19 Spatially continuous data analysis and modelling

Saveliev, A.A., Mukharamova, S.S., Chizhikova, N.A., Budgey, R. and Zuur, A.F.

19.1 Spatially continuous data

In the previous chapter, we explored techniques to analyse data collected on a lattice. In this chapter, we will consider techniques to model continuous spatial data. The term *continuous* does not mean that the variable of interest is continuous, but merely that the variable can be measured in any location in the study area. Such continuously distributed variables are widely used in ecology and geoscience. Examples are relief elevation and bathymetry, temperature, moisture, soil nutrients, and subsurface geology. Spatially continuous data are often referred to as *geostatistical data* (Bailey and Gatrell 1995). The set of statistical techniques that can be used for analysing and modelling this type of data is called *geostatistics.*

Mathematical tools used to analyse and model geostatistical data are similar to those applied in time series analysis. Methods from both fields consider discrete observations, e.g., time points or spatial locations, and use correlation structure to describe the data dependence with tools like the auto-correlation function for time series and the variogram for geostatistical data. However, there are two important differences. Data in time series have only one direction, which is from past to more recent. However, there are multiple directions in space. Additionally, the main aim in time series analysis is to determine what has happened and occasionally to predict future values. In general, there is no need for data interpolation at intermediate time points. In geostatistics, the prediction of values at new locations is one of the prime aims of the analysis.

Geostatistical analysis usually consists of the same steps of the traditional statistical analysis workflow, and the following steps will be discussed in this chapter:

- Exploratory data analysis.
- Analysis of spatial correlation. This step is usually called the structural analysis or variogram modelling.

- Prediction of the data values at unsampled locations and estimation of the prediction uncertainty using *kriging*. Note that the goal is now shifting from parameter estimation to prediction of data.

To illustrate some of the geostatistical methods in subsequent sections of this chapter, we will refer to a simple relief elevation data from an area in the Volga River Valley in Russia. The data are part of the SRTM data set; see Rabus et al. (2003). We randomly selected a subset of 172 locations from our main data set for this area to illustrate analysis methods. The subset of points represents about 20% of the total relief data (26 × 33 grid) for this river valley. The location of samples within this relief surface are shown in Figure 19.1. A more challenging data set (bird radar data; see also Chapter 10) is analysed at the end of this chapter, and Chapter 37 shows a case study.

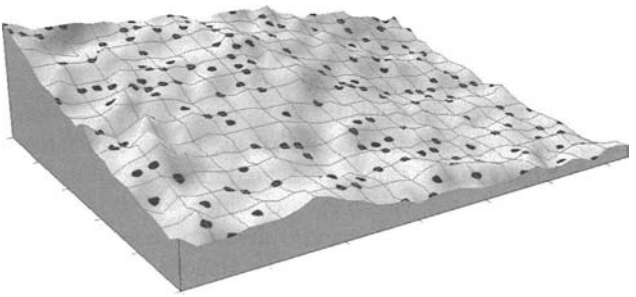

Figure 19.1. Relief surface and sampled locations (dots). The area is 2600 by 3300 meters.

19.2 Geostatistical functions and assumptions

In geostatistics, the term *regional variable* refers to some phenomenon that takes place in geographical space, i.e., on a plane (Isaaks and Srivastava 1989; Pannatier 1996). The phenomenon under study is modelled as a *random function* $Z(\boldsymbol{x})$ that depends on the spatial coordinates $\boldsymbol{x}$ on a plane and contains information about (i) some statistical regularity, which is also called the spatial trend, and (ii) local randomness. The random variables $Z(\boldsymbol{x})$ and $Z(\boldsymbol{x} + \boldsymbol{h})$, which are obtained at the points $\boldsymbol{x}$ and $\boldsymbol{x} + \boldsymbol{h}$, are separated by a vector $\boldsymbol{h}$. Figure 19.2 shows an example of the vector $\boldsymbol{h}$. It places the point $\boldsymbol{x}_1$ with coordinates (1,2) at the point (8,4). For neighbouring sites (i.e., those with small values of $\boldsymbol{h}$), $Z(\boldsymbol{x})$ and $Z(\boldsymbol{x} + \boldsymbol{h})$ may be statistically dependent (spatially correlated). Finding the spatial dependence of these variables is the aim of geostatistical analysis.

In geostatistics, we have a model for the population and we need to estimate its parameters using sample data just like we do in linear regression (Chapter 5). In spatial analysis, $z(\boldsymbol{x})$ is considered a realisation of the random variable $Z(\boldsymbol{x})$, so we ultimately want to characterise the population $Z(\boldsymbol{x})$ using the observed value $z(\boldsymbol{x})$.

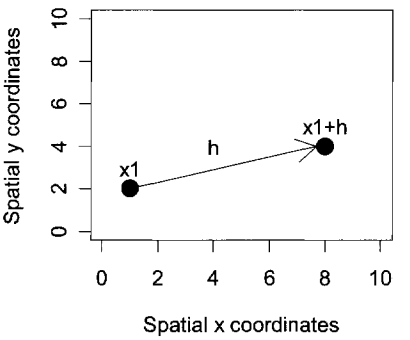

Figure 19.2. Illustration of the separation vector $\boldsymbol{h}$ that places the point x_1 with coordinates (1,2) at (8,4). The length of $\boldsymbol{h}$ is 7.3 (using the Pythagoras theorem), and $\boldsymbol{h}$ adds 7 and 2 units, respectively, to the x and y coordinate of x_1.

A random variable like $Z(\boldsymbol{x})$ is typically described by a density function. However, we have multiple spatial locations $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k$, and at each location we have a population $Z(\boldsymbol{x}_i)$ and an observed value $z(\boldsymbol{x}_i)$. In our illustration in Figure 19.1, we have a similar situation. Sample position on the landscape is designated by $\boldsymbol{x}_i$ (and determined by longitude and latitude). The population value of whatever we measure (relief, rainfall, or number of birds) at the site is designated $Z(\boldsymbol{x}_i)$, and the observed value of the measured variable is $z(\boldsymbol{x}_i)$. The cumulative distribution function for $Z(\mathbf{x}_1),\ldots,Z(\boldsymbol{x}_k)$ is given by[1]

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_k}(z_1,\ldots z_k) = P(Z(\mathbf{x}_1) < z_1,\ldots,Z(\mathbf{x}_k) < z_k) \tag{19.1}$$

[1] A refresher in density curves. We have explored density functions already when we examined the normal, Poisson, and binomial distributions in Chapters 5 and 6. Remember the simple exercises from a basic course in statistics that ask you to determine the probability of getting a particular number of items given a specified distribution function. For example, if the number of birds is Poisson-distributed with a mean of 5, what is the probability that we count not more than 10 birds in a particular location $\boldsymbol{x}$? The question is formulated mathematically as find $P(Z(\boldsymbol{x}) < 10)$, if $Z(\boldsymbol{x})$ follows a Poisson distribution with mean 5. The probability $P(Z(\boldsymbol{x}) < 10)$ is also called the cumulative distribution function $F(\boldsymbol{x})$; and, for this specific example, the notion for the distribution function is $F(10)= P(Z(\boldsymbol{x}) < 10)$. We have a similar notation in spatial statistics; see equation (19.1). Suppose that k represents the total number of sample locations in Figure 19.1. If we count birds at all of these locations and assume that bird abundances in the k locations are independent, then the cumulative density function in equation (19.1) can be simplified using the probability rule $P(A,B) = P(A) \times P(B)$. However, if the k locations are spatially dependent, then this is incorrect and, instead, we would use the cumulative distribution function in equation (19.1).

This equation is also called the *spatial distribution* of the random function $Z(\mathbf{x})$ (Chiles and Delfiner 1999). In a normal distribution, the density curve is fully specified by the mean (first moment) and the variance (second moment). Linear geostatistics also uses the first two moments of the random function. The first moment is the expectation or mean. The second moment is represented by the covariance or variogram. The mathematical notations for the mean and covariance are as follows:

$$m(\mathbf{x}) = E[Z(\mathbf{x})]$$

$$Cov(\mathbf{x}_1, \mathbf{x}_2) = E[(Z(\mathbf{x}_1) - m(\mathbf{x}_1))(Z(\mathbf{x}_2) - m(\mathbf{x}_2))]$$

The definition of the covariance is similar to a time series. However, instead of finding the association or relationship between two points in time, we want to find the relationship between two points in space. The association is quantified by coherence of the $Z(\mathbf{x}_1)$ and $Z(\mathbf{x}_2)$ deviate from their respective means.

On a real spatial landscape with continuous variables, we cannot sample such that we have values for every point on the landscape. Sometimes we need to estimate values for a regional variable $z(\mathbf{x})$ at a point where it is not measured. To estimate values, we use a geostatistical interpolation process called *kriging* (Chiles and Delfiner 1999). To interpolate values by kriging, however, we need to find an unbiased estimator for $Z(\mathbf{x})$ that has minimal uncertainty. This requires estimating parameters, and structural analysis or variography is used for this. Its main tool is the variogram. The variogram is a commonly used tool in geostatistics that measures how much of the variation in a measured variable is due to the spatial location of the sample itself. The variogram is defined as follows:

$$\gamma(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} E[(Z(\mathbf{x}_1) - Z(\mathbf{x}_2))^2]$$

The variogram or $\gamma(\mathbf{x}_1,\mathbf{x}_2)$ measures spatial dependence. Usually, if sample sites are located close to each other in space, the variable of interest will have similar values at each site and the difference calculated for the variogram value will be small. A low value of $\gamma(\mathbf{x}_1,\mathbf{x}_2)$ may indicate dependence between $Z(\mathbf{x}_1)$ and $Z(\mathbf{x}_2)$, whereas a large value may indicate independence.

Structural analysis is used to create a quantitative model for spatial correlation. In the geostatistical estimation process, certain conditions are imposed on the random function and its spatial correlation. These assumptions allow us to apply spatial analysis tools like variography and kriging. The most important geostatistical assumptions require that the random function $Z(\mathbf{x})$ is *stationary* and *ergodic,* and that it follows a multivariate *normal* distribution. The distribution assumption means that for any set of k variables, $Z(\mathbf{x}_1),\ldots,Z(\mathbf{x}_k)$ must be multivariate normally distributed.

The ergodicity assumption requires that the average over all possible realisations is equal to the average over the single realisation. For example, whether you examine 10 maps to get information about an area or just examine 1, the resulting impressions should be similar if map information follows the ergodicity assump-

tion. The ergodicity assumption allows parameter estimations based on only one realisation of the random variables.

