

# Chapter 3

## The Variogram and Modelling

**Abstract** Accurate estimates of variograms are needed for reliable prediction by kriging and subsequent mapping and for optimizing sampling schemes. Sample variograms are usually computed by the method of moments at a sequence of lags, and one or more ‘authorized’ functions are fitted to them. A variogram may be computed along transects or on grids at regular intervals or in bins from irregularly scattered data. Accuracy of the variogram depends on the size of sample, the number of lags at which it is estimated and the lag interval relative to the spatial scale of variation, the marginal distribution of the variable, anisotropy and trend. Robust estimators can deal with extreme values, outliers. Variograms may be bounded (for second-order stationary processes) or unbounded (intrinsically stationary only), and there are few simple authorized functions for modelling them. The parameters of the models summarize the spatial variation and are needed for subsequent kriging. Computing the variogram in at least three directions can identify anisotropy if it is present. Diagnostics including residual mean squares and the Akaike Information Criterion help in the selection of the best fitting model.

**Keywords** Experimental variogram • Method of moments • Model parameters • Lag interval • Spatial scale • Marginal distribution • Anisotropy • Outliers • Robust estimators • Nugget variance • Sill variance • Model diagnostics

The variogram is the cornerstone of many geostatistical applications. The experimental variogram and any model fitted to it should be accurate. Only then can the model describe the variation reliably. Kriging requires a variogram, and it is by ensuring its accuracy that you will eventually obtain minimum-variance predictions by kriging. If the variogram describes the variation poorly then the kriged predictions are likely to be poor also, and they might have little or no validity no matter how ‘pretty’ the map. The term ‘cartographic pornography’ has been used by those who realize that no confidence can be placed in many of the beautiful smooth maps that exist because of sparsity of the data that underlies them (see Chap. 4). Further, the parameters of the variogram model may be used later for sample design and the kriged estimates for decision-making; computing experimental variograms and modelling them should not be treated in a cavalier fashion.

This chapter illustrates the essential steps in obtaining reliable experimental variograms by Matheron's (1965) method-of-moments (MoM) and modelling them. In some geostatistics packages and several GISs, computing the variogram and kriging from the data is automated. As a consequence of such a 'black box' approach, the variogram is computed and modelled, and the parameter values from the model are inserted into the kriging equations without any intervention or assessment by the user. As a result the user has no idea of the variogram's form (it might even be pure nugget) or whether the model is a good fit. There are many other reasons for poor variograms and their models, for example too few data (Webster and Oliver 1992), unsuitable models, poor fitting, faulty processing and misunderstanding. These are matters that form the basis of this chapter. Our aim is to prevent researchers from wasting time on analyses for which their data are unsuitable, and to guide them through the stages that will ensure that their variograms are 'fit-for-purpose'.

### 3.1 The Experimental Variogram

The first task in turning theory into practice is to estimate the variogram from sample data, say  $z(\mathbf{x}_1), z(\mathbf{x}_2), \dots$ , where  $\mathbf{x}_1, \mathbf{x}_2, \dots$  denote the positions of the sample in two-dimensional space. We assume that those positions have been selected without bias. They need not be random, as in design-based estimation, because we treat the variables as the outcomes of random processes. Therefore, we can take a relaxed attitude to the sampling design, which may be systematic, random, nested or some combination (see Chap. 5). The usual equation to compute the variogram is Matheron's method of moments (MoM) estimator:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2, \quad (3.1)$$

where  $z(\mathbf{x}_i)$  and  $z(\mathbf{x}_i + \mathbf{h})$  are the observed 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  $\mathbf{h}$ . By changing  $\mathbf{h}$  we obtain an ordered set of semivariances; these constitute the experimental or sample variogram. The way that Eq. (3.1) is implemented as an algorithm depends on whether the data are regularly spaced in one dimension, are on a regular grid or are irregularly distributed in two dimensions.

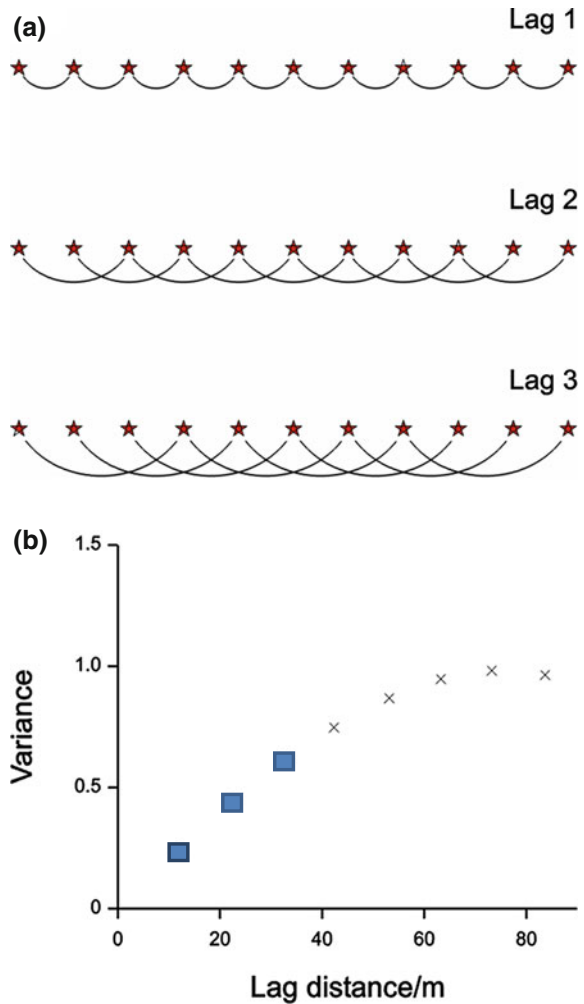
#### 3.1.1 Computing the Variogram from Regular Sampling in One Dimension

Regular sampling in one dimension may be horizontal or vertical (e.g. down boreholes or through the atmosphere) along transects. The lag,  $\mathbf{h}$ , becomes a scalar

$h = |\mathbf{h}|$  that replaces  $\mathbf{h}$  in Eq. (3.1). Semivariances,  $\hat{\gamma}(h)$ , can be computed only at multiples of the sampling interval. Figure 3.1a shows how the comparisons between pairs of points are made; first for  $h = 1$  and then for  $h = 2, 3, \dots$ . This results in a set of semivariances,  $\hat{\gamma}(1), \hat{\gamma}(2), \hat{\gamma}(3), \dots$ , i.e. a one-dimensional experimental variogram which we can plot as a graph of  $\hat{\gamma}(h)$  against  $h$  as in Fig. 3.1b. There may be positions along a transect where, for various reasons, there are no observations. These missing data do not present a problem; they simply result in fewer comparisons for Eq. (3.1).

Transects may be aligned in several directions, for example to identify anisotropy when at least three directions should be used (see Sect. 3.2.5). The same procedure may be used to compute these variograms, and Eq. (3.1) will provide a separate set of estimates for each direction.

**Fig. 3.1** **a** Comparisons for computing a variogram for three lag intervals from a regular sample every 10 m along a transect and **b** semivariances plotted against the first three lag intervals to form the sample variogram (other possible semivariances shown as crosses)



### 3.1.2 Computing the Variogram from Regular and Irregular Sampling in Two Dimensions

Data from regular grid sampling in two dimensions can be analysed in one of three ways. First, the grid can be treated as a series of transects in two dimensions—it is one way in which you can investigate anisotropy, i.e. directional differences, in the variation. The variogram can be computed as above, but in several directions of the grid separately, for example, along the rows and columns of the grid and on the diagonals. Second, the variogram can be computed in two dimensions as follows:

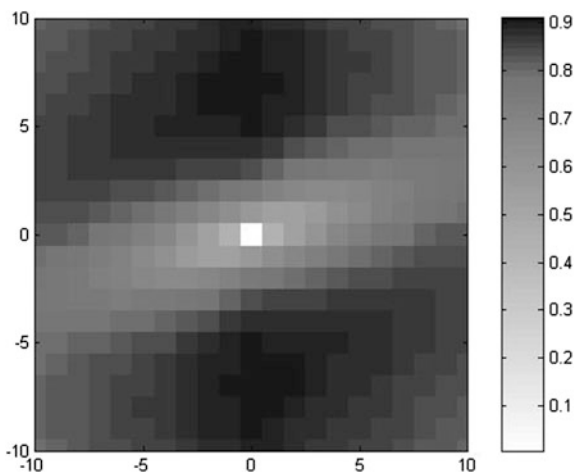
$$\hat{\gamma}(p, q) = \frac{1}{2(m-p)(n-q)} \sum_{i=1}^{m-p} \sum_{j=1}^{n-q} \{z(i, j) - z(i+p, j+q)\}^2, \quad (3.2)$$

$$\hat{\gamma}(p, -q) = \frac{1}{2(m-p)(n-q)} \sum_{i=1}^{m-p} \sum_{j=q+1}^n \{z(i, j) - z(i+p, j-q)\}^2,$$

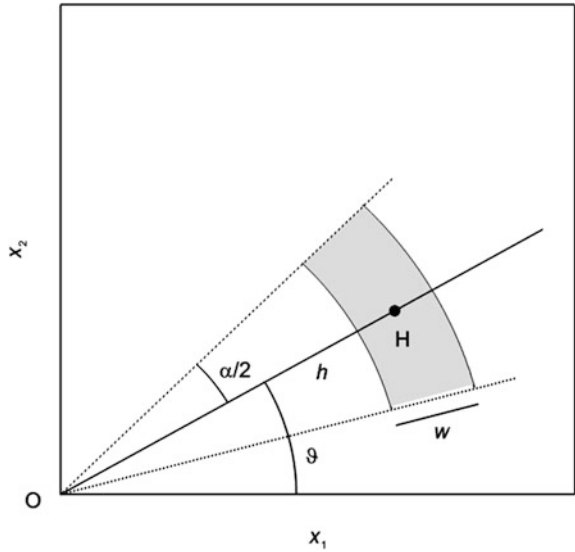
where  $p$  and  $q$  are the lags along the rows and down the columns of the grid, respectively. In general, the lag interval is that of the grid. The variogram is computed for lags from  $-q$  to  $q$  and from 0 to  $p$ . The output from this is then plotted as a two-dimensional variogram as in Fig. 3.2.

