Excerpt from Chapter 13 of:

Legendre, P. and L. Legendre. 1998. Numerical ecology, 2nd English edition. Elsevier Science BV, Amsterdam. xv + 853 pages.

13.2 Maps

The most basic step in spatial pattern analysis is the production of maps displaying the spatial distributions of values of the variable(s) of interest. Furthermore, maps are essential to help interpret spatial structure functions (Section 13.1).

Several methods are available in mapping programs. The final product of modern computer programs may be a contour map, a mesh map (such as Figs. 13.13b and 13.16b), a raised contour map, a shaded relief map, and so on. The present Section is not concerned with the graphic representation of maps but instead with the way the mapped values are obtained. Spatial interpolation methods have been reviewed by Lam (1983).

Geographic information systems (GIS) are widely used nowadays, especially by geographers, to manage complex data corresponding to points, lines and surfaces in space. The present Section is not an introduction to these complex systems. It only aims at presenting the most widespread methods for mapping univariate data (i.e. a single variable *y*). The spatial analysis of multivariate data (multivariate matrix **Y**) is deferred to Sections 13.3 to 13.5.

Beware of non-additive variables such as pH, logarithms of counts of organisms, diversity measures, and the like (Subsection 1.4.2). Maps of such variables, produced by trend-surface analysis or interpolation methods, should be interpreted with caution; the interpolated values only make sense by reference to sampling units of the same size as those used in the original sampling design. Block kriging (Subsection 2) for blocks representing surfaces or volumes that differ from the grain of the observed data simply does not make sense for non-additive variables.

1 — Trend-surface analysis

Trend-surface analysis is the oldest method for producing smoothed maps. In this method, estimates of the variable at given locations are not obtained by interpolation, as in the methods presented in Subsection 2, but through a regression model calibrated over the entire study area.

In 1914, Student proposed to express observed values as a polynomial function of time and mentioned that it could be done for spatial data as well. This is also one of the most powerful tools of spatial pattern analysis, and certainly the easiest to use. The objective is to express a response variable *y* as a nonlinear function of the geographic coordinates X and Y of the sampling sites where the variable was observed:

$$y = f(\mathrm{X},\ \mathrm{Y})$$

In most cases, a polynomial of X and Y with cross-product terms is used; trend-surface analysis is then an application of polynomial regression (Subsection 10.3.4) to spatially-distributed data. For example a relatively complex, but smooth surface might be fitted to a variable using a third-order polynomial with 10 parameters (b_0 to b_9):

$$\hat{y} = f(\mathrm{X}, \mathrm{Y}) = b_0 + b_1\mathrm{X} + b_2\mathrm{Y} + b_3\mathrm{X}^2 + b_4\mathrm{XY} + b_5\mathrm{Y}^2 + b_6\mathrm{X}^3 + b_7\mathrm{X}^2\mathrm{Y} + b_8\mathrm{XY}^2 + b_9\mathrm{Y}^3 \quad \textbf{(13.18)}$$

Note the distinction between the response variable *y*, which may represent a physical or biological variable, and the Cartesian geographic coordinate Y. Using polynomial regression, trend-surface analysis produces an equation which is linear in its parameters, although the response of *y* to the explanatory variables in matrix **X** = [X, Y] may be nonlinear. If variables *y*, X and Y have been centred on their respective means prior to model fitting, the model has an intercept of 0 by construct; therefore parameter b_0 does not have to be fitted and it can be removed from the model.

Numerical example. The data from Table 10.5 are used here to illustrate the method of trend-surface analysis. The dependent variable of the analysis, *y*, is Ma, which was the log-

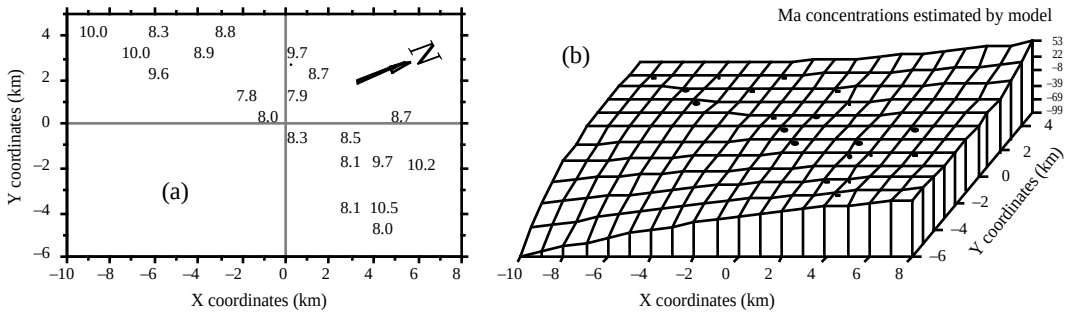

Figure 13.13 Variable Ma (log-transformed concentrations of aerobic heterotrophic bacteria growing on marine agar at salinity of 34 psu) at 20 sites in the Thau lagoon on 25 October 1988. (a) Map of the sampling sites with respect to arbitrary geographic coordinates X and Y. The observed values of Ma, from Table 10.5, are also shown. The *N* arrow points to the north. (b) Trend-surface map; the vertical axis gives the values of Ma estimated by the polynomial regression equation. Dots represent the sampling sites.

transformed (ln(x + 1)) concentrations of aerobic heterotrophic bacteria growing on marine agar at salinity of 34 psu. The explanatory variables are the X and Y geographic coordinates of the sampling sites (Fig. 13.13a). The steps of the calculations are the following:

• Centre the geographic coordinates on their respective means. The reason for centring X and Y is given in Subsection 10.3.4; the amount of variation explained by a trend-surface equation is not changed by a translation (centring) of the spatial coordinates across the map.