As to stationarity, this is a slightly more complex issue. The random function *Z(x)* (the variable of interest) is stationary when its spatial cumulative distribution function does not vary when the points are translated by the vector **h**:

$$F_{\mathbf{x}_1,\ldots,\mathbf{x}_k}(z_1,\ldots z_k) = F_{\mathbf{x}_1+\mathbf{h},\ldots,\mathbf{x}_k+\mathbf{h}}(z_1,\ldots z_k)$$

The stationarity assumption would hold, for example, if we count birds in k locations, and the probability that we count certain values at these k locations is the same at k other locations, translated by **h**, elsewhere in the study area. In practice, this form of stationarity is unrealistic. This definition of stationarity is also called strong stationarity, and as the name suggests we can also define weak stationarity or second-order stationarity. There is even a third definition of stationarity, namely intrinsic stationarity.

Second-order stationarity means that the mean does not depend on location and the covariance is only a function of the separation vector $\boldsymbol{h}$. In mathematics:

$$E[Z(\mathbf{x})] \equiv m \quad \text{and} \quad Cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})] = C^*(\mathbf{h})$$

Strong stationarity implies weak stationarity but not the other way around. It is important to realise that so far, we have not said anything about the direction of $\boldsymbol{h}$. Indeed, the $\boldsymbol{h}$ in C^* is causing a translation of points in the same direction. This is visualised in Figure 19.3A, in which we connected points that are on a distance $h = 2$, but all connecting lines have the same direction. The covariance C^* at $h = 2$ can be estimated using only the points that are connected in Figure 19.3-A. *Isotropy* (discussed in more detail below) means that the direction is not important, only the distance between the points.

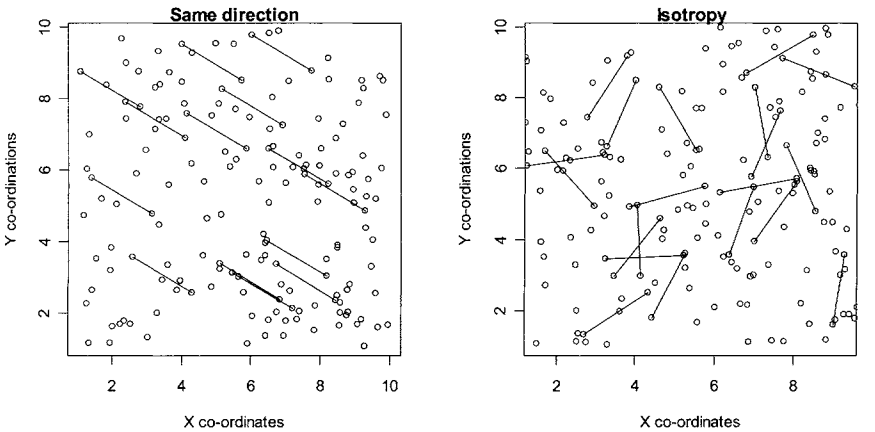

Figure 19.3. A: Points that are on a distance h with the same direction are connected. B: Points that are on a distance h in *any* direction are connected. In both graphs $h = 2$ was used.

Second-order stationarity and isotropy means that the covariance is calculated by

$$Cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})] = C(\|\mathbf{h}\|)$$

The notation $\|\mathbf{h}\|$ stands for Euclidean distances (Chapter 10). We omitted the * above the C to emphasise the effect of isotropy. In the spatial statistical literature one often uses $C()$ for both $C^*()$ and $C()$. The assumption of isotropy implies which one it is. Under the second-order stationarity and isotropy assumptions, the variance of $Z(\mathbf{x})$ is constant and location independent:

$$C(\mathbf{0}) = Cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{0})] = \text{var}(Z(\mathbf{x}))$$

If a time series is not stationary, one can look at differences between time points (first differences, see Chapter 16). In spatial statistic, we can do something similar, we can look at the difference $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$, and assume that this is second-order stationary. This is called intrinsic stationary. This is a weak form of stationarity, which assumes that the mean can have a local linear drift and the variogram only depends on the separation vector between points. These assumptions are represented mathematically as follows:

$$E[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})] \equiv m \quad \text{and} \quad Var[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h}).$$

Note that intrinsic stationarity only uses the first two moments. Again, we can assume that $\gamma(\mathbf{h})$ does not depend on the direction (isotropy). Strong stationarity, as the name suggests, is much stronger because it makes a statement on the spatial cumulative distribution functions.

Before some spatial analyses like variography and kriging can be applied on a data set, the three assumptions given above must be verified. Multivariate normality can be verified by making QQ-plots or histograms of individual variables. The intrinsic hypothesis can be verified by checking for a spatial trend. Ergodicity cannot be verified, unless samples are taken repeatedly in time and space.

19.3 Exploratory variography analysis

The aim of variography (Isaaks and Srivastava 1989; Pannatier 1996; Chiles and Delfiner 1999) is to estimate the strength and direction of spatial data dependency. As stated, the basic tool of variography analysis is the variogram or semivariogram (these terms are interchangeable). A variogram measures the spatial dependence between the values of the variable of interest in the points (locations) $\boldsymbol{x}$ and $\boldsymbol{x} + \boldsymbol{h}$, where $\boldsymbol{h}$ is the separating vector or lag (Figure 19.2).

Experimental variogram

In Section 19.2, we defined a variogram for the population, but how do we calculate such a population variogram using sample data? First, we need to calculate a *variogram estimator* using sample values just like we use samples to estimate population parameters in linear regression. This estimator is called an *experimental* variogram and it is calculated by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2\mathrm{N}(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{x}_i + \mathbf{h}) - z(\mathbf{x}_i)]^2$$

We used the notation $\hat{}$ to emphasise that it is an estimator based on sample data, not the population variogram. All the pairs of the sampled points that are separated by a specific *lag* $\boldsymbol{h}$ are used to calculate $\hat{\gamma}(\mathbf{h})$. If there is spatial dependence, then points close to each other will tend to have similar values and $\hat{\gamma}(\mathbf{h})$ will be small. If $\hat{\gamma}(\mathbf{h})$ is large, it indicates spatial independence.

$\boldsymbol{h}$ can be referred to as either a vector or as a number denoting the vector's length in this analysis. The context should be clear enough in the following discussions to indicate whether we mean the vector itself or its length. Both the length of $\boldsymbol{h}$ and its direction are important in this variogram.

The number of points that are separated from each other by a distance $\boldsymbol{h}$ is denoted by $N(\boldsymbol{h})$. In general, the number of points that are exactly separated by a distance of $\boldsymbol{h}$ will be very small so a small tolerance around the value of $\boldsymbol{h}$ is usually used instead of exact values. To illustrate this characteristic, we simulated the spatial distribution of 100 points (the coordinates were drawn from a univariate distribution), and we plotted the results in Figure 19.4. The figure contains four panels, each with a precise $\boldsymbol{h}$ value. In each panel, we connected the points that were *exactly* separated by a distance $\boldsymbol{h}$, where $h = 1$ (upper left panel), 2, 3 and 4 (lower right panel). As you can see, there are only two points that have exactly a distance of 1. There are no other connected points in any of the other panels. Obviously, we cannot calculate the variogram using only two points. Therefore, we need a small tolerance adjustment around the lag distance. We arbitrarily selected the interval $h - 0.05$ to $h + 0.05$, where h is the lag. The effect of introducing a lag tolerance is shown in Figure 19.5. Note that there are now several connected points in $\mathrm{N}(\boldsymbol{h})$ (which is therefore larger).

The variogram $\hat{\gamma}(\mathbf{h})$ is calculated by subtracting the observed values $z(\boldsymbol{x}_i)$ at both ends of a line from each other and squaring the difference. This is then repeated for all points that are connected by the given lag. If there are not enough connecting lines, the tolerance interval can be increased.

If we calculate an experimental variogram using all the points that are connected by the lines in Figure 19.4, we assume that the spatial relationships are the same in all directions, which is referred to as *isotropy*. To assume that the spatial dependence is the same in every direction, or *omnidirectional*, is a reasonable starting point for any spatial analysis. The resulting experimental variogram is considered as an average variogram of all spatial directions.

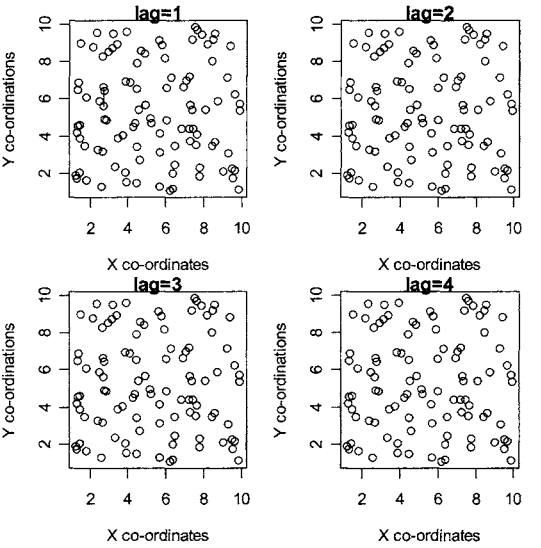

Figure 19.4. 100 Randomly selected sites. Points that are exactly 1 (upper left), 2 (upper right), 3 (lower left) or 4 (lower right) units separated are connected by a line. The numbers in the title are the lags. As can be seen (with some effort) from the upper left panel, there is only one combination of points that are exactly separated by the lag = 1.

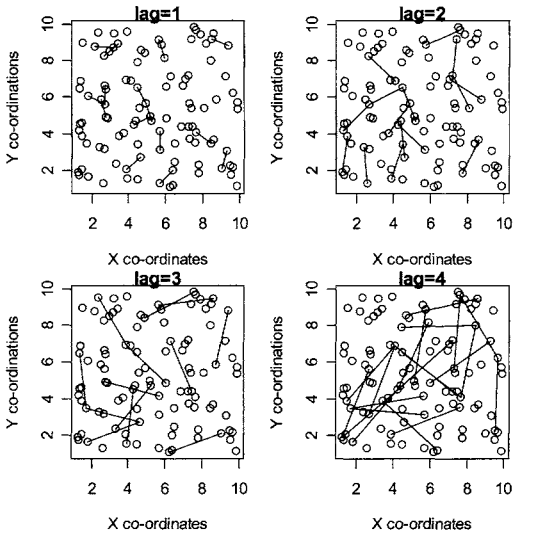

Figure 19.5. The same 100 randomly selected sites as in Figure 19.3. Points that are separated by a distance between $h - 0.05$ and $h + 0.05$ are connected by a line. The lag represents the value of $\boldsymbol{h}$.

How do you verify, however, that the variogram has the same spatial relationships in each direction? The easiest method is to calculate *directional* experimental variograms, each focusing on a specific direction, and then compare the variograms with each other to see whether the shape changes. This can be done by calculating $\hat{\gamma}(\mathbf{h})$ for only those lines that point in the same direction, say from north to south or from east to west. During these tests, however, we may end up with the same problem as we encountered when using lag values without tolerances (Figure 19.4), namely that there are only a few lines pointing exactly in the right direction. The solution is to again use a tolerance range, but this time to apply it on the direction. Usually, optimum values of the tolerance around distances and angles are estimated iteratively and details can be found in Cressie (1993). Another way to verify isotropy is using a variogram surface, which is discussed next.