Third, the variogram can be computed over all directions (omnidirectional) for both regular and irregular sampling designs. For a grid the initial nominal lag interval should be that of the grid spacing, whereas for irregularly scattered data the choice is wider because the observations may be separated by potentially unique lags in both distance and direction. Figure 3.3 explains how we can obtain semivariances over all directions in two dimensions by placing the lags into bins. We choose a nominal lag interval in both distance and direction as shown in grey in the figure. The width in distance is designated  $w$ , which for irregularly scattered data could be the average

**Fig. 3.2** Two-dimensional anisotropic experimental variogram of a simulated field of 100 000 values computed to 11 intervals on the principal axes



**Fig. 3.3** The geometry in two dimensions for discretizing the lag into bins by distance and direction



distance between neighbouring sampling points. The angular width is the angle  $\alpha$ . All pairs of comparisons that fall within that bin and contribute to  $\hat{\gamma}$  are attributed to its centroid at H with nominal lag  $h$ ,  $\vartheta$  and where  $\vartheta$  is the lag direction. The lag is usually incremented in steps of  $w$  and  $\vartheta$  so that each paired comparison falls into one and only one bin. To compute the omnidirectional variogram, the angular width of the bins is set to  $\alpha = \pi$  (i.e.  $180^\circ$ ). We may also compute the variogram in a set of lag directions,  $\vartheta$ , (see Sect. 3.3.2).

## 3.2 Factors Affecting the Reliability of Experimental Variograms

### 3.2.1 Sample Size

The accuracy of the variogram depends primarily on one's having enough data at a suitable density or separating interval. It also depends on the design or configuration of the sample because of the way that the variogram is usually computed. The random function model (see Chap. 2) enables us to have the multiple realizations required by theory; we treat each comparison between any pair of data as a single realization of the process. Therefore, for every lag interval we require many comparisons to ensure reliability of the estimated semivariances. At the shortest lags or separating distances, we might have rather few paired comparisons for two-dimensional data. As the lag interval between data increases, however, the number of comparisons increases (see Table 3.1). At some distance that depends on the number of data the number of pairs for comparison starts to decrease, although the

**Table 3.1** Lag intervals, semivariances and counts for  $\log_{10} K^+$  at Broom's Barn Farm

Lag/m	Semivariance	Counts	Lag/m	Semivariance	Counts
48.0	0.00726	1 545	55.3	0.00818	53
92.5	0.00971	2 793	100.1	0.01245	90
131.9	0.01128	3 217	135.7	0.01439	124
169.3	0.01280	3 592	167.7	0.01715	137
212.6	0.01488	5 456	210.3	0.01808	223
253.0	0.01661	4 562	252.9	0.01617	190
293.0	0.01777	5 524	292.2	0.01647	230
334.8	0.01905	5 792	337.2	0.02171	243
374.8	0.01936	5 226	374.6	0.02059	193
414.3	0.01996	5 699	414.6	0.01616	236
453.0	0.01960	4 918	452.0	0.01949	174
492.2	0.02016	5 447	490.8	0.02363	236
<b>534.5</b>	<b>0.01930</b>	<b>5 484</b>	<b>534.3</b>	<b>0.01786</b>	<b>226</b>
575.6	0.01881	4 513	574.7	0.01706	187
614.5	0.01877	4 189	615.1	0.02181	167
653.5	0.01806	3 524	653.9	0.02363	139
693.2	0.01866	3 572	692.4	0.02228	142
734.9	0.01792	3 136	733.5	0.01774	123
775.1	0.01766	2 689	773.5	0.01832	119

numbers might still be much larger than for the first few lags (Table 3.1). The larger numbers do not imply greater reliability, however, because individual data are used repeatedly, and the estimated semivariances are more or less correlated with one another. As a result, you should not rely on the number of comparisons as a guide to the reliability of your variogram when you have too few data to ensure accuracy.

We illustrate the effect of sample size with data on exchangeable potassium in the topsoil (0–23 cm) from a survey at Broom's Barn Farm (an experimental farm of 80 ha in Suffolk, England), which was first analysed by Webster and McBratney (1987). Table 3.2 summarizes the statistics. There were 434 sampling sites at an interval of 40 m on a square grid. The data were transformed to common logarithms ( $\log_{10}$ ) because the skewness coefficient is 2.04 (see Table 3.2 and an explanation in

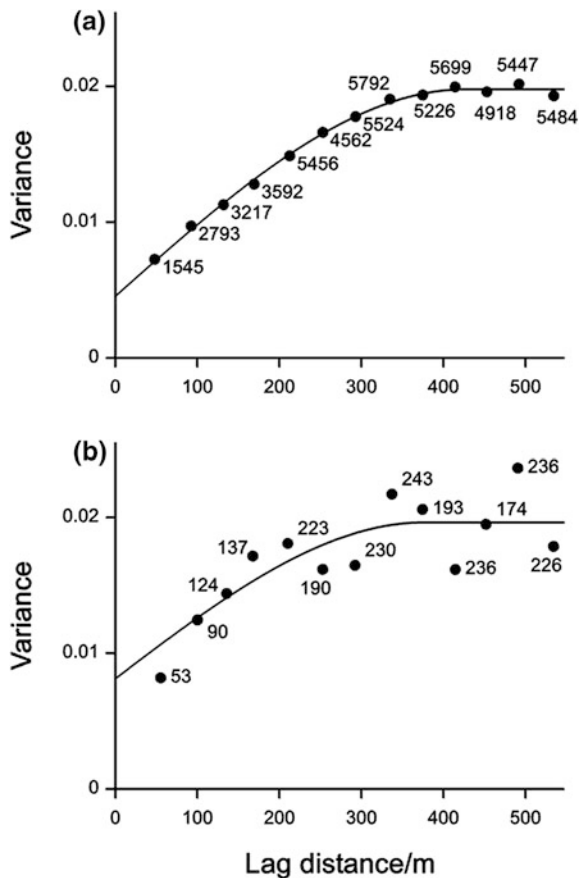
**Table 3.2** Summary statistics of potassium at Broom's Barn Farm

	$K^+/\text{mg l}^{-1}$	$\log_{10}(K^+)$	$K^+/\text{mg l}^{-1}$	$\log_{10}(K^+)$
Number of data	434	434	87	87
Minimum	12.0	1.0792	14.0	1.146
Maximum	96.0	1.9823	70.0	1.845
Mean	26.31	1.3985	26.7	1.404
Median	25.0	1.3979	26.0	1.415
Standard deviation	9.039	0.1342	9.403	0.138
Variance	81.71	0.0180	88.42	0.019
Skewness	2.04	0.39	1.760	0.395

Sect. 3.2.4). Figure 3.4a shows the experimental variogram of the full set of data (symbols); the estimates lie on a smooth curve. The exchangeable data were sub-sampled to 87 sites, and Fig. 3.4b shows a much more erratic experimental variogram. The result is therefore likely to be less reliable, and it is not clear what kind of curve would fit it best.

