48, 173–175.
- Webster, R., Oliver, M.A., 2007. *Geostatistics for Environmental Scientists*, second edition. John Wiley & Sons, Chichester.