Variogram surface

Recall from Figure 19.2 that the vector $\boldsymbol{h}$ adds a horizontal and vertical replacement to the x-y coordinates and has a directional angle for the change. Instead of using a specific angle and length of $\boldsymbol{h}$, however, we can express the same concept as a replacement in terms of West-East and North-South map units. For example, Figure 19.6 shows a replacement of two (with some tolerance) units in the West-East direction and two units in the South-North direction. For this replacement, we can calculate the experimental variogram as described in the previous paragraph but allowing for a small tolerance around the length and direction of $\boldsymbol{h}$, defined on a grid. This exercise is done for each possible grid cell and results in a experimental variogram with values in tabular (grid surface) form; e.g., see Table 7.5 in Isaaks and Srivastava (1989).

Instead of trying to analyse all of the numbers in such a table/figure for spatial relationships, we make a contour plot of the experimental variogram values called an *experimental variogram surface* (Isaaks and Srivastava 1989). This graph is used to detect violations in isotropy or *anisotropy*.

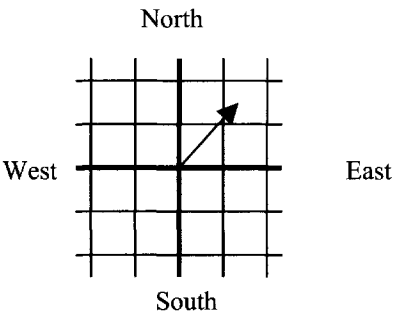

Figure 19.6. Example of replacement along the West-East and North-South axes.

Anisotropy indicates dependence in the data that changes in different directions. Recall that in Chapter 10 we used bird abundance data measured by a radar. If the birds use the wind direction for their flight path, then the dependencies are anisotropic. Plant abundances measured along an altitude gradient are also anisotropic because their spatial dependencies will be different based on map directions. When we calculate experimental variograms for different directions, they should be referred to as *experimental directional variograms*. In practice, we will use covariates to model the trends in the data and hope that the residuals are isotropic.

h-Scatterplots

An *h*-scatterplot can be used to visualise violation of the stationary and multivariate normality assumptions. Instead of plotting two different variables against each other, we take a particular site *i* and determine which other sites are on a distance *h* (scalar value), again allowing for a small tolerance around h. We then plot our observed value of $z(\boldsymbol{x}_i)$ against the values of $z(\boldsymbol{x}_i + \boldsymbol{h})$, for all the stations that are on a distance *h*. Let us explain this in a bit more detail using our lag values from Figure 19.5. First, we take a particular point $\boldsymbol{x}_i$ and make a list of sites $\boldsymbol{x}_j$, $\boldsymbol{x}_{j+1}$, $\boldsymbol{x}_{j+2}$, etc., that are all at the same distance *h* from $\boldsymbol{x}_i$. In the *h*-scatterplot, we plot $z(\boldsymbol{x}_i)$ versus $z(\boldsymbol{x}_j)$, $z(\boldsymbol{x}_i)$ versus $z(\boldsymbol{x}_{j+1})$, $z(\boldsymbol{x}_i)$ versus $z(\boldsymbol{x}_{j+2})$, etc. This gives a vertical band of points because we are using the same coordinate $z(\mathbf{x}_i)$ along the *x*-axis. We repeat this process for all sites, which gives us layers of vertical dots. Obviously, the distribution of points in this graph depends on the value of h. It is common to make a couple of *h*-scatterplots for different values of h and plot them next to each other to see how *h* values affect the distribution.

If the values of $z(\boldsymbol{x}_i)$ and $z(\boldsymbol{x}_{j+k})$ are similar for all *i* and *k*, (implying spatial dependence), then most points should lie close to a 45-degree line halfway between the horizontal and vertical axes and starting at the origin. For small *h*, we expect to see a narrow band along this 45-degree line. When *h* is larger, the spatial dependence decreases, and the cloud of points is likely to get 'fatter' (Figure 19.7-A). If the univariate distribution of $Z(\boldsymbol{x})$ is highly skewed to the right, then small values of $z(\boldsymbol{x}_i)$ correspond to large values of $z(\boldsymbol{x}_i+\boldsymbol{h})$ and vice versa. These relationships form lines of dots adjacent to the axes of the *h*-scatterplot and are referred to as a 'butterfly wing' pattern (Figure 19.7-B). Note that a histogram ignores the spatial dependences and is therefore less suitable to verify stationarity. A single cloud located above or below the diagonal line indicates a systematic drift in the direction of the vector $\boldsymbol{h}$. It means that the value of $z(\boldsymbol{x}_i+\boldsymbol{h})$ has systematic bias from $z(\boldsymbol{x}_i)$. If the cloud is located above the diagonal, values of $z(\boldsymbol{x}_i)$ systematically increase in the direction of $\boldsymbol{h}$ (Figure 19.7C). Clouds of points far from the diagonal line in the scatter plot are usually formed by pairs belonging to different statistical populations. For example, if we have two populations with different means (as in Figure 19.7-D), the *h*-scatterplot has four individual clouds: The two clouds on the diagonal correspond to values of $z(\boldsymbol{x}_i)$ and $z(\boldsymbol{x}_i+\boldsymbol{h})$ from the same populations; the cloud in the right bottom corner corresponds to values of $z(\boldsymbol{x}_i+\boldsymbol{h})$ from the smaller mean population and $z(\boldsymbol{x}_i)$ from the larger mean population; and the cloud in the left top corner corresponds to values of $z(\boldsymbol{x}_i+\boldsymbol{h})$ from the larger mean population

and $z(\mathbf{x}_i)$ from the smaller mean population. In this case, $Z(\mathbf{x})$ cannot be considered stationary.

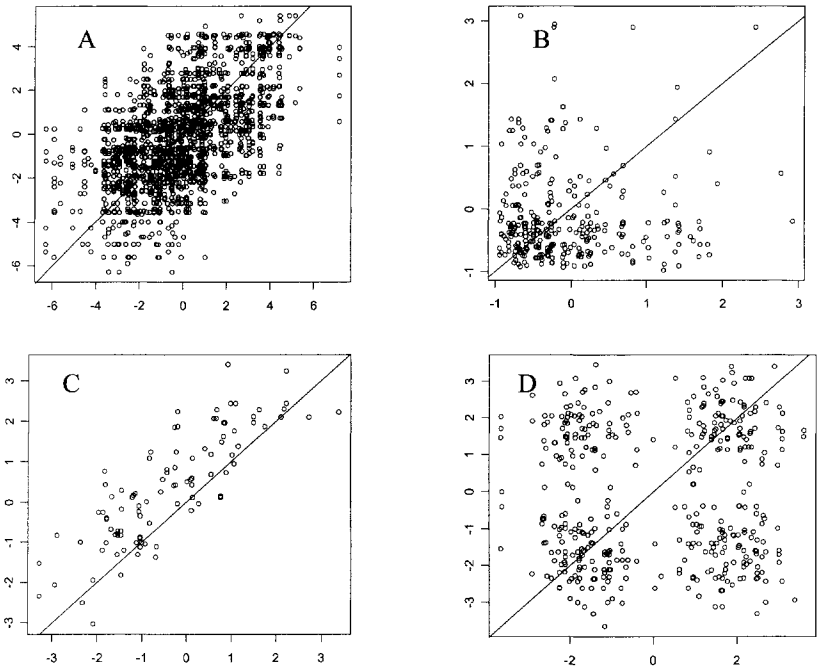

Figure 19.7. *h*-Scatterplots for various different simulated data sets. A: *h*-Scatterplot for the data drawn from the multivariate Gaussian distribution. B: Data violating the multivariate Gaussian distribution properties: the 'butterfly wing' for the highly skewed distribution. C: Non-stationarity caused by the mean drift. D: Two populations.

Illustration: Volga River relief data — detecting trends in spatial data

The shape of the relief data in Figure 19.1 suggests a spatial trend. To eliminate this trend, a linear regression model using the longitude (X) and latitude (Y) coordinates as covariates was fitted. The model gave the following results:

	Estimate	Std.Error	*t*-value	*p*-value
Intercept	86.970	0.420	209.54	<0.001
X	-0.005	0.0002	-32.11	<0.001
Y	0.008	0.0002	39.34	<0.001
Residual standard error: 1.97 on 169 degrees of freedom				
Multiple R-Squared: 0.941				
F-statistic: 1348 on 2 and 169 df, *p*-value: <0.001				

The residual standard deviation is about 2 m. The trend surface model was subtracted from the observed values, and the residuals were normally distributed (as indicated by histograms and QQ-plots). The residuals were plotted against their locations in Figure 19.8. In this example we have access to a much larger part of the data, but we only used 20% in the trend fitting. This allows us to calculate the trend residuals from the true elevation values at the 'un-sampled' (the other 80% of the data) locations. Although the trend was fitted at discrete sample points, we present all the trend residuals because it is visually easier to interpret. White or dark grey areas indicate large positive or large negative residuals. The distribution of the colours indicate areas of spatial dependence. If there are clear directional patterns formed by the shades of colour, we have evidence of anisotropy.

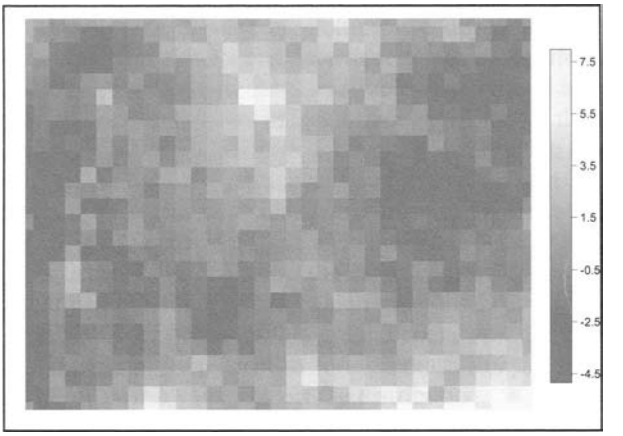

Figure 19.8. Residual pattern for the linear trend model (residuals for all locations are shown). The horizontal and vertical axes represent spatial coordinates, and the grey-scaling indicates the relative value of the residuals. Large white or dark grey areas, or directional patterns, indicate violation of stationarity and isotropy.