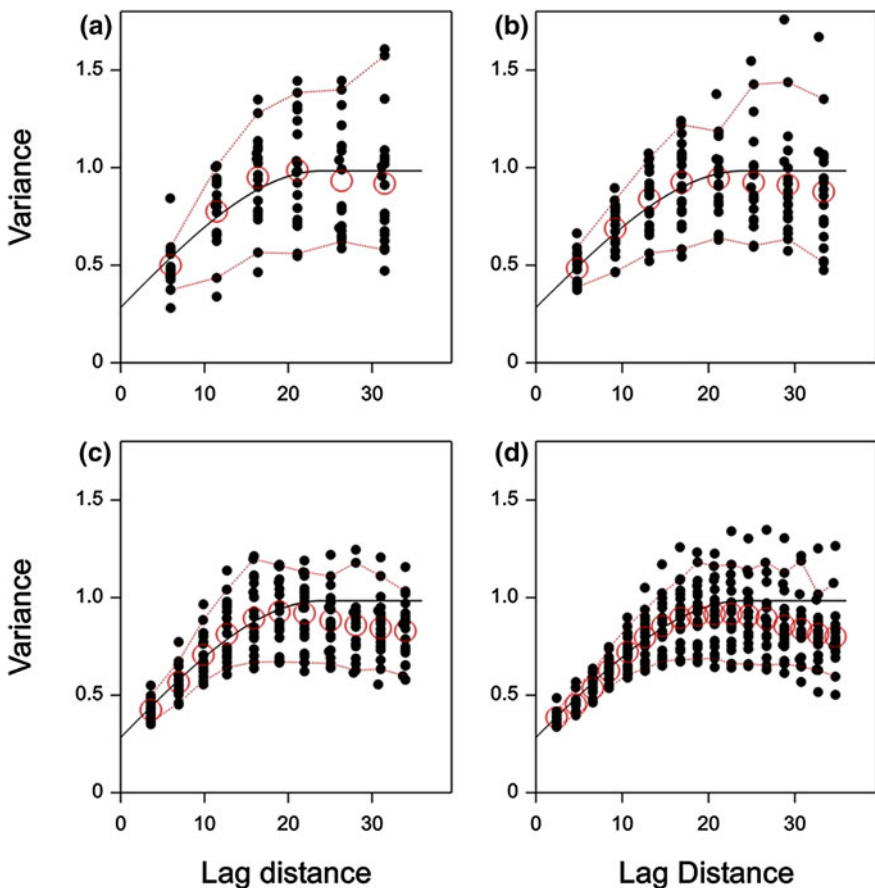
Figure 3.4 gives the number of comparisons for each computed semivariance from both the full set of 434 sites and from the sub-sample of 87 sites. The semivariences computed on 434 data, Fig. 3.4a, have more than 1 000 comparisons at the first lag, and they increase to more than 5 000 at the longest lags. The variogram computed from 87 data, Fig. 3.4b, computed with the same step and bin width as in Fig. 3.4a, has many fewer comparisons at all lags. Nevertheless, the number of comparisons at some of the longer lags exceeds 200. Many authors have been misled into thinking that they can obtain reliable estimates of  $\gamma(\mathbf{h})$  based on 50 comparisons, or even fewer; they cannot, as is clear from this result with 87 data—variograms computed on small sets of data are unreliable (see Webster and Oliver 1992).

**Fig. 3.4** Experimental variograms of the common logarithm of exchangeable potassium,  $\log_{10} K^+$ , in the topsoil of Broom’s Barn Farm, Suffolk; **a** computed from data at 434 sampling quadrats and **b** computed from all 87 quadrats. The *numbers* attached to the *points* are the numbers of paired comparisons from which the semivariences are computed. The *lines* are the best fitting spherical models



Over the years we have seen many erratic variograms computed on too few data, in some cases as few as 25. Twenty years ago we explored the sampling fluctuation in variograms (Webster and Oliver 1992). We concluded that one should aim for 150 data where variation is isotropic and set 100 as a minimum. Brus and de Gruijter (1994) came to a similar conclusion via a different route, and the message is reinforced with examples in Webster and Lark (2013) and in Oliver and Webster (2014).

For this chapter we have revisited the matter by repeated independent sampling from a much larger correlated random field of  $400 \times 400 = 160\,000$  with an isotropic spherical variogram:  $0.283 + 0.700 \times \text{sph}(h|24)$  and variance of 1.0. See Eq. (3.10) for a full definition of the function. Figure 3.5 shows the results of 15



**Fig. 3.5** Experimental variograms computed from repeated sampling on grids of  $7 \times 7$ ,  $9 \times 9$ ,  $12 \times 12$  and  $18 \times 18$  points. The *solid lines* are those of the isotropic spherical model fitted to the exhaustive experimental variogram, and the *dashed lines* join the 5 and 95 % quantiles, and the *circles* are the mean values at the lags. The model is  $\gamma(h) = 0.283 + 0.700 \times \text{sph}(h|24)$  for a field with variance 1.0

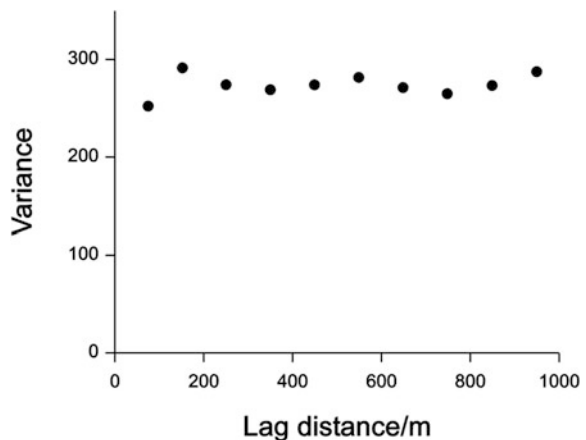


independent repeated samplings for four grids of sizes  $7 \times 7 = 49$ ,  $9 \times 9 = 81$ ,  $12 \times 12 = 144$  and  $18 \times 18 = 324$  sampling points. Evidently with 144 points the estimates of  $\gamma(h)$  at the shorter lag distances lie close to the variogram used to generate the field, the solid curve in the figure, but diverge at the longer lags. The same is true for the much larger samples of 324 points. Sample sizes in the range 100–150 should be adequate where the variogram is required for kriging, but estimates of sill variances will be erratic. Where variation is anisotropic, i.e. not the same in all directions, more data are required to identify it and define it mathematically. We deal with anisotropy below (Sect. 3.2.5).

### 3.2.2 Sampling Interval and Spatial Scale

The choice of a suitable sampling interval depends on the scale of variation that the practitioner wishes to resolve, e.g. experimental plot, field, farm, catchment, administrative region and so on. If you have rough variograms of the properties of interest or variograms from related ancillary data such as aerial images then choose a sampling interval that will give you at least five estimates of  $\gamma(h)$  within the effective range. Alternatively, you can use an accurate existing variogram of a property of interest to determine the kriging errors and so determine an optimal sampling interval for kriging, see Chap. 5 (Burgess et al. 1981; Webster and Lark 2013). If the lag interval exceeds half the range or effective range of variation the resulting variogram is likely to be flat; it will not capture the correlated structure and so will not describe adequately the spatial variation present, as in Fig. 3.6. The experimental variogram of topsoil sand in this figure was computed from a stratified random sample of the soil of the Wyre Forest, England (Oliver and Webster 1987). The average distance between neighbouring sampling points was 165 m, and the experimental variogram was computed with a lag interval of 75 m. The resulting

**Fig. 3.6** Experimental variogram of topsoil sand from a stratified random survey in the Wyre Forest, England. The variogram is pure nugget

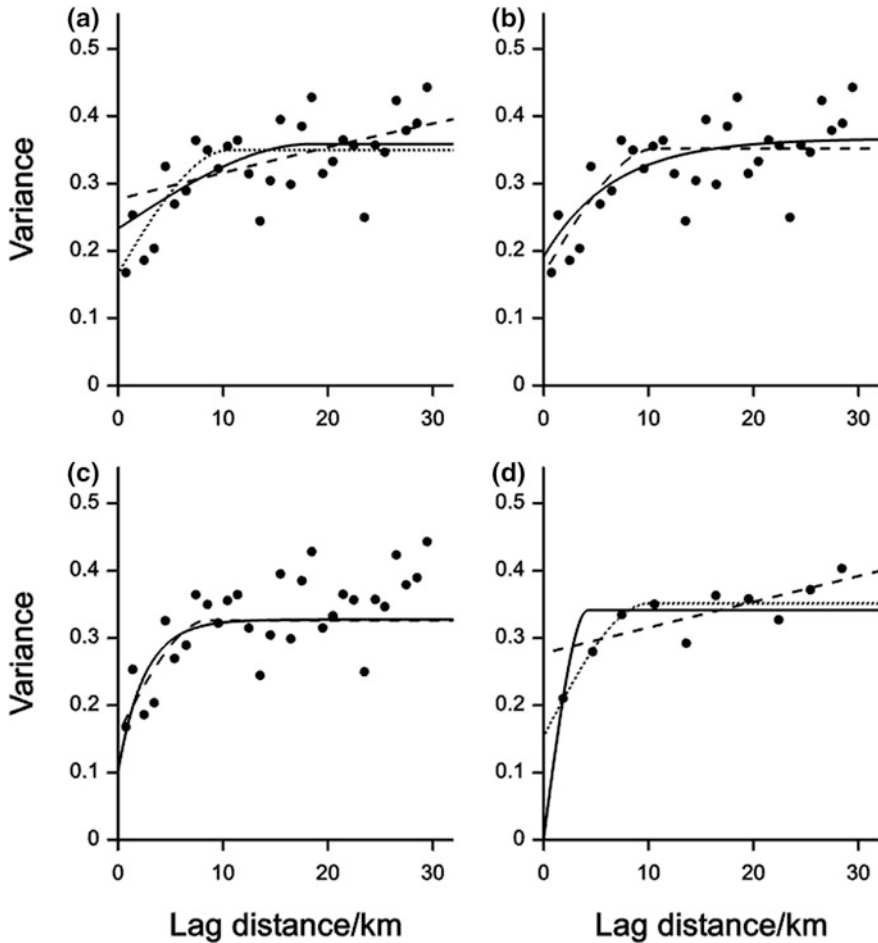


variogram appears as ‘pure nugget’—it shows no spatial structure. Further surveys revealed that the range of spatial dependence of topsoil sand here was approximately 70 m. In other words, all of the variation occurs over distances less than 70 m, which is much less than the average sample spacing in the first survey.

### 3.2.3 Lag Interval and Bin Width

As mentioned above, where data are on a regular grid or at equal intervals on transects the natural step is one interval. Where they are irregularly scattered, the comparisons must be grouped by distance as described in Fig. 3.3. The practitioner must choose both the length of the step,  $h$ , and the limits,  $w$ , 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 the width of bins requires judgement. If the steps are short and the bins narrow then there will be many estimates of  $\gamma(\mathbf{h})$ , which can lead to a ‘noisy’ variogram because the semi-variances are calculated from few comparisons. If in contrast the steps are large and the bins wide then there might be too few estimates of the semivariances to reveal the form of the variogram. The choice is thus a compromise; it is not one that should be automated. The practitioner should graph the experimental values, as in Fig. 3.7, so that the selection can be made objectively.