• Determine the order of the polynomial equation to be used. A first-degree regression equation of Ma as a function of the geographic coordinates X and Y alone would only represent the linear variation of Ma with respect to X and Y; in other words, a flat surface, possibly sloping with respect to X, Y, or both. With the present data, the first-degree regression equation was not significant (R^2 = 0.02), meaning that there was no significant linear geographic trend to be described in the data. A regression equation incorporating the second-degree monomials (X^2, XY and Y^2) together with X and Y would be appropriate to model a surface presenting a single large bump or trough. Again, this did not seem to be the case with the present data since the second-degree equation was not significant (R^2 = 0.39). An equation incorporating the third-degree, fourth-degree, etc. terms would be able to model structures of increasing complexity and refinement. The cost, however, is a loss of degrees of freedom for every new monomial in the equation; trend-surface analysis using high-order equations thus requires a large number of observed sampling sites. In the present example, the polynomial was limited to the third degree, for a total of 9 terms; this is a large number of terms, considering that the data only contained 20 sampling sites.

• Calculate the various terms of the third-degree polynomial, by combining the variables X and Y as follows: X^2, X×Y, Y^2, X^3, X^2×Y, X×Y^2, Y^3.

• Compute the multiple regression equation. The model obtained using all 9 regressors had $R^2 = 0.87$, but several of the partial regression coefficients were not significant.

• Remove nonsignificant terms. The linear terms may be important to express a linear gradient; the quadratic and cubic terms may be important to model more complex surfaces. Nonsignificant terms should not be left in the model, however, except when they are required for comparison purpose. Nonsignificant terms were removed one by one (backward elimination, Subsection 10.3.3) until all terms (monomials) in the polynomial equation were significant. The resulting trend-surface equation was highly significant ($R^2 = 0.81$, $p < 0.0001$):

$$\hat{y} = 8.13 - 0.16\ XY - 0.09\ Y^2 + 0.04\ X^2Y + 0.14\ XY^2 + 0.10\ Y^3$$

Remember, however, that tests of significance are too liberal with autocorrelated data, due to the non-independence of residuals (Subsection 1.1.1).

• Lay out a regular grid of points (X', Y') and, using the regression equation, compute forecasted values ($\hat{y}'$) for these points. Plot a map (Fig. 13.13b) using the file with (X', Y', and $\hat{y}'$). Values estimated by a trend-surface equation at the observed study sites do not coincide with the values observed at these sites; regression is not an exact interpolator, contrary to kriging (Subsection 2).

Different features could be displayed by rotating the Figure. The orientation chosen in Fig. 13.13b does not clearly show that the values along the long axis of the Thau lagoon are smaller near the centre than at the ends. It displays, however, the wavy structure of the data from the lower left-hand to the upper right-hand corner, which is roughly the south-to-north direction. The Figure also clearly indicates that one should refrain from interpreting extrapolated data values, i.e. values located outside the area that has actually been sampled. In the present example, the values forecasted by the model in the lower left-hand and the upper right-hand corners (–99 and +53, respectively) are meaningless for log bacterial concentrations. Within the area where real data are available, however, the trend-surface model provides a good visual representation of the broad-scale spatial variation of the response variable.

Examination of the residuals is essential to make sure that the model is not missing some salient feature of the data. If the trend-surface model has extracted all the spatially-structured variation of the data, given the scale of the study, residuals should look random when plotted on a map and a correlogram of residuals should be non-significant. With the present data, residuals were small and did not display any recognizable spatial pattern.

A cubic trend-surface model is often appropriate with ecological data. Consider an ecological phenomenon which starts at the mean value of the response variable *y* at the left-hand border of the sampled area, increases to a maximum, then goes down to a minimum, and comes back to the mean value at the right-hand border. The amount of space required for the phenomenon to complete a full cycle — whatever the shape it may take — is its extent (Section 13.0). Using trend-surface analysis, such a phenomenon would be correctly modelled by a third-degree trend surface equation. A polynomial equation is a more flexible mathematical model than sines or cosines, in that it does not require symmetry or strict periodicity.