To assess whether the residuals are independent (i.e., whether we violate the assumption of independence) in the linear regression model, the experimental variogram for the residuals was calculated for 20 lags and with a lag distance of 200 m. The resulting variogram is shown in Figure 19.9-B. Because no anisotropy was detected on the variogram surface (Figure 19.9-A), the omnidirectional variogram was used. Under second-order stationarity, it can be shown (Schabenberger and Pierce 2002) that $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$; hence, the variogram is always smaller than the variance. In this case, the residual experimental variogram value exceeds the theoretical limit (residual variance) for lags above 1000 m (Figure 19.9-B). The reason for this is the deviation from the Gaussian model for larger lags as can be seen in Figure 19.9-D. At lags above 1000 m, the *h*-scatterplot reveals the 'butterfly wing' pattern with large values of $z(\boldsymbol{x})$ corresponding to small values of $z(\boldsymbol{x} + \boldsymbol{h})$ and vice versa. This means that the values are negatively corre-

lated at this distance and the average value of $[z(\boldsymbol{x}_i + \boldsymbol{h}) - z(\boldsymbol{x}_i)]^2$ becomes larger than the variance. This may suggest some spatial structure like a hill, which has a typical size in the area of about 1500 m (implying violation of the second-order stationarity assumption).

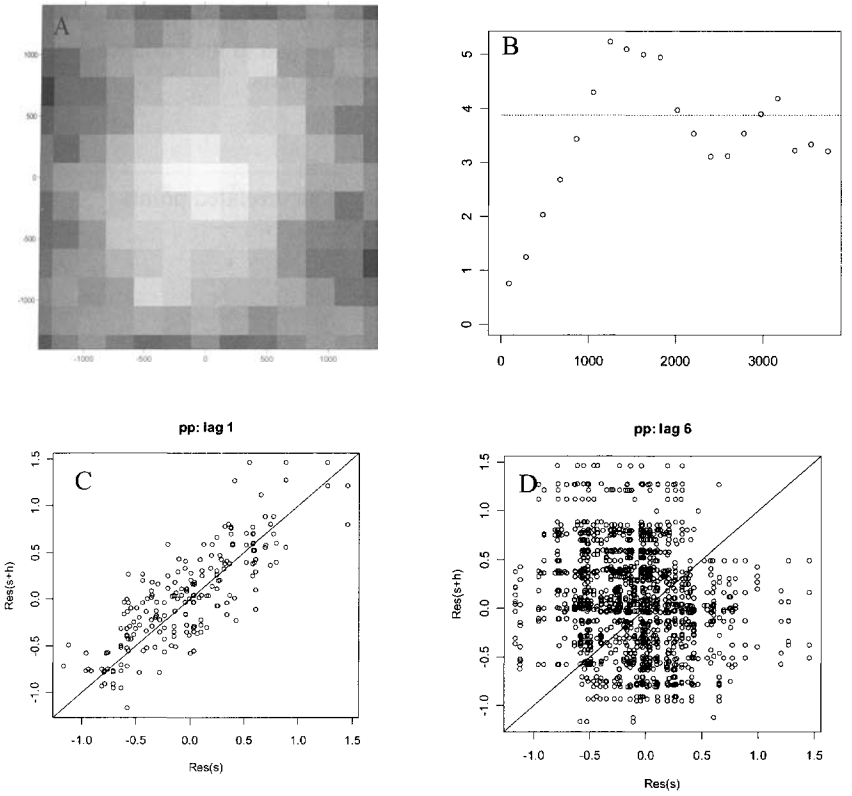

Figure 19.9. A: Experimental variogram surface for residuals from the linear trend. The variogram values for various West-East and North-South displacements were calculated, and a grey-level was used to visualise the values. B: Experimental variogram with the residuals variance drawn. *h*-Scatterplots for lag = 200 m (C) and for lag = 1200 m (D).

We will now work towards models for the variogram. Note that the experimental variogram in Figure 19.9-B only contains dots, and these are based on distances between the sampled locations. To analyse the entire area, we need to know the value of the variogram at intermediate distance values. In order to obtain these values, we need to fit a model on the experimental variogram and obtain a fitted line through the points in the experimental variogram. This model has parameters that need to be estimated. Once we know the parameters, the model can also be

used in kriging or regression type models (Chapter 37). We first discuss possible line shapes that fit the points and then discuss their associated models.

A typical variogram has low values for small distances $\boldsymbol{h}$ (points close to each other tend to be similar) and increasing values for larger $\boldsymbol{h}$ that eventually reach an asymptotic value beyond which sites are independent (Figure 19.10). The asymptotic value itself is called *sill*, and the distance $\boldsymbol{h}$ at which it is reached is the *range*. If $Z(\mathbf{x})$ exhibits stationarity, then the sill is equal to the variance. See Figure 19.10 for an illustration of each of these variogram parts.

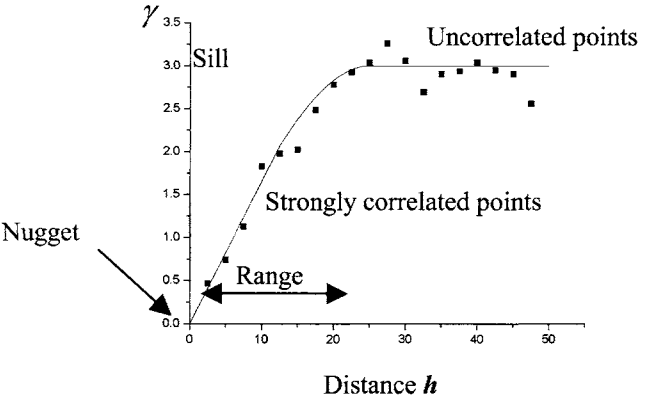

Figure 19.10. Variogram with fitted line. The sill is the asymptotic value, and the range is the distance at which this is reached. Points on a distance larger than the range are uncorrelated. The nugget effect occurs if $\hat{\gamma}(\mathbf{h})$ is not 0 for small $\boldsymbol{h}$.

What kinds of patterns are usually depicted in the experimental variogram? Typically, the variogram takes on one of four shapes. The first shape looks like part of a *parabolic* function (Figure 19.11-A). It shows a smooth pattern for small $\boldsymbol{h}$. The second shape looks *linear* for small $\boldsymbol{h}$ values (Figure 19.11-B). This is most common in areas that are small in comparison with the range of the phenomenon. The experimental variogram may not reach the asymptotic value and has only increasing values of $\boldsymbol{h}$. Note that both the parabolic and the linear patterns start at the origin. The third shape for the variogram may have distinct patterns that start above the origin (0,0). This result occurs when $\hat{\gamma}(\mathbf{h})$ is not 0 for small $\boldsymbol{h}$, and it is also known as the *nugget effect (*Figure 19.11-C). It represents the discontinuity of the regional variable caused by spatial structure at distances less than the minimum lag (also called the microstructure) or measurement error. The fourth shape of the variogram may resemble a *flat* line (Figure 19.11-D). It occurs when there is no correlation between the values taken at two locations at any distance. This is an extreme case of an absence of spatial structure.

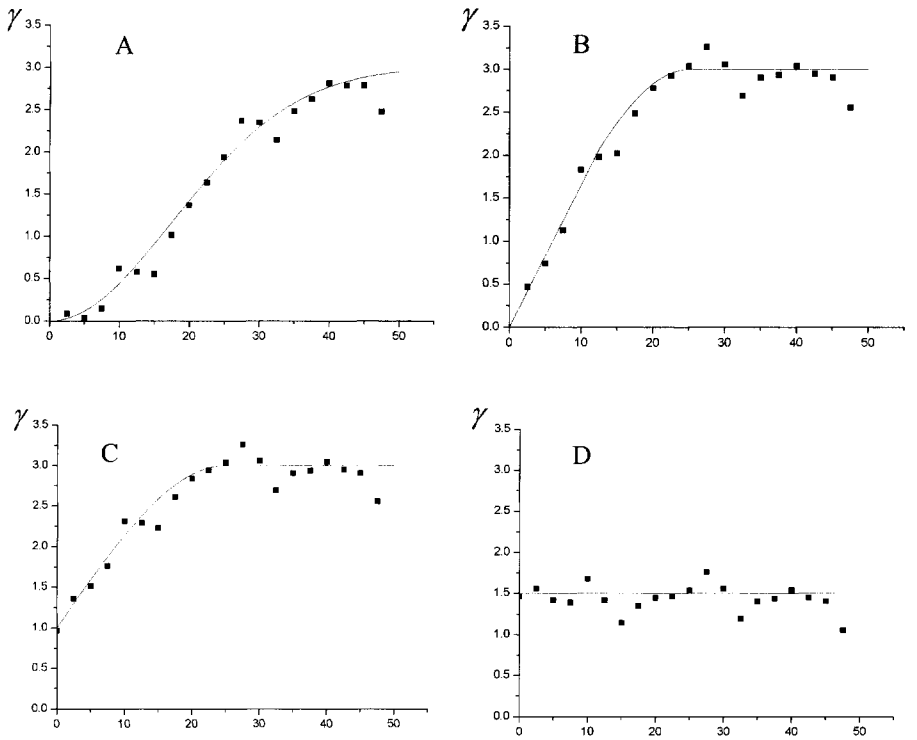

Figure 19.11. Variogram behaviour at short distances examples: parabolic (A), linear (B), nugget (C) and pure nugget (D). The horizontal axis in each graph contains values of ***h***.

Variogram models

The shape of the experimental variogram allows one to define and fit a variogram model or a *theoretical variogram*. Only a certain number of basic models meet the special requirements of the variogram modelling (Chiles and Delfiner 1999). The most commonly used are the spherical, power and wave models; Gaussian models sometimes better fit the experimental variogram but are hard to implement in kriging. The spherical model $\gamma_{sph}(h,r,\sigma)$ and the Gaussian model $\gamma_{sph}(h,r,\sigma)$ are defined by:

$$\gamma_{sph}(h,r,\sigma^2) = \begin{cases} \sigma^2(1.5\frac{h}{r} - 0.5\frac{h^3}{r^3}) & \text{if } h < r \\ \sigma^2 & \text{otherwise} \end{cases}$$

$$\gamma_G(h,r,\sigma^2) = \sigma^2(1 - e^{-(h/r)^2})$$

The wave model $\gamma_w(h,\phi,\sigma^2)$ is defined by:

$$\gamma_w(h,\varphi,\sigma^2) = \sigma^2(1-(\varphi/h)\sin(h/\varphi)),$$

The nugget model is given by

$$\gamma_{ne} = \begin{cases} 0 & \text{if } h = 0 \\ c & \text{otherwise} \end{cases}$$

Within each of these models, r is the range, σ^2 is the variance or sill value, and φ is the period parameter. The variable $\boldsymbol{h}$ defines the distance between the sites and is the separating lag. The unknown parameters in each of these models are the range and sill (variance). The nugget effect is modelled by adding a known constant. If the area under investigation is small in comparison with the correlation range, the basic model for the variogram without asymptotic behaviour can be used, such as a power function with $0 < a < 2$:

$$\gamma_{pow}(h,a,\sigma^2) = \sigma^2 h^a$$

More examples of variogram models are presented in Chiles and Delfiner (1999) and Cressie (1993). Figure 19.12 shows some of the patterns that can be obtained for these models. To decide which model is most appropriate for your data set, you need to consider both the shape of the experimental variogram and the spatial characteristics of the phenomenon under study. Improper variogram selection may result in unstable kriging results. Cross-validation (see also Chapters 7 and 9) may be used to test whether the selected variogram model is appropriate.