We illustrate the effect of lag interval and bin width with irregularly scattered data on cadmium concentrations in the soil of a region to the south east of the Madrid metropolitan area, Spain (Vázquez de la Cueva et al. 2014). The region is 35 km from west to east by 30 km from north to south. The topsoil (0–15 cm) was sampled at 125 sites. The design comprised two superimposed grids, one at 5-km intervals and the other at 1-km intervals. From the possible 1 116 nodes 74 were chosen at random, and 51 points were added 200 m from the 74. At each site five cores of soil were taken from a circle of radius 5 m and bulked for laboratory analysis. Table 3.3 summarizes the statistics. The coefficient of skewness of 1.71 indicates a long upper tail in the distribution (see Sect. 3.4) that might be reduced by transformation. After transformation to natural logarithms the skewness is somewhat reduced to  $-1.11$ , but remains outside the generally advised limits (Sect. 3.4). The experimental variograms in Fig. 3.7 were computed from the natural logarithms of cadmium concentrations: those in Fig. 3.7a–c were computed with a lag interval of 1 km and that in Fig. 3.7d with an interval of 3 km. The variogram computed at 1-km intervals is very erratic because of the small number of comparisons in each estimate. There is no clear indication from the experimental values of the kind of model that will fit best. The sequence of points in Fig. 3.7d computed with a lag interval of 3 km is now smoother and has a clearer structure. This example shows that the lag interval and bin width give different pictures of the spatial correlation: contrast Fig. 3.7a–c with d.



**Fig. 3.7** Experimental variograms of cadmium in the topsoil of a region south east of Madrid computed from 125 sampling points, **a–c** at 1-km intervals with bins 1 km wide, and **d** at 3-km intervals with bins 3 km wide. Models have been fitted by GenStat as follows: **a** spherical model to 30 km with range set initially to 25 km (*dashed*), iterated once (*dotted*) and iterated twice (*solid*); **b** spherical model to 30 km with range set initially to 10 km (*dashed*) and exponential model fitted to 30 km with distance parameter set initially to 3 km (*solid*); **c** spherical (*dashed*) and exponential (*solid*) models fitted to 15 km with same initial values for distance parameters as in (**b**); **d** spherical models with range initially set to 3 km (*dashed*) and iterated once (*solid*) and with range set initially to 10 km (*dotted*)

### 3.2.4 Statistical Distribution

Geostatistical analysis does not require data to follow a normal distribution. However, variograms comprise sequences of variances, and these can be unstable where data are strongly skewed and contain outliers. If your data do not have a

**Table 3.3** Summary statistics of cadmium in soil of Madrid region

		Cd/mg kg <sup>-1</sup>	ln(Cd)
Number of data	125		
Minimum		0.005	-5.298
Maximum		0.48	-0.734
Mean		0.137	-2.138
Median		0.11	-2.207
Standard deviation		0.0802	0.589
Variance		0.00643	0.347
Skewness		1.71	-1.107

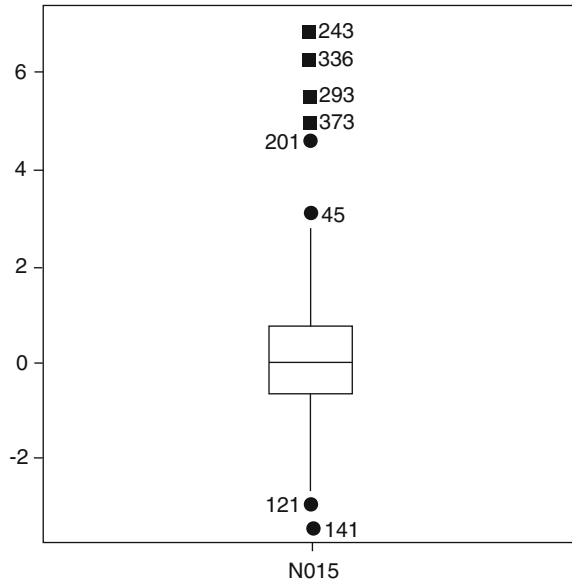
near-normal distribution and have a skewness coefficient outside the limits  $\pm 1$ , because of a long tail, you should consider transforming them. So, transform the data in some appropriate way, say by taking logarithms, and examine variograms computed on both raw and transformed values. Do the resulting variograms differ substantially apart from a scaling factor? In some cases the answers will be 'no'; in others 'yes'.

Kerry and Oliver (2007a) explored the effects of varying skewness and sample size on simulated random fields with asymmetry. Their results showed that for a large sample size of 1 600 data (on a 5-m grid), the change in shape of the variogram with increasing asymmetry was small, even for a skewness coefficient of 5. For a sample size of 400 (on a 10-m grid), the change in shape of the variograms was not large with increasing skewness and transformation. With 100 data (20-m grid), the semivariances at the first two lags proved to be similar to the generating function of the simulated field, but beyond that they departed progressively as the skewness increased, and for the skewness coefficient of 5 the variogram appeared as pure nugget. Our advice is to transform if it makes a difference to the variogram, but otherwise work with the original data (Table 3.2).

The variogram is sensitive to outliers in the data, i.e. *unexpectedly* large or small values beyond the limits of the main distribution. Box-plots, Fig. 3.8, are an ideal way to identify outliers. All outliers should be investigated and considered as potentially erroneous values before they are allowed to remain as part of the data set. For contaminated sites, however, the largest values will be of most interest. We mentioned above that the same data can contribute to several estimates of  $\gamma(\mathbf{h})$ , and so outliers inflate the averages. If there are few outliers relative to the whole data, removing them often reduces skewness, and this is a reasonable approach. The values removed can be returned to the data for kriging if desired. Transformation often fails to improve the distribution when outliers are present and can even make matters worse. The alternative is to use one of the robust estimators, such as those of Cressie and Hawkins (1980), Dowd (1984) and Genton (1998).

Cressie and Hawkins's (1980) estimator,  $\hat{\gamma}_{\text{CH}}(\mathbf{h})$ , is based on taking the fourth root of the squared differences and dampens the effect of outliers from the secondary process. It is given by

**Fig. 3.8** Box-plot computed from a field of 400 values simulated with a spherical variogram function with zero nugget and contaminated with five outliers resulting in a skewness coefficient of 1.5, where *filled squares* represent the far outliers which are three times beyond the interquartile range and *filled circles* are near outliers



$$2\hat{\gamma}_{CH}(\mathbf{h}) = \frac{\left\{ \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} |z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})|^2 \right\}^4}{0.457 + \frac{0.494}{m(\mathbf{h})} + \frac{0.045}{m^2(\mathbf{h})}} \quad (3.3)$$

The denominator in Eq. (3.3) is a correction based on the assumption that the underlying process to be estimated has normally distributed differences over all lags.

Dowd’s (1984) estimator,  $\hat{\gamma}_D(\mathbf{h})$ , and Genton’s,  $\hat{\gamma}_G(\mathbf{h})$ , estimate the variogram for a dominant intrinsic process in the presence of outliers. Dowd’s estimator is given as

$$2\hat{\gamma}_D(\mathbf{h}) = 2.198 \{ \text{median}(|y_i(\mathbf{h})|) \}^2, \quad (3.4)$$

where  $y_i(\mathbf{h}) = z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})$ ,  $i = 1, 2, \dots, m(\mathbf{h})$ . The term within the braces of Eq. (3.4) is the median absolute pair difference (MAPD) for lag  $\mathbf{h}$ , which is a scale estimator only for variables where the expectation of the differences is zero. The constant is a correction that scales the MAPD to the standard deviation of a normally distributed population.

Genton’s (1998) estimator,  $\hat{\gamma}_G(\mathbf{h})$ , is based on the scale estimator,  $Q_{Nh}$ , of Rousseeuw and Croux (1992). The estimator,  $Q_{Nh}$ , is given by

$$Q_{Nh} = 2.219 \left\{ |X_i - X_j|; i < j \right\}_{\left(\frac{H}{2}\right)}, \quad (3.5)$$

where the constant 2.219 is a correction for consistency with the standard deviation of the normal distribution, and  $H$  is the integral part of  $(N/2) + 1$ . Genton’s (1998)

estimator uses Eq. (3.5) as an estimator of scale applied to the differences at each lag; it is given by

$$2\hat{\gamma}_G(\mathbf{h}) = \left[ 2.219 \left\{ |y_i(\mathbf{h}) - y_j(\mathbf{h})|; i < j \right\}_{\left(\frac{H}{2}\right)} \right]^2, \quad (3.6)$$

but with  $H$  being the integral part of  $\{m(\mathbf{h}/2)\} + 1$ .