For the variogram models that we have discussed so far, we assumed isotropy. If the data show anisotropy, there are two possible types. The first is directional (or geometrical) anisotropy. In this case, the sill is the same for all $\boldsymbol{h}$ directions, but the ranges may vary. In that case we need to calculate directional variograms for several directions instead of the omnidirectional variogram. Usually the anisotropy direction (direction of higher correlation), direction perpendicular to it and one or two intermediate directions are used. Then we need to allow for different ranges within the model and this can be done by a scaling of $\boldsymbol{h}$ according to the directions (Cressie 1993; Chiles and Delfiner 1999). The extreme case of the directional anisotropy is *zonal anisotropy* in which the sill varies with the $\boldsymbol{h}$ direction (Chiles and Delfiner 1999).

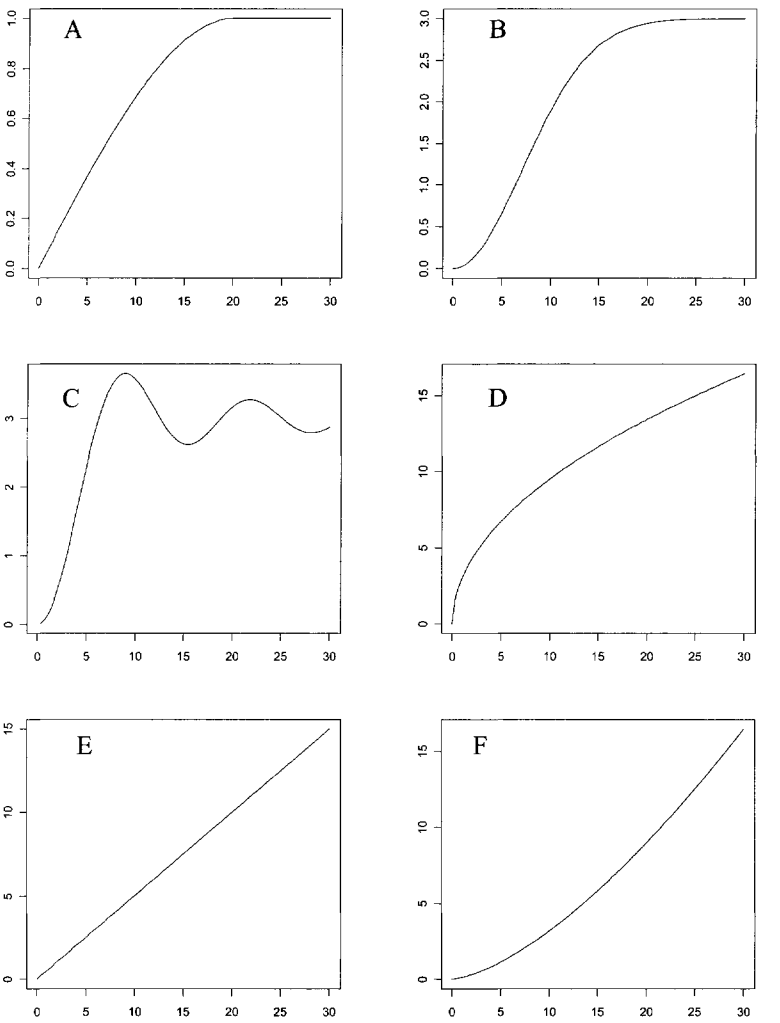

Figure 19.12 Typical variogram model shapes. A: Spherical (r = 20), B: Gaussian (r = 10), C: Wave (φ =2), D: Power (a = 0.5), E: Linear (power with a = 1), F: Power (a = 1.5) models.

Illustration: Volga River relief data — fitting a variogram model

For our illustration data, we fitted two theoretical variogram models to the experimental variogram for the relief data in Figure 19.1. We chose to test the spherical and wave models. For the spherical model, we obtained the following variogram-model parameters: range = 1300, nugget = 0.17 and sill = 3.7. For the wave model, the parameters were period = 310, nugget = 0.6 and sill = 3. The fit-

ted variogram models are shown in Figure 19.13. The spherical variogram incorporates the medium range behaviour and does not include the large-scale residuals pattern; the wave variogram models take into account the large-scale residuals pattern, but it gives a larger nugget effect.

To estimate the variogram model parameters, we used a direct minimization of the sum of squared difference between the experimental variogram $\hat{\gamma}(\mathbf{h})$ and its theoretical model $\hat{\gamma}(\mathbf{h}, \theta)$ as follows:

$$\sum_i (\hat{\gamma}(\mathbf{h}_i) - \hat{\gamma}(\mathbf{h}_i, \theta))^2$$

θ is the vector of the parameters of variogram model. Calculations were performed over all the lags $\mathbf{h}_i$. Other robust methods for variogram model fitting can be found in Cressie (1993). The best method to test whether either of these models fit the relief data is to substitute the models into kriging equations and to compare the kriging results, which we will do in the next section. Alternatively, we could use independent sample tests or modified cross-validation tests to select the model that best fits the phenomenon at both sampled and unsampled locations.

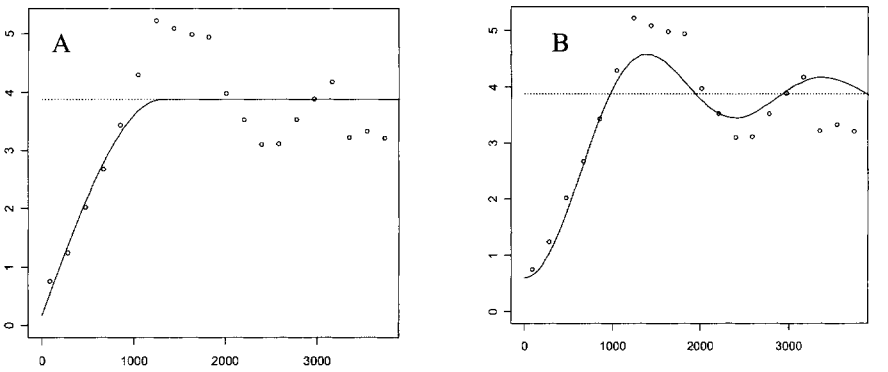

Figure 19.13. Fitted variogram models for Volga River relief data. A: Fitted spherical variogram model with range = 1300, nugget = 0.17, and sill = 3.7. B: Wave variogram model with period = 310, nugget = 0.6 and sill = 3.8. The residuals variance is drawn as a dotted line.

19.4 Geostatistical modelling: Kriging

Kriging is one of those magical terms used in geostatistics that sounds exotic but is simply a term for the process of interpolating data values from areas that have been sampled on a landscape to areas that have not. In Figure 19.14, x_0 is surrounded by locations that have values for some variable of interest but x_0 itself has not been sampled. In nine sites, the variable of interested was measured (e.g.,

rainfall, relief, number of birds, and tree height) and we would like to predict the value at the site x_0. However, assigning a value to x_0 cannot be done with a simple interpolation from the nearest site. It involves considering the data's gradient in all directions around each sample point near x_0. We may even want to predict the variable of interest between each sampled site in the study area and create a contour plot of all possible values between them. For this interpolation process, kriging employs a weighted-average approach to assign values to unsampled areas based on their proximity to each other.

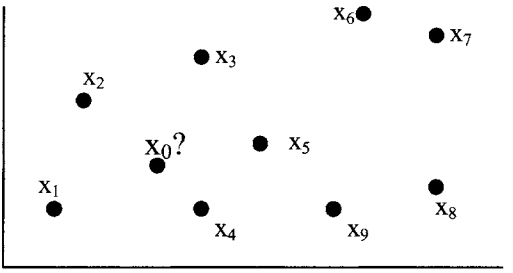

Figure 19.14. Nine sites labelled as x_1 to x_9, and at each site the variable of interest was measured: $z(x_1)$ to $z(x_9)$. The site labelled x_0 was not sampled and we want to predict its z value.

Statistically, if we have sampled N locations $x_1, x_2, \ldots, x_N$, at each location x_i, we have realisations $z(x_i)$ of the variables $Z(x_i)$. Our question then is, what is value of Z at an unsampled location x_0? To estimate the value $z(\mathbf{x}_0)$, it seems sensible to use information from neighbouring sites. Due to spatial dependence it is likely that the values in the points x_1 to x_5 contain more information on the variable $Z(x_0)$ than the sites x_6 to x_9, simply because the first five locations are closer. So, it makes sense to use some sort of weighted average of information from nearby locations to obtain the estimator for $z(x_0)$. The weighted average is denoted by $z^*(x_0)$ and can be estimated by

$$z^*(x_0) = \mu(x_0) + \sum_i \lambda_i (z(x_i) - \mu(x_i)) \tag{19.2}$$

where x_0 is the unsampled site, $\mu(\mathbf{x}_i)$ is the value of the mean in x_i, and λ_i are weighting functions that determine the importance of neighbouring sites. Sites that are far away will have a λ_i close to zero and for sites close to x_0, λ_i is relative large. As defined in Section 19.2, $Z(x_i)$ is the random variable and $z(\mathbf{x}_i)$ is the observed value. $z^*(x_0)$ is the predicted value of $Z(x_0)$ at the unsampled location. Equation (19.2) tells us that the value $z^*(x_0)$ is estimated as the trend plus a weighted sum of deviations from the trend at nearby sites.

Equation (19.2) looks similar to the linear regression model, where $z^*(x_0)$ is the Y, $\mu(x_0)$ the intercept, $z(x_i) - \mu(x_0)$ the X, and the λ_is the slopes. However, this ex-

pression differs from a linear regression equation in that the coefficients λ_i are not fixed constants. Each are allowed to differ for each location $\mathbf{x}_0$.

There are three questions that we now need to address. What are the values of λ_i? Do we actually know the trend values $\mu(\boldsymbol{x}_i)$? Can we ensure that $z^*(\boldsymbol{x}_0)$ is precise? The last question implies that we need to find an expression for the variance of $z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)$ that is as small as possible. You might wonder how the variogram is used in this analysis. The variogram is used to calculate the λ_is and in the expression of the variance of $z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)$. We also need to determine the mean or trend value $\mu(\boldsymbol{x}_i)$ in equation (19.1). There are three basic forms of kriging (Goovaerts 1997; Chiles and Delfiner 1999), namely:

1. Simple kriging (SK). Here we assume that the mean $\mu(\boldsymbol{x})$ is known.
2. Ordinary kriging (OK) in which we assume that the mean $\mu(\boldsymbol{x}) = \mu$ is unknown but constant.
3. Universal kriging (UK); $\mu(\boldsymbol{x})$ is unknown and varies over the area.