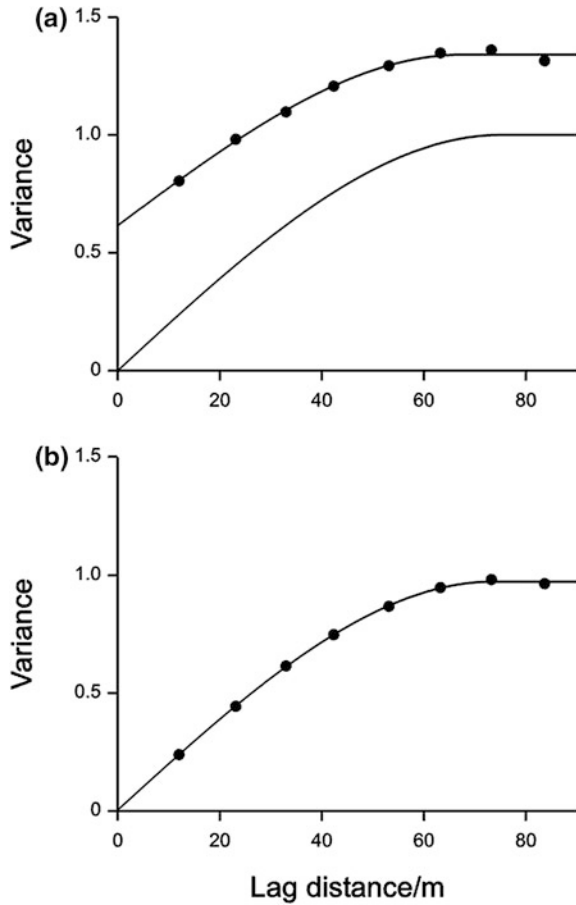
Kerry and Oliver (2007b) examined the effects of outliers and sample size in detail with fields of simulated data. They concluded that skewness caused by outliers must be dealt with regardless of the number of data. Furthermore, their results indicated that practitioners should act when skewness exceeds 0.5 rather than the limits mentioned above which are those generally used. Although the robust estimators provided a reasonable solution, they did not perform equally well in all the situations Kerry and Oliver examined. They therefore recommended the removal of outliers before computing the variogram as the current ‘best practice’ where outliers are randomly located and will not be returned to the data for kriging. Where outliers are crucial to the investigation, as on contaminated sites, practitioners should compute several robust variograms and compare them by cross-validation.

A field of 400 values was simulated on a 10-m grid by a spherical function with zero nugget, a sill of 1 and range of 75 m (Kerry and Oliver 2007b), i.e.  $c_0 = 0$ ,  $c = 1$  and  $r = 75$  m, see Eq. (3.10). Five of the values were contaminated by another process to give a skewness coefficient of 1.5. Figure 3.8 shows the box-plot of these values; the outliers are  $>4$ . An experimental variogram was computed from all the values and modelled, Fig. 3.9a. The nugget variance has increased dramatically to 0.617 showing the effect of adjacent disparate values. The sill variance is 1.341, which is an expression of the increase in variance, and the range has decreased to 67.5 m. The dashed line in Fig. 3.9a is the generating function of the simulated field. Figure 3.9b shows the experimental variogram and model for the same values, but with the outliers removed. The nugget variance is zero, the sill variance is almost 1.0 and the range is 73.6 m. This result shows how important it is to deal with outliers in data.

### 3.2.5 Anisotropy

Variation can vary from one direction to another, i.e. it can be anisotropic. You should therefore check your data for fluctuations in directional variation. In many instances the anisotropy is such that it could be made isotropic by a simple linear transformation of the spatial coordinates. Imagine that the region sampled is placed on a rubber sheet, which could be stretched in the direction in which variation seemed shortest. If the stretching eventually produces variation that is the same in that direction as that in the perpendicular direction then the anisotropy is known as *geometric*. The equation for the transformation is

**Fig. 3.9** Experimental variograms (*symbols*) and fitted models (*solid lines*) computed from a field of 400 values simulated with a spherical variogram function with zero nugget (*dashed line*): **a** contaminated with five outliers resulting in a skewness coefficient of 1.5 and **b** with the outliers removed



$$\Omega(\vartheta) = \{A^2 \cos^2(\vartheta - \varphi) + B^2 \sin^2(\vartheta - \varphi)\}^{1/2}, \quad (3.7)$$

where  $\Omega$  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 direction of least continuity or greatest variation, the minimum. For an unbounded variogram, the roles of  $A$  and  $B$  are reversed, and  $A$  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. Figure 3.12 shows an example in which there are differences in the ranges for a bounded variogram (see Sect. 5.1).

Anisotropy can also occur as preferentially orientated zones with different means that result in changes in variance with change in direction and fluctuations in the sill. This is known as *zonal anisotropy*.

### 3.2.6 Trend

In Chap. 1 we mentioned trend. We consider it briefly here in relation to the variogram, but Chap. 6 is devoted to the matter, and readers should turn to that chapter for detail. We can always calculate an experimental variogram by Eq. (3.1), but it estimates the theoretical variogram  $\gamma(\mathbf{h})$  only where the underlying process is random. If there is trend then this equation gives a false summary of the random part of the process. Typically, where trend is present the experimental variogram increases without bound, and if it dominates then the experimental sequence becomes increasingly steep as the lag distance increases (see Fig. 6.11). If you obtain such a result then examine your data by fitting simple linear and quadratic polynomials on the coordinates. Alternatively, map the data by some simple graphical procedure before doing a statistical analysis; if the map shows gradual continuous change across the region then there is trend with more or less patchiness superimposed.

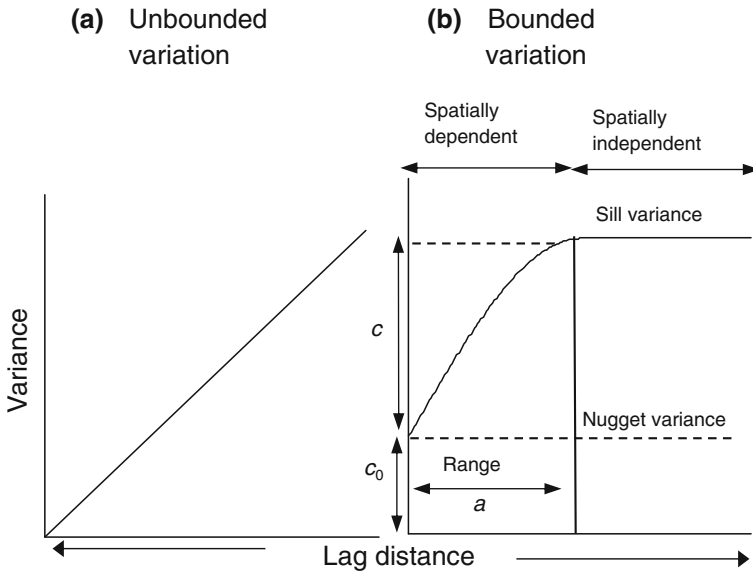
## 3.3 Modelling the Variogram

The experimental variogram consists of semivariances at a finite set of discrete lags. These semivariances are estimates based on samples; they are therefore subject to error, which itself varies from one estimate to the next. In addition, the underlying function is continuous for all  $\mathbf{h}$ , Eqs. (3.5) and (3.8). The next step in variography is to fit a smooth curve or surface to the experimental values, one that describes the principal features of the sequence (see Sect. 3.3.1) while ignoring the point-to-point erratic fluctuation. Not any plausible-looking curve or surface will serve; it must have a mathematical expression that can legitimately describe the variances of random processes. It must guarantee non-negative variances of combinations of values, and there are only a few simple functions that do so. They are known as *conditional negative semi-definite* (CNSD) because the matrices to which they contribute are themselves conditional negative semi-definite (see Webster and Oliver 2007, for a full account).

### 3.3.1 Principal Features of the Variogram

- (1) An increase in variance with increasing lag distance from the ordinate  
 In Fig. 3.10a the variogram shows a monotonic increase in variance as the lag distance increases. The slope shows the change in the *spatial autocorrelation* or *dependence* between sampling points as the separation distance increases. In other words at short lag intervals,  $|\mathbf{h}|$ , the semivariances,  $\gamma(|\mathbf{h}|)$ , are small indicating that values of  $Z(\mathbf{x})$  are similar, and as  $|\mathbf{h}|$  increases they become increasingly dissimilar on average.





**Fig. 3.10** Examples of: **a** unbounded and **b** bounded variogram models with annotations to illustrate the parameters of a bounded model function

- (2) An upper bound, the *sill variance*  
 If the process is second-order stationary then the variogram will reach an upper bound, the *sill variance*, after the initial increase as in Fig. 3.10b. For some variograms the sill remains constant, whereas for others it is an asymptote, which we explain below. The sill variance is also the *a priori* variance,  $\sigma^2$ , of the process.
- (3) The *range* of spatial correlation or dependence  
 A variogram that reaches its sill at a finite lag distance has a *range*, which is the limit of spatial correlation where the autocorrelation becomes 0, Fig. 3.10a. Places further apart than this are spatially uncorrelated or independent. Variograms that approach their sills asymptotically have no strict ranges; in practice, however, we use an effective range at the lag distances where they reach 0.95 of their sills.
- (4) *Unbounded* variogram  
 The variogram may increase indefinitely with increasing lag distance as in Fig. 3.10a. It describes a process that is not second-order stationary, and the covariance does not exist. The variogram, however, does exist and fulfils Matheron's (1965) intrinsic hypothesis (see Sect. 2.1, Chap. 2).
- (5) A positive intercept on the ordinate, the *nugget variance*  
 The variogram often approaches the ordinate with a positive intercept known as the *nugget variance*, Fig. 3.10b. Theoretically, when  $\mathbf{h} = \mathbf{0}$  the semivariance should also be 0 (see Chap. 2). The term 'nugget' in this context was

coined in gold mining because gold nuggets appear to occur at random and independently of one another. They represent a discontinuity in the variation, an uncorrelated component, because the gold content no longer relates to that at neighbouring sites. For properties that vary continuously in space, such as the amount of water vapour in the atmosphere or the pH of the soil, the nugget variance arises from measurement error (usually a small component) and variation over distances less than the shortest sampling interval.

(6) Directional variation *anisotropy*

Spatial variation might vary according to direction, as mentioned above, and we need to be able to take this into account in our analysis and modelling.

### 3.3.2 Variogram Model Functions

There are two principal kinds of function, namely bounded and unbounded (Fig. 3.10). We give the equations and illustrate the three most popular models; power function (unbounded), spherical (bounded) and exponential (asymptotically bounded). If none of these appears to fit the experimental values then more complex functions may be fitted. Such functions may be any combination of simple CNSD functions; these combinations are themselves CNSD.

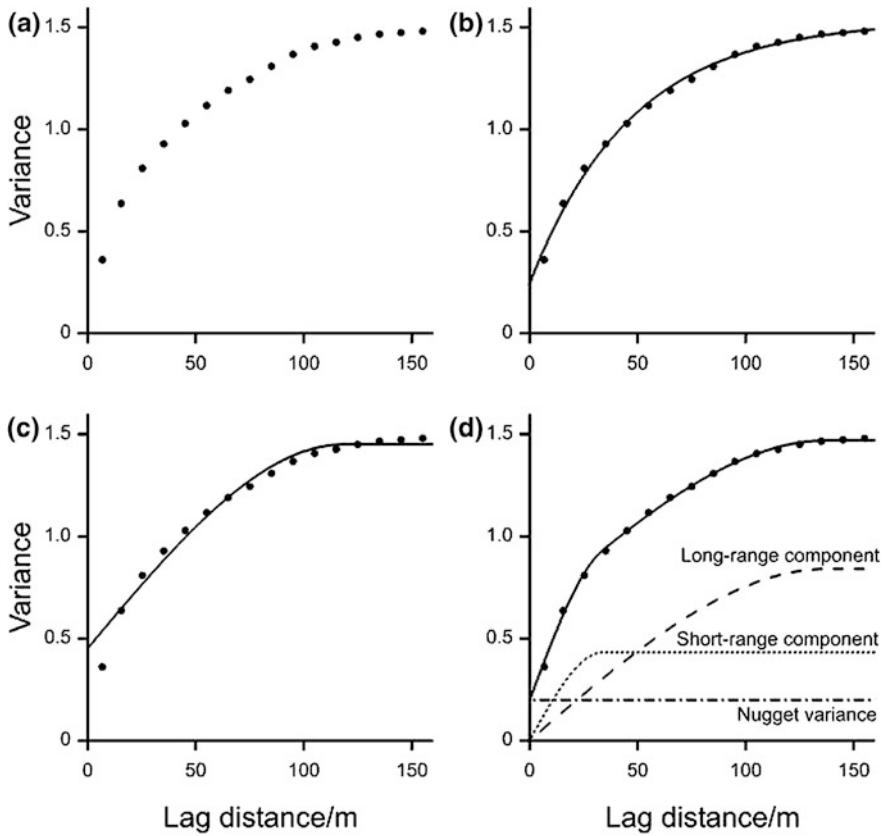
Theoretically, the variogram model should intercept the ordinate at the origin according to theory as in Eq. (3.8) and Fig. 3.10a. In practice the experimental variogram frequently, indeed usually, appears to approach the ordinate at some positive finite value. To make the curve fit one adds a nugget component to the simple function as in Eqs. (3.8)–(3.11) and Fig. 3.10b where there is a nugget variance and a structured component. A more complex function is required where there are two or more distinct scales of spatial dependence, i.e. a nested model. We illustrate this scenario with two spherical functions, one nested within the other, plus a nugget variance, Eq. (3.12) and Fig. 3.11. We describe the models in their isotropic form; they are symmetric about zero lag, but we define them for  $|\mathbf{h}| \geq 0$  only.

The equations for the four models are as follows.

*Power function.* This is an unbounded function

$$\gamma(h) = gh^\beta \quad \text{for } 0 < \beta < 2, \quad (3.8)$$

where  $g$  describes the intensity of the variation and  $\beta$  describes the curvature. If  $\beta = 1$ , the variogram is linear and  $g$  represents the gradient. The limits 0 and 2 are excluded because  $\beta = 0$  indicates constant variance for all  $h > 0$  and  $\beta = 2$  that the function is parabolic with zero gradient at the origin. The latter means that the process is not random. Figure 3.10a gives an example of an unbounded variogram with no nugget variance as in the equation above. If such a function had a positive intercept at the ordinate the equation would be



**Fig. 3.11** Wheat yield recorded in 1999 in Football Field on the Shuttleworth Estate **a** the experimental variogram (*symbols*), **b** the *solid line* is an exponential function fitted to the experimental values, **c** the *solid line* is a spherical function fitted to the experimental values and **d** the best fitting nested spherical model (*solid line*). The model was decomposed to illustrate the individual model components as shown by the *ornamented lines*

$$\gamma(h) = c_0 + gh^\beta \quad \text{for } 0 < \beta < 2, \tag{3.9}$$

*Spherical model.* This is

$$\gamma(h) = \begin{cases} c_0 + c \left\{ \frac{3h}{2r} - \frac{1}{2} \left( \frac{h}{r} \right)^3 \right\} & \text{for } 0 < h \leq r \\ c_0 + c & \text{for } h > r \\ 0 & \text{for } h = 0, \end{cases} \tag{3.10}$$

where  $c_0$  is the nugget variance,  $c$  is the variance of spatially correlated component and  $r$  is the range of spatial dependence. Figure 3.10b illustrates a spherical

variogram with annotations of the main features as described above. The quantity  $c_0 + c$  is known as the *sill variance*.

*Exponential model.* This is

$$\gamma(h) = \begin{cases} c_0 + c\{1 - \exp(-\frac{h}{a})\}, & \text{for } 0 < h \\ 0 & \text{for } h = 0 \end{cases} \quad (3.11)$$

where  $a$  is the distance parameter. This function approaches its sill asymptotically, and so it does not have a finite range. For practical purposes it is usual to assign an effective range,  $a'$ , which is approximately equal to  $3a$ . Figure 3.11b shows an example of a fitted exponential function.

*Nested spherical.* This is

$$\gamma(h) = \begin{cases} c_0 + c_1 \left\{ \frac{3h}{2r_1} - \frac{1}{2} \left( \frac{h}{r_1} \right)^3 \right\} + c_2 \left\{ \frac{3h}{2r_2} - \frac{1}{2} \left( \frac{h}{r_2} \right)^3 \right\} & \text{for } 0 < h \leq r_1 \\ c_0 + c_1 + c_2 \left\{ \frac{3h}{2r_2} - \frac{1}{2} \left( \frac{h}{r_2} \right)^3 \right\} & \text{for } r_1 < h \leq r_2 \\ c_0 + c_1 + c_2 & \text{for } h > r_2 \\ 0 & \text{for } h = 0, \end{cases} \quad (3.12)$$

where  $c_1$  and  $r_1$  are the sill and range of the short-range component of the variation, and  $c_2$  and  $r_2$  are the sill and range of the long-range component. A nugget component can also be added as above. The yield of wheat in Football Field, Shuttleworth Estate, Bedfordshire, was recorded in 1999, and an experimental variogram was computed from the values. Figure 3.11a shows the experimental values and Fig. 3.11b and c the fitted exponential and spherical functions, respectively. It is clear that the spherical function, Fig. 3.11c, fits poorly and that the exponential model, Fig. 3.11b, fits reasonably. The fit of the latter emphasizes the small change in slope evident in the experimental variogram at about lag 30 m and another change at around lag 140 m. Figure 3.11d shows the nested spherical function, which provides a near-perfect fit to the experimental values with a smaller nugget variance than the exponential model and follows the values closely. The variously ornamented lines in Fig. 3.11d show the components of the nested model; the nugget, short-range and long-range. Table 3.4 gives the parameters of these models; they show that the spherical function has a larger nugget variance than the other two models and a smaller range of spatial dependence. The parameters of the exponential model are closer to those of the nested spherical with a smaller nugget variance and an approximate effective range ( $3a$ ) of 140 m. The diagnostics in Table 3.4 reflect the visual observations. The residual sum of squares (RSS) is much larger for the spherical function than for the exponential and nested spherical models, and that for the exponential is larger than for the nested model.