The exact definitions of the estimator in equation (19.2) and the variance $z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)$ depend on whether we know the mean $\mu(\boldsymbol{x})$ and if not whether it is constant or varies over the area. Once we know which type of kriging we have, finding the unknown weights λ_i follows the standard least squares approach. It is a matter of taking derivatives of the variance $z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)$, setting them to zero, and solving linear equations. In terms of matrix algebra, the solution for the known-mean case (SK) looks simple:

$$\boldsymbol{\lambda} = \mathbf{K}^{-1}\mathbf{k}$$

where $\mathbf{K}$ and $\mathbf{k}$ are a matrix and vector of covariance terms (which are closely related to the variogram model fitted to the experimental variograms), respectively, and $\boldsymbol{\lambda}$ contains all the weighting factors. Although the mathematical expressions used to calculate for the variance of $z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)$ are slightly more complex, calculations for simple kriging use the following relationships (supposing a linear dependence on covariates):

$$z_0^* = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m + u_0$$
$$u_0 = \sum_j \lambda_j u_j + \varepsilon$$
$$\varepsilon \sim N(0, \sigma_{err}^2)$$

This technique is very similar to the spatial moving average (SMA) model discussed in Chapter 18. The X_i are covariates (e.g., spatial coordinates) giving the trend model mean $\mu(\boldsymbol{x})$, and u_0 is the residual from the trend model. The regression parameter β_i models the effect of the explanatory variables on the $Z(\mathbf{x}_0)$, u_i is spatially correlated noise and λ_i are the kriging (averaging) coefficients for it. SK differs from an SMA model, however, because the averaging coefficients λ_i are not derived from the fixed neighbourhood matrix $\boldsymbol{W}$ but are newly calculated for each location $\mathbf{x}_0$ from the variogram model.

Illustration: Volga River relief data — selecting the variogram model

In Section 19.3, we identified two potential variograms that might model the residuals of the Volga River data well. In this section, we illustrate how to choose between the two variogram models using kriging. Recall that the following linear regression model was applied prior to our variogram model calculations in Section 19.3:

$$\text{Elevation}(\boldsymbol{x}) = \alpha + \beta_1 \text{ Latitude}(\boldsymbol{x}) + \beta_2 \text{ Longitude}(\boldsymbol{x}) + Z(\boldsymbol{x}) \quad (19.3)$$

Another way to express this relationship is:

$$H(\boldsymbol{x}) = H_{LM}(\boldsymbol{x}) + Z(\boldsymbol{x}),$$

where $H(\boldsymbol{x})$ is elevation at location $\boldsymbol{x}$ and $H_{LM}(\boldsymbol{x})$ is the spatial trend as modelled by the covariates in the linear regression model. The mean of the residuals of a linear regression model is always zero. However, $z(\boldsymbol{x})$ may be locally unequal to zero. We will apply two types of kriging on the relief data to estimate the trend residuals $z(\boldsymbol{x})$ at each location, namely: the wave variogram tested with ordinary kriging and the spherical variogram tested with universal kriging.

Wave variogram tested with ordinary kriging.

To apply this method, we assume that the expectation of the residuals $z(\mathbf{x})$ is an unknown constant, denoted by μ. We assume that large-scale smooth spatial changes in Figure 19.7 are caused by a stochastic process, and the changes are modelled using the wave variogram. In this case, equation (19.2) becomes

$$z^*(\mathbf{x}_0) = \mu + \sum_i \lambda_i (z(\mathbf{x}_i) - \mu)$$

Note that the spatial trend in the residuals is a constant μ. In this case we use ordinary kriging equations with the wave variogram fitted to calculate both μ and λ_i, and this allows us to estimate $z^*(\boldsymbol{x}_0)$. It is common to denote the estimator by $z^*_{OK}(\mathbf{x})$.

Spherical variogram tested with universal kriging.

In this approach, we assume that large-scale smooth spatial changes in Figure 19.7 are the result of the residual mean varying over the territory (non-stationarity). Therefore, equation (19.2) becomes

$$z^*(\mathbf{x}_0) = \mu_x(\mathbf{x}_0) + \sum_i \lambda_i (z(\mathbf{x}_i) - \mu_x(\mathbf{x}_i))$$

where $\mu_x()$ means that the mean values in $Z(\mathbf{x})$ differ for each location $\mathbf{x}$ and needs to be estimated simultaneously with the λ_i in the kriging system. It is common to model $\mu_x(\boldsymbol{x})$ as a polynomial function of spatial coordinates x and y (e.g., latitude and longitude):

$$\mu_{\mathbf{x}}(\mathbf{x}) = \beta_{\mathbf{x},0} + \beta_{\mathbf{x},1}x + \beta_{\mathbf{x},2}y + \beta_{\mathbf{x},3}xy + \beta_{\mathbf{x},4}x^2 + \beta_{\mathbf{x},5}y^2$$

Universal kriging (UK) with the local mean estimation and fitted spherical variogram can be used to estimate both the local trend and the λ_is, and it is denoted by $z^*_{UK}(\mathbf{x})$.

Figure 19.15-A shows the same relief elevation graph as in Figure 19.1, except that we now used a contour graph to plot all available data. Recall that in the examples in this chapter we used only 20% of the data to fit models. The fit of the linear regression model in equation (19.2) is shown in Figure 19.15-B. The results of the two kriging methods applied on the residuals, as described above are presented in Figure 19.15-C and Figure 19.15-D. So, panels B–D were obtained with 20% of the data, and panel A shows the true picture. Note that the kriged relief is smoother than the original source data (panel A). This is a generic aspect of the kriging estimations.

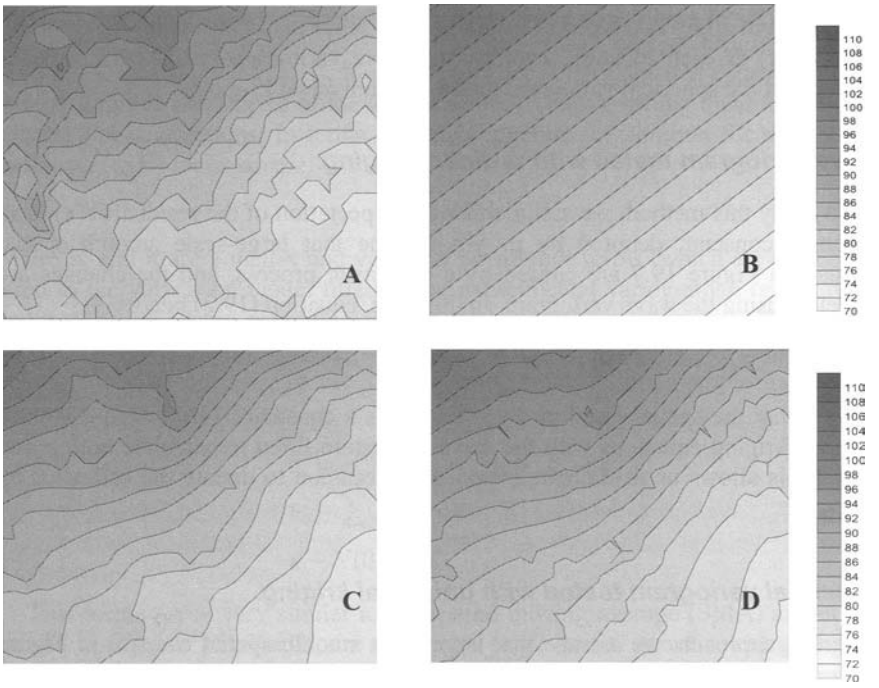

Figure 19.15. A: Contour plot for the original elevation relief data. B: Linear trend model. C: OK with spherical variogram. D: UK with wave variogram.

Because we have access to the detailed relief data for the area, we can compare the original relief structure and the kriged relief structure. The linear trend was used to calculate residuals from the true elevation $z(\boldsymbol{x})$. Differences between $z(\boldsymbol{x})$

values and the residuals estimations $z^*(\boldsymbol{x})$ obtained from kriging are calculated using $z(\boldsymbol{x}) - z^*(\boldsymbol{x})$. For our first approach of testing the wave variogram with ordinary kriging, we obtained a difference mean of 0.011 and a variance of 1.45. In the second approach using universal kriging with the spherical variogram model, the difference mean was −0.001 and the variance 1.17. Both types of kriging incorporated most of the spatial variation of the residuals and gave similar results. However, the spherical model showed slightly better prediction in the error mean and variance.

19.5 A full spatial analysis of the bird radar data

In Chapter 10, multivariate techniques were applied on a bird radar data set. A spatial grid was used to calculate numbers of birds per hour for 3 days. In this section, we apply the geostatistical analysis process that was introduced in this chapter and Chapter 18 to model the spatial relationships in the bird radar data. For this analysis, we will use totals per grid for only the first day. Recall from Section 19.1 that a full geostatistical analysis follows three steps: data exploration, analysis of spatial correlation, and kriging.

Data exploration

Moran's I index is one test for auto-correlation in spatial data. The Moran's I statistic for the bird radar data is 0.842, which indicates that there is positive correlation in bird abundance at neighbouring sites (Chapter 10). A permutation test also indicates that the spatial correlation of bird counts is significant ($p < 0.001$). So the closer the sites are located to each other, the more similar are their bird counts.

Moran spatial correlogram explores the correlation of bird counts at sites against the distances that separate each site. For the bird radar data, the correlogram in Figure 19.16 shows that bird counts at sites that are spaced 3000 m or less apart are significantly positively correlated. If bird counts are separated by 4000 m or more, they are significantly negatively correlated.

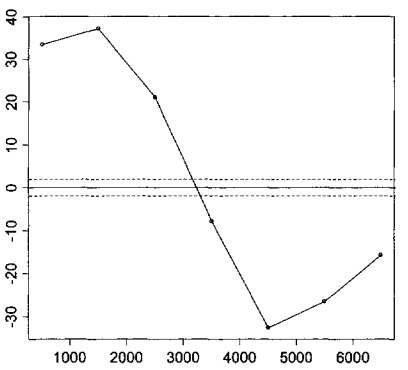

Figure 19.16. Moran's $I^{(k)}$ calculated for birds count at 7 distance bands (k = 0, 1000, 2000, 3000, 4000, 5000, 6000, 7000). Expectation (solid line) and critical values (dashed lines) are also given.