If your models have the same number of parameters and the ones fitted seem to fit well then choose the one with the smallest residual sum of squares (RSS) or smallest mean square. You may wish to fit more complex models, but you should

**Table 3.4** Parameters of models fitted to yield from Football Field, Shuttleworth Estate, Bedfordshire, UK recorded in 1999

Model type	Estimates of parameters						Diagnostics			
	$c_0$	$c_1$	$c_2$	$a_1/m$	$a_2/m$	$r/m$	RSS	RMS	% variance	AIC
Exponential	0.2437	1.291				47.33	3 802	292.5	99.5	96.86
Spherical	0.4516	0.9994		118.4			14 896	1 146	97.8	118.7
Nested spherical	0.1975	0.4318	0.8415	33.88	137.8		1 052	95.63	99.8	82.97

be cautious because you can always diminish the RSS by increasing the number of parameters in the fitted model. For example, the double spherical model with nugget has five parameters, whereas the simpler single spherical model with nugget has only three. Are the two additional parameters justifiable? To ensure parsimony in our fitting we can compute an estimate of the Akaike Information Criterion (AIC) (see Webster and Oliver 2007, for more detail) if, as in our comparisons above, the models have unequal numbers of parameters as for the nested spherical model. The AIC is estimated by

$$AIC = \left\{ n \ln \left( \frac{2\pi}{n} \right) + n + 2 \right\} + n \ln R + 2p, \tag{3.13}$$

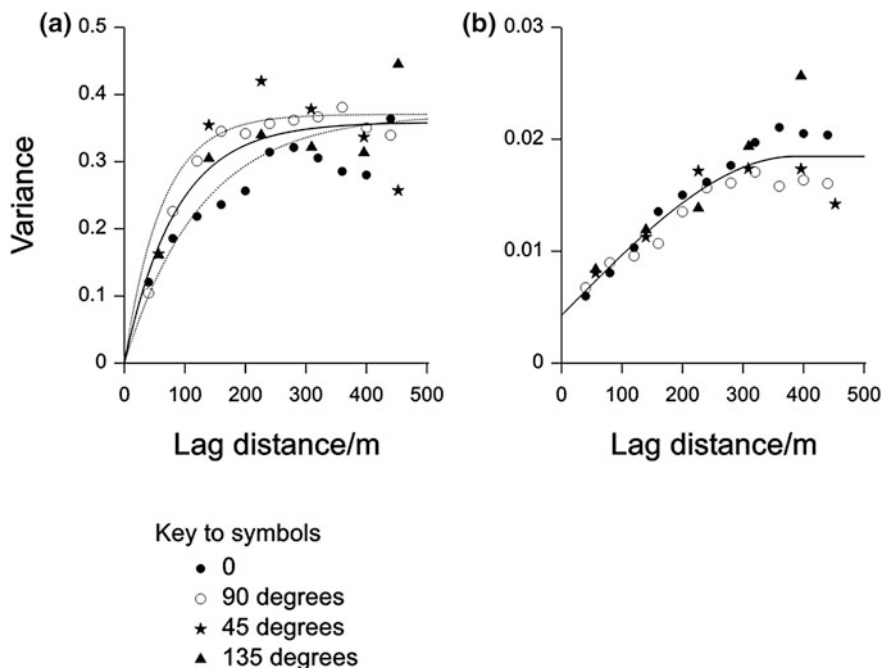
where  $n$  is the number of points on the variogram (16 in this example),  $p$  is the number of model parameters and  $R$  is the mean square of the residuals (RMS in Table 3.4). The quantity in braces is constant for any one experimental variogram, and so we need compute only

$$\hat{A} = n \ln R + 2p. \tag{3.14}$$

We then choose the model for which  $\hat{A}$  is the least. In Table 3.4,  $\hat{A}$  is markedly smaller for the nested spherical model than for the exponential and spherical functions, and so we would choose the more complex function as providing the best fit.

*Anisotropic model*

To examine data for both types of anisotropy compute the variogram in at least four directions to start with: along the rows, down the columns and on the principal diagonals if data are on a rectangular grid (see Fig. 3.12). The semivariances can be plotted in these directions, and no information is lost. For irregularly scattered data, we have to group the separations by direction as well as distance as in Fig. 3.3. The angle,  $\alpha$ , within which data are included in estimating the semivariance should allow complete cover to start with, i.e.  $\pi/4$  for four angles, which will include all data in those directions. Note, however, that this procedure loses some directional information. If it reveals directional variation then reduce  $\alpha$  to identify the direction of strongest anisotropy, but realize that the smaller  $\alpha$  becomes the fewer will be the



**Fig. 3.12** Experimental variograms computed in four directions: **a** pH; the *solid line* is the isotropic exponential model and the *dotted lines* form the envelope of the fitted anisotropic exponential model and **b**  $\log_{10} K^+$  with the fitted isotropic spherical function

number of comparisons and the greater will be the error in the estimated semi-variances. Choosing  $\alpha$  is therefore a compromise between a stable estimate based on many comparisons that will underestimate the directional effect with a wide angle and one that is subject to greater error but reflects the anisotropy more closely.

Figure 3.12a, b shows the experimental variograms of pH and exchangeable potassium (as  $\log_{10} K^+$ ), respectively, at Broom's Barn Farm computed in four directions. The directional variogram of pH shows a longer range of variation in the north–south ( $90^\circ$ ) direction and a shorter range in the east–west (0) direction, whereas for  $\log_{10} K^+$  no anisotropy is evident. The directional variogram for pH has been fitted with an anisotropic exponential function:

$$\gamma(h, \vartheta) = c_0 + c\{1 - \exp[-|\mathbf{h}|/\Omega(\vartheta)]\}, \quad (3.15)$$

where  $|\mathbf{h}|$  is the modulus of the lag and  $\Omega(\vartheta)$  is defined in Eq. (3.7). The model parameters are given in Table 3.4 and Fig. 3.12a shows the envelope of the model as the dotted lines. An isotropic exponential function was also fitted; the parameters of this are given in the table and the model is the solid line in Fig. 3.12a.

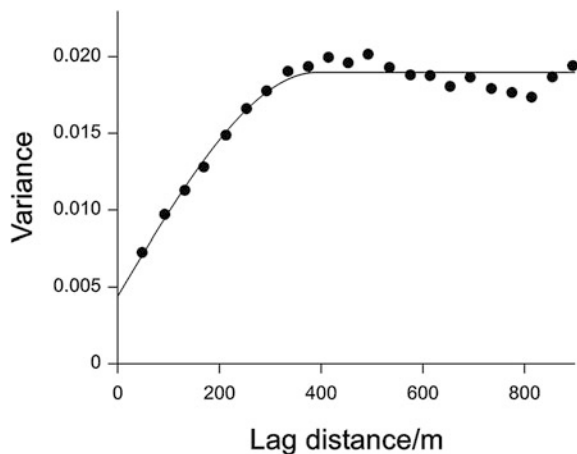
### 3.4 Factors Affecting the Reliability of Variogram Models

There are operational aspects that we need to consider when computing the experimental variogram and fitting models. They include the effects of poor choice of lag or bin interval and of maximum lag, and sample size on the reliability of the model parameters that will then be used for kriging. The experimental variogram should be computed and modelled only as far as it is reliably estimated. We recommend that you compute it to a maximum lag of no more than a third to one half of the extent of the data. Table 3.1 shows how the number of comparisons (counts) starts to decrease after a certain lag distance. It is at this lag distance (about 530 m) that the semi-variances also start to depart from the smooth curve; this is a sign that the estimates are becoming increasingly unreliable (Fig. 3.13). Table 3.5 shows how the model parameters of the fitted spherical models also change for  $\log_{10} K^+$  when the model was fitted to a maximum lag of 900 m compared with 550 m.

#### 3.4.1 Fitting Models

Fitting models remains controversial in geostatistics, yet it is one of the most important stages to get right. Some practitioners fit models by eye, which we do not recommend because the observed semivariances may fluctuate too much from point to point and their accuracy is not constant, which makes this approach unreliable. Fitting models with ‘black box’ software can also produce poor results because there is no choice, judgement or control over the process. We recommend a procedure that involves both visual inspection and statistical fitting in steps as follows.

**Fig. 3.13** Experimental variogram computed and modelled to a maximum lag of 900 m for  $\log_{10} K^+$  at Broom’s Barn Farm