Sometimes, techniques are applied on spatial data that ignore the spatial correlations. However, ignoring spatial auto-correlation structure may cause important exploratory variables to be missed or an inadequate model to be selected to describe the data. As an example of how neglecting spatial correlations affects a linear model, we use the bird radar data in the following linear regression:

$$Count_i = \alpha + \beta_1\ Latitude_i + \beta_2\ Longitude_i + \varepsilon_i$$

where the response variable *Count* of birds at a site i is a function of *latitude* and *longitude* (of the site) plus an error component ε_i that is assumed to be *independently* normally distributed with a mean 0 and a variance σ^2. A log-transformation of counts was used to ensure homogeneity. Also latitude and longitude were centred by subtracting the mean latitude and longitude, respectively. This reduces collinearity. The full output of this linear regression model is not given here. Both latitude ($p < 0.001$) and longitude ($p = 0.029$) were significantly different from 0 at the 5% level. The R^2 and AIC were 14% and 1072.56, respectively. Moran's I test applied to the residuals showed that there was significant spatial auto-correlation. The test statistics was $I = 0.834$ ($p < 0.0001$), so the p-values for the t- and F-statistics produced by this linear regression model were unreliable.

We also applied an additive model on the bird radar data in which longitude and latitude were used as smoothers (Chapter 7). The R^2 for this additive model was 0.73, and the AIC was 724.35. The additive model is preferred over the linear regression model because its AIC is lower. ANOVA tests indicated that both smoothers were highly significant (for both: $p < 0.0001$). The Moran's I test statistic for residuals is 0.544 ($p < 0.0001$); so, in this case, we cannot trust the p-values for the t- and F-statistics either.

The linear regression and additive models indicate violation of the independence assumptions. One possible approach to incorporate the spatial dependence within the model is to use the simultaneous autoregressive model (SAR) or the

spatial moving average model (SMA) that are described in Chapter 18. For the bird count data, the SAR model specification is:

$$Counts_i = \alpha + \rho \sum_j w_{ij} Counts_j + \beta_1 Latitude_i + \beta_2 Longitude_i + \varepsilon_i$$

The SAR models bird counts at site i as a function of latitude, longitude and bird counts at nearby sites. The output of the SAR model (not shown here) showed that latitude and longitude were not significant at the 5% significance level, and ρ was 0.94 ($p < 0.001$). The large value of ρ confirms the spatial dependence. The AIC was 535.22. The SMA model is given by

$$Counts_i = \alpha + \beta_1 Latitude_i + \beta_2 Longitude_i + u_i$$
$$u_i = \lambda \sum_j w_{ij} u_j + \varepsilon_i$$

In the SMA, bird counts are modelled as a function of latitude, longitude, and error components of nearby sites. In the SMA model, latitude was significantly different from zero ($p = 0.026$), λ was 0.95 ($p < 0.001$) and the AIC was 530.8. The AICs for the linear regression, additive model, SAR, and SMA model are, respectively, 1072.56, 724.35, 535.22 and 530.87. Therefore, the SMA model fits the data best.

Analysis of spatial correlation

Two assumptions made for most geostatistical modelling tools are normality and stationarity. Before we proceed with the geostatistical analysis of bird counts, these assumptions must be verified using data for this area. The spatial trend (if present) in the count data can be subtracted from the data in order to remove any dependence of site locations or other covariates. This means that the following model (in words) is applied:

$$\text{Counts}_i = F(\text{Latitude}_i, \text{Longitude}_i, \text{Other Covariates}_i) + \text{residuals}_i \qquad (19.4)$$

F stands for 'Function of' and can be modelled as in linear regression or additive modelling, among others. For the bird radar data, we only have latitude and longitude and there are no other covariates. In geostatistics, F is confusingly called the trend. The entire model in equation (19.4) is sometimes called a trend model. Earlier in this section, we applied two trend models to relate the spatial coordinates of sites with bird counts. These were the linear regression (LM) and additive model (AM). We will now use the residuals of these models as a variable in the geostatistical analysis and modelling. Hence, the normality and stationarity assumptions are for the residuals and not the raw data. QQ-plot, histograms, and graphs with fitted values versus residuals were made to verify normality and homogeneity for the residuals of the LM and AM models. These plots (not shown here) show a clear violation of normality and homogeneity for the regression residuals and to a lesser extend also for the AM residuals.

Possible ways to overcome violation of normality and homogeneity were discussed in Chapters 4 and 5 and are (i) to apply a transformation on the bird count

data, (ii) to model birds counts using a Poisson distribution, (iii) add more covariates to the models (or interactions), and (iv) apply smoothing methods, among others. In this example, we applied a log-transformation on the count data. We chose a log transformation and not a square root because of the results of the log transformation produced better results in the QQ-plots. Histograms and QQ-plots of residuals after the log-transformation (not shown here) of the bird count data have a 'more' normal distribution than they had without the transformation.

The stationarity assumption for the residuals is tested with *h*-scatterplots. Residuals of the linear model show strong departure from stationarity; i.e., a swarm of points is above the diagonal line (Figure 19.17). The figure shows a strong drift in the data and indicates that we have underestimated the trend. So, we need to eliminate the trend from the data. One option to do this is to use the additive model, which is more flexible than the linear model.

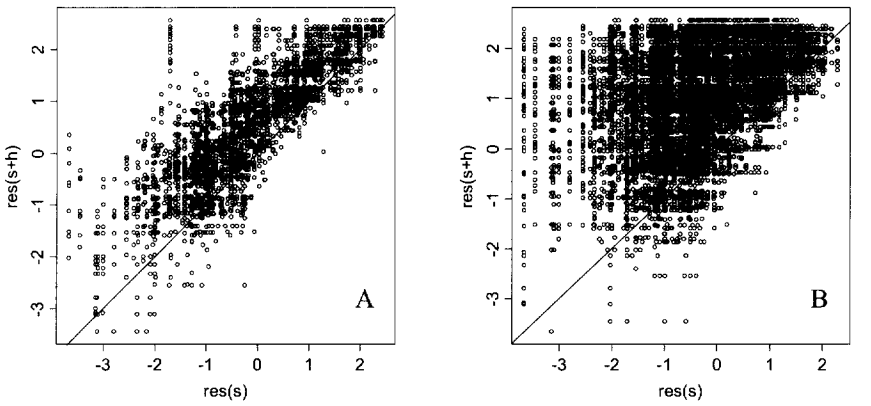

Figure 19.17. Omnidirectional *h*-scatterplots (all directions of lag ***h*** are used). A: Plot for LM residuals for lag h = 1300 m. B: Plot for LM residuals for lag h = 3900 m. The value of the residual at **s** is along the horizontal axis, and residuals of all points that are ***h*** units away from **s** are plotted along the vertical axis. This is done for each site in the study area. The bird data were log-transformed.

h-Scatterplots for AM residuals (Figure 19.18) look better than for the LM residuals. There is no strong butterfly effect. Although we have points that are far from the diagonal line, there are only a few of them. A drift is present, but it is not as strong as for the LM residuals. *h*-Scatterplots for different directions (Figure 19.19) confirm that the drift is weak. Therefore, we proceed by assuming that the residuals are stationary and normally distributed.

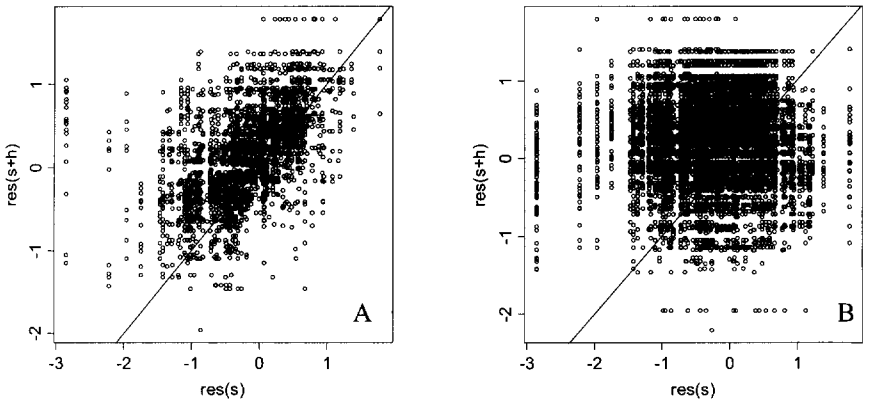

Figure 19.18. Omnidirectional h-scatterplots (All directions of lag $\boldsymbol{h}$ are used). A: Plot for AM residuals for lag h = 1300 m. B: Plot for AM residuals for lag h = 3900 m. The bird data were log-transformed.

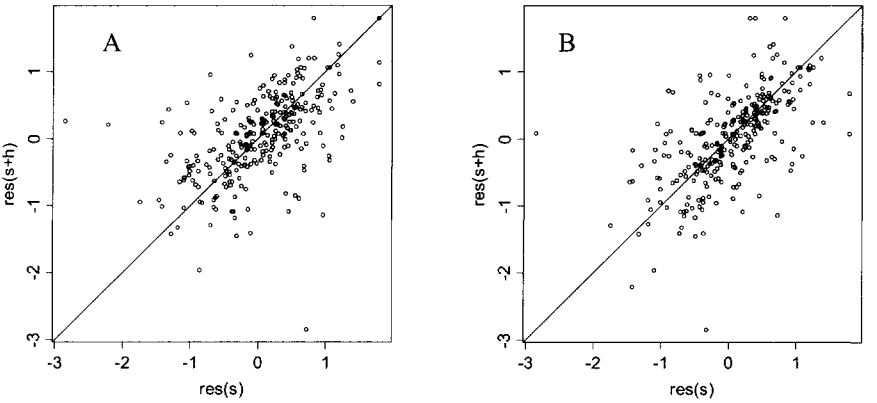

Figure 19.19. Directional h-scatterplots (lag $\boldsymbol{h}$ is taken only along selected directions). A: Plot for AM residuals for lag h= 1300 m at 90° direction. B: Plot for AM residuals for lag h = 1300 m at 0° direction. The value of the residual at **s** is along the horizontal axis, and residuals of all points that are $\boldsymbol{h}$ units away from **s** are plotted along the vertical axis. The bird data were log-transformed.

In the previous paragraph the following model (in words) was fitted on the log-transformed bird data (the value of 1 was added to each observation to avoid problems with logs of zero):

$$\text{Birds} = \text{intercept} + f(\text{Latitude}) + f(\text{Longitude}) + \text{noise}$$

where $f()$ was either a smoothing function (additive modelling) or a parametric term (linear regression). The noise component is assumed to be independently normally distributed. If $f()$ is a smoothing function, the model above is called the AM trend model. We can refine the AM trend model by taking into account the spatial auto-correlation directly within the error structure of the noise as this produces better smoothing curves.

In Chapter 18 we introduced linear regression models with correlated errors, LM(ce), to do this, but the same can be done for additive models. We call it an additive model with correlated error AM(ce). The estimation process makes use of generalised least squares (GLS), and details can be found in Pinheiro and Bates (2000). Within the GLS, we need to specify the model for the correlation structure and we tried two different correlation structures. The first model with a spherical residual auto-correlation structure gave an AIC of 553, and the second model with an exponential correlation structure has AIC = 556. Therefore, we decided to use the spherical correlation structure because it had a lower AIC.