**Table 3.5** Parameters of models fitted to exchangeable potassium in the topsoil at Broom's Barn Farm, England

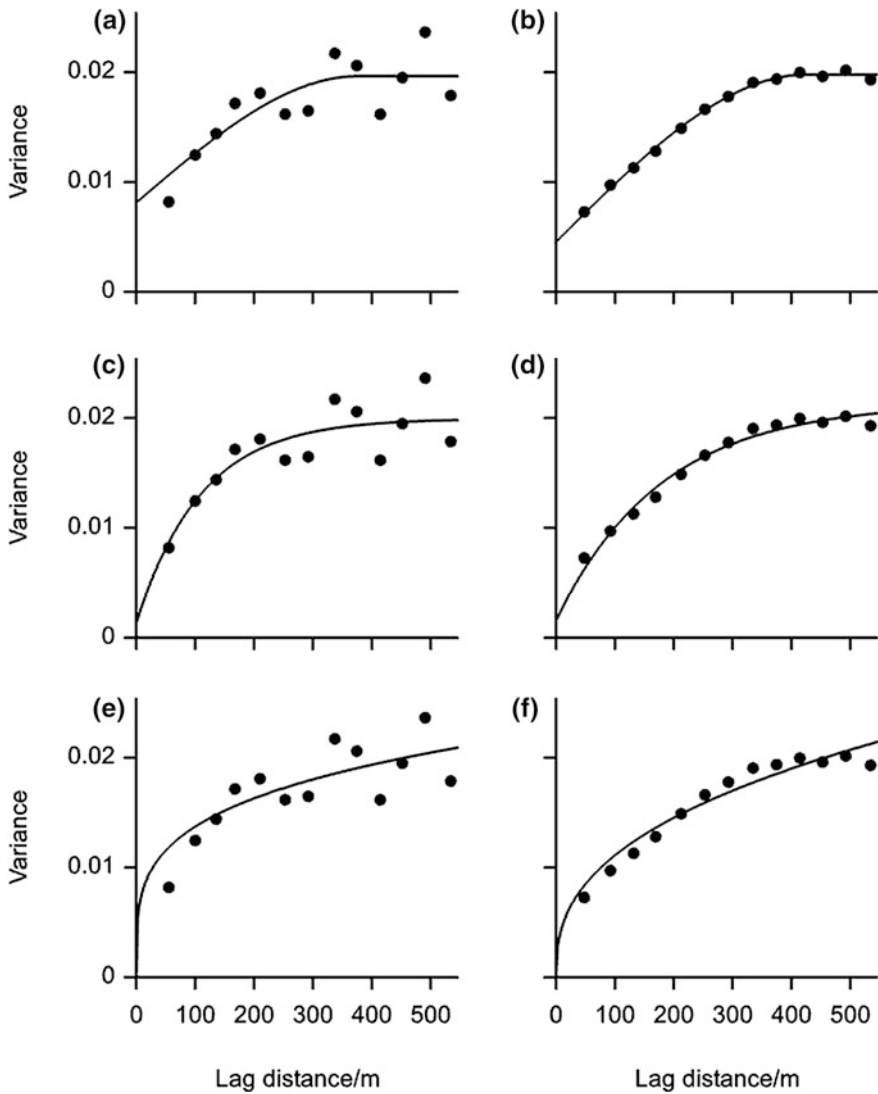
Model type	Estimates of parameters				Diagnostics			
	$c_0$	$c$	$a/m$	$r/m$	RSS	RMS	% variance	
Spherical— $\log_{10} K^{+87}$	0.008113	0.01151	376.4		0.01184	0.001184	40.9	
Spherical— $\log_{10} K^{+434}$	0.004535	0.01524	426.0		0.003683	0.000368	99.4	
Exponential— $\log_{10} K^{+87}$	0.001351	0.01860		110.4	0.01127	0.001127	43.8	
Exponential— $\log_{10} K^{+434}$	0.001578	0.01981		180.0	0.01943	0.001943	96.7	
	$w$	$\alpha$						
Power— $\log_{10} K^{+87}$	0.01092	0.2484			0.01254	0.001140	43.1	
Power— $\log_{10} K^{+443}$	0.00778	0.3876			0.05727	0.005206	91.2	
Spherical— $\log_{10} K^{+434}$ Maximum lag 900 m	0.004399	0.01458	393.2				96.4	
Exponential—pH	0	0.3588		85.2	13.67	0.5256	78.7	
Model type	Estimates of parameters							
Anisotropic exponential—pH	$c_0$	$c$	$\varphi$ (rad)		$A/m$	$B/m$	$A/B$	
	0	0.371	2.735 (156.7°)		126.9	61.96	2.048	
					RSS	RMS	% variance	
					9.255	0.3856	84.4	



1. First, plot the experimental variogram, the black discs in Fig. 3.14.
2. Choose several models with a similar shape and fit each in turn by weighted least squares, the curves in Fig. 3.14.
3. Plot the fitted models on the graph of the experimental variogram and assess whether the fit looks reasonable. If all plausible models seem to fit well, choose the one with the smallest residual sum of squares (RSS) or smallest mean square. If the models have unequal numbers of parameters as for the nested spherical model then compute the Akaike Information Criterion (AIC) and choose the model for which the AIC is least as above.

Figure 3.14a, c and e shows the experimental variogram computed from  $\log_{10} K^+$  with 87 data. None of the three models chosen, power, spherical and exponential Eqs. (3.9)–(3.11), respectively, and displayed above in Sect. 3.3.2, appears to fit well. Without the diagnostic information in Table 3.5 it would be difficult to choose between them. The exponential function has the smallest residual mean square (RMS) and accounts for the most variance, albeit only 44 %. The difference between the parameters of the two bounded functions, spherical and exponential, Eqs. (3.10) and (3.11), is marked, especially in relation to the nugget variance,  $c_0$ . The power function, Eq. (3.7), provides the next best fitting model, although it is clear from the variogram of the full set of data, Fig. 3.14b, that the underlying process is second-order stationary and requires a bounded function. For the same functions fitted to the experimental variogram of the full data, 434 sites, the best fitting function is clearly the spherical one which has a very small RMS and accounts for 99.4 % of the variance (Table 3.5). The exponential and power functions fit less well both visually and from the diagnostic values. The importance of an adequate sample size is clear from this example, which illustrates the poor fit of all functions to the experimental values from the sample of 87 and the small percentage variance accounted for compared with those for the full set.

Finally, we compare the effect of choice of lag interval and bin width for the data on the cadmium in the soil near Madrid, again with data from Vázquez de la Cueva et al. (2014). Figure 3.7 shows experimental variograms computed with a lag interval of 1 km in Fig. 3.7a–c and of 3 km in Fig. 3.7d. The variogram computed with a lag of 1 km is so erratic that none of the functions provides a good fit. Several of the exponential and spherical models fitted appear to be as good as any other, whereas for Fig. 3.7d it is clear that the model represented by the dotted line (spherical with range of 10 km and no iteration) provides the best fit. Table 3.6 lists the parameters of the functions fitted to the two experimental variograms. Different initial values for the non-linear parameter,  $r$ , for the spherical model were used and also different numbers of iterations which give increasing weight to values near to the origin. Because the experimental semivariograms are based on different numbers of paired comparisons,  $m(\mathbf{h})$  in Eq. (3.1), and because confidence in the estimate of variance decreases as its value increases, we generally weight the semivariograms by the number of counts when fitting the models. The inverse relation between the reliability of an estimate of variance and the variance itself led Cressie (1985) to propose a more elaborate weight, which has the form



**Fig. 3.14** Experimental variograms computed from 87 data for  $\log_{10} K^+$  Broom's Barn Farm, Suffolk and fitted with: **a** spherical model, **c** exponential model and **e** power function, and experimental variograms computed from 434 data and fitted with: **b** spherical model, **d** exponential model and **f** power function

$$m(\mathbf{h}_j) / \gamma^{*2}(\mathbf{h}_j), \quad (3.16)$$

where  $\gamma^{*2}(\mathbf{h}_j)$  is the value of semivariance predicted by the model. The quantity  $\gamma^{*2}(\mathbf{h}_j)$  is inserted into the weighting vector and the fitting is repeated, and the

**Table 3.6** Models fitted to experimental variograms of cadmium in soil of Madrid region

Fitting	Estimates of parameters				Diagnostics		
	$c_0$	$c$	$a/\text{km}$	$r/\text{km}$	MSE	MSDR	Correlation
<i>Lag interval 1 km</i>							
Spherical to 30 km, initial 25 km	0.2766	0.2011	76.0		0.3422	1.112	0.144
After one iteration	0.2329	0.1259	18.5		0.3480	1.201	0.145
After two iterations	0.1665	0.1833	10.6		0.3581	1.341	0.167
Spherical to 30 km, initial 10 km	0.1654	0.1868	10.75		0.3594	1.350	0.164
Exponential to 30 km, initial 3 km	0.1907	0.1762		6.63	0.3424	1.219	0.1898
Spherical to 15 km, initial 10 km	0.1575	0.1685	8.57		0.3392	1.284	0.228
Exponential to 15 km, initial 3 km	0.1012	0.2261		2.54	0.3371	1.292	0.241
<i>Lag interval 3 km</i>							
Spherical to 30 km, initial 3 km	0.2754	0.2062	76.8		0.3422	1.115	0.144
After iteration	0	0.3415	4.32		0.3602	2.125	0.258
Spherical to 30 km	0.1536	0.1978	10.08		0.3574	1.364	0.181

$c_0$  is the nugget variance,  $c$  is the sill variance of the correlated structure and  $a$  is the range of the spherical model and  $r$  is the distance parameter of the exponential model  
MSE is the mean squared error and MSDR is the mean squared deviation ratio

whole process is iterated to convergence, i.e. until there is no perceptible change in  $\hat{\gamma}^{*2}(\mathbf{h}_j)$ . However, McBratney and Webster (1986) discovered that one iteration was usually sufficient, and so in GenStat, for example, only one repeat is programmed.

The second iteration is a refinement of the former proposed by McBratney and Webster (1986):

$$m(\mathbf{h}_j)\hat{\gamma}(\mathbf{h}_j)/\hat{\gamma}^{*3}(\mathbf{h}_j), \tag{3.17}$$

where  $\hat{\gamma}(\mathbf{h}_j)$  is the observed value of the semivariance at  $\mathbf{h}_j$ . Both iterations give more weight to estimates close to the origin, which is usually desirable for kriging.

The initial value of the non-linear parameter,  $r$ , for the spherical model, can seriously affect the final model fitted: contrast the three curves in Fig. 3.14d. The weights given to the semivariances,  $\hat{\gamma}(h)$ , can also seriously affect the final model if the distance parameter is chosen poorly initially: see Fig. 3.7a and d and the parameters in Table 3.6.

Another fairly popular way of choosing models for variograms is by cross-validation. This procedure involves leaving out each and every value in the data in turn and kriging the value there using the surrounding data and the given model parameters. The kriged values  $\hat{Z}(\mathbf{x}_i)$  are compared with the observed ones  $z(\mathbf{x}_i)$ . The mean squared error (MSE) between the predictions and the observed values, the mean squared deviation ratio (MSDR) and the median of the squared deviation ratio are calculated and used as criteria of the goodness of the models. The precise nature of these quantities will be apparent when we have described kriging and so they are defined at the end of the next chapter, Chap. 4.