Recall that the spherical correlation structure is fully specified once we know the range and sill. Numerical optimisation gave a range of 3764 m and nugget effect (nugget/sill) ≈ 0.08.

Due to software implementations, AM(ce) models can only be used on spatial data with isotropic auto-correlation processes, so we need to explore the residuals of the trend model and verify that we indeed have isotropy. We use a variogram surface to determine the presence and directions of anisotropy; see Section 18.3 for an explanation. A variogram surface for the AM trend residuals is plotted in Figure 19.20. Anisotropy is present, but it is not strongly pronounced. Until the ranges of 4000–5000 m the direction with the largest auto-correlation observed is approximately 135°.

We can also explore anisotropy using a directional variogram. It shows variogram values for lags taken only in specified directions. Directional variograms are presented for residuals of the AM trend model in Figure 19.20. Usually four directions are tested: the first along the direction with maximum spatial auto-correlation detected by the variogram surface (in this example, at 135°); the second, perpendicular to the maximum (at 45°); and the third and fourth at intermediate directions (at 0° and 90° in this example). Isotropy would be confirmed if all directional variograms are the same. If they are not, our models possess anisotropy; and we need to look at the values of the nugget, sill, and range of the auto-correlation process.

The directional variograms in Figure 19.21 are not the same, and show strong evidence for zonal anisotropy. Zonal anisotropy means that the sill varies with direction (recall that in geometrical anisotropy sills of all variograms are the same but are reached at different ranges). It is near 0.9 in the 45° direction and near 0.7 in the 135° direction. This may also be caused by the weak drift indicated in the h-scatterplots above.

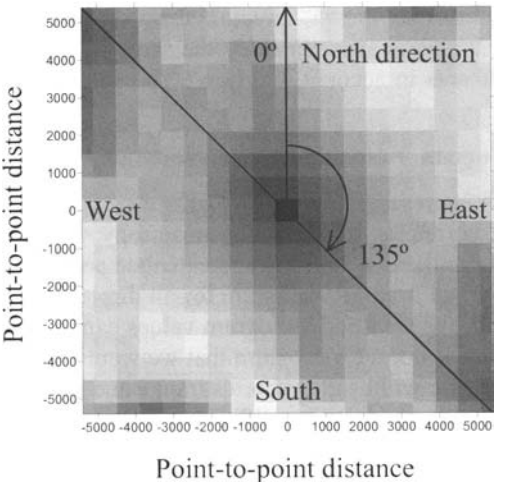

Figure 19.20. Variogram surface. Darker cells indicate lower variogram values. The direction with the largest auto-correlation observed is approximately 135°.

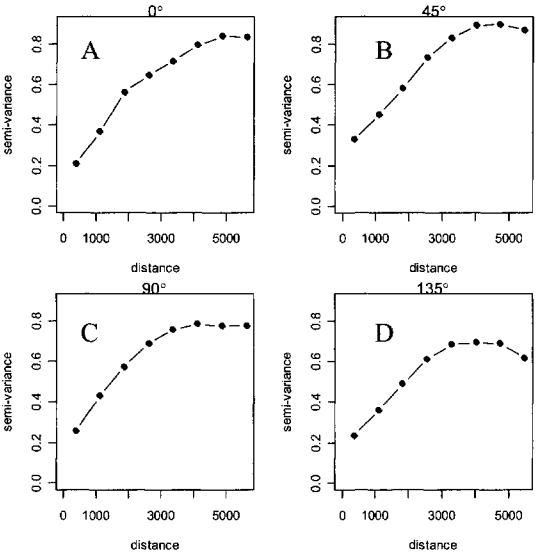

Figure 19.21. Directional variograms for AM residuals for lag distances at four directions selected. A: 0°, B: 45°, C: 90°, D: 135°. Differences may be caused due to different flight paths of the birds.

Figure 19.21 shows two problems; the range differs per direction (shorter range for the 90° and longer for the 0° direction), and the sill differs per direction (ap-

proximately 0.2 higher for the 45° direction than for the 135° direction). The good news is that the nugget is approximately the same in each direction. We need to take the two problems in account, but how? The solution is to use the following mechanism:

Anisotropy variogram = omnidirectional variogram + deviation in direction *i*

The variogram is presented as the sum of two components: An omnidirectional variogram and a correction allowing for anisotropy. For the first component we will just pretend that there is isotropy and calculate an omnidirectional variogram. This can be seen as an average variogram for all directions. Because we will add a second component, the resulting variogram values can only increase. The omnidirectional variogram is not the variogram that we would obtain by clumping all the directions. Instead, we will use the shortest range and lowest sill of the directional variograms to guarantee a positive difference between the experimental and the omnidirectional variograms. Once we have the experimental directional variograms, we can easily calculate the difference between the experimental variogram in direction *i*, and fit an anisotropic variogram on the difference (component 2). It is the deviation in direction *i*. The process is illustrated for the bird radar data in Figure 19.22. Hollow dots represent experimental variograms in each of the four directions. The dotted line in each panel is the omnidirectional variogram model. Note that it has the same shape for all directions. The heavy dashed line is the anisotropic component, and it models the difference between these two. The solid line is a sum of two components, and it fits the experimental variogram reasonably well in each of the four directions. A variogram was applied on the difference. Full technical details can be found in Chiles and Delfiner (1999) or Cressie (1993).

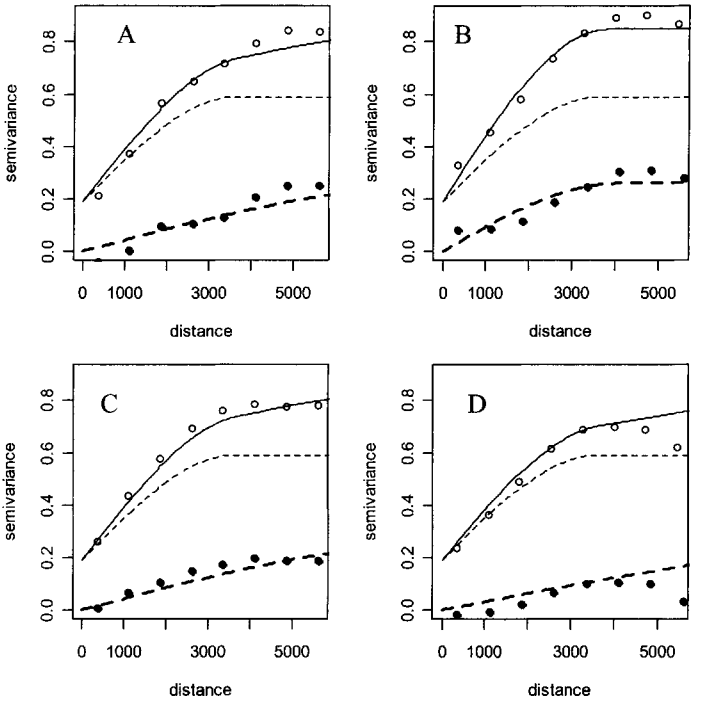

Figure 19.22. Hollow dots represent empirical variograms at the four directions selected. A: 0°, B: 45°, C: 90°, D: 135°. Dotted lines represent the behaviour of the isotropic (general) variogram component. Filled dots represent the difference between empirical variogram and isotropic general variogram component. Dashed lines represent the behaviour of anisotropic variogram component. Solid lines represent the behaviour of the final variogram model that is the sum of isotropic and anisotropic components.

Kriging

We will now incorporate the resulting variogram model into a kriging analysis that interpolates the sampled bird counts into unsampled areas and allows us to predict how birds are distributed over this area.

Kriging combines information from the trend models and the variogram model of trend residuals to predict the spatial distribution of a variable of interest. In this case, we want to predict how birds are spatially distributed over the entire area when only a minimal number of actual bird counts have been done. Figure 19.22 shows how the bird-count information differs using just the AM trend model (Figure 19.23-A), the predicted trend residuals from the AM model after kriging (Figure 19.23-B), and the combined results from the trend and residuals predictions (Figure 19.23-C). Because we used the log-transformed bird counts, we also included a prediction based on the back-transformed values on the original count

scale (Figure 19.23-D). Each map shows different characteristics of the bird count data, and each analysis refines our predictions about how birds may be distributed in this area.

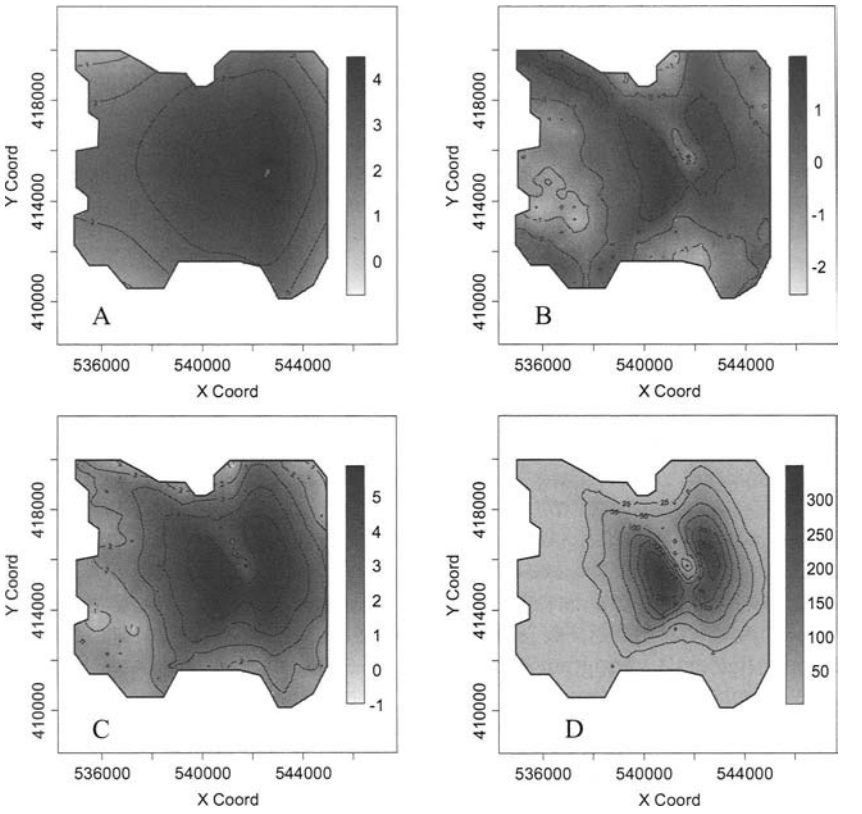

Figure 19.23. Maps of predicted bird counts. A: Prediction based on AM trend model. B: Prediction based on kriged AM residuals. C: Prediction based on log(bird counts) as a sum of trend and residuals maps. D: Prediction obtained by back-transformation of log-counts in map C. A–C: contour level step 1. D: contour level step 50.