The degree of the polynomial which is appropriate to model a phenomenon is predictable to a certain extent. If the extent is of the same order as the size of the study

area (say, in the X direction), the phenomenon will be correctly modelled by a polynomial of degree 3 which has two extreme values, a minimum and a maximum. If the extent is larger than the study area, a polynomial of degree less than 3 is sufficient; degree 2 if there is only one maximum, or one minimum, in the sampling window; and degree 1 if the study area is limited to the increasing, or decreasing, portion of the phenomenon. Conversely, if the scale of the phenomenon controlling the variable is smaller than the study area, more than two extreme values (minima and maxima) will be found, and a polynomial of order larger than 3 is required to model it correctly. The same reasoning applies to the X and Y directions when using a polynomial combining the X and Y geographic coordinates. So, using a polynomial of degree 3 acts as a filter: it is a way of looking for phenomena that are of the same extent, or larger, than the study area.

An assumption must be made when using the method of trend-surface analysis: that all observations form a single statistical population, subjected to one and the same generating process, and can consequently be modelled using a single polynomial equation of the geographic coordinates. Evidence to that effect may be available prior to the analysis. When this is not the case, the hypothesis of homogeneity may be supported by examining the regression residuals (Subsection 10.3.1). When there are indications that values in different regions of the geographic space obey different processes (e.g. different geology, action of currents or wind, or influence of other physical variables), the study area should be divided into regions, to be modelled by separate trend-surface equations.

Polynomial regression, used in the numerical example above, is a good first approach to fitting a model to a surface when the shape to be modelled is unknown, or known to be simple. In some instances, however, it may not provide a good fit to the data; trend-surface analysis must then be conducted using nonlinear regression (Subsection 10.3.6), which requires that an appropriate numerical model be provided to the estimation program. Consider the example of the effect of some human-generated environmental disturbance at a site, the indicator variable being the number of species. The response, in this case, is expected to be stronger near the impacted site, tapering off as one gets farther away from it. Assume that data were collected along a transect (a single geographic coordinate X) and that the impacted site is near the centre of the transect. A polynomial equation would not be appropriate to model an inverse-squared-distance diffusion process (Fig. 13.14a), whereas an equation of the form:

$$\hat{y} = b_0 + \frac{b_1 X^2}{b_2 X^2 + 1}$$

would provide a much better fit (Fig. 13.14b). The minimum of this equation is b_0; it is obtained when X = 0. The maximum, b_1/b_2, is reached asymptotically as X becomes large in either the positive or negative direction. For data collected in different directions around the impacted site, a nonlinear trend-surface equation with similar properties would be of the form:

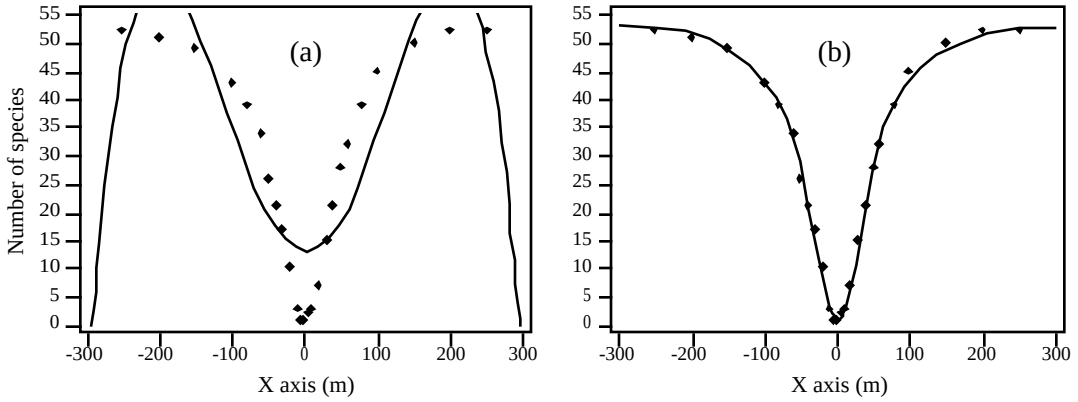

Figure 13.14 (a) Artificial data representing the number of species around the site of an environmental disturbance (located at X = 0) are not well-fitted by a 4th-order polynomial equation of the X coordinates ($R^2 = 0.7801$). (b) They are well-fitted by the following inverse-squared-distance diffusion equation: $\hat{y} = 1 + [0.0213X^2 / (0.0004X^2 + 1)]$ ($R^2 = 0.9975$).

$$\hat{y} = b_0 + \frac{b_1X^2 + b_2Y^2}{b_3X^2 + b_4Y^2 + 1}$$

where X and Y are the coordinates of the sites in geographic space.

Trend-surface analysis is appropriate for describing broad-scale spatial trends in data. It does not produce accurate fine-grained maps of the spatial variation of a variable, however. Other methods described in this Chapter may prove useful to model variation at finer scales. In some studies, the broad-scale trend itself is of interest; this is the case in the numerical example above and in Ecological application 13.2. In other instances, and especially in studies that cover large geographic expanses, the broad-scale trend may be already known and understood; students of geographic variation patterns may want to conduct analyses on detrended data, i.e. data from which the broad-scale trend has been removed. Detrending a variable may be achieved by computing the residuals from a trend-surface equation of sufficient order, as in time-series analysis (Section 12.2).

If there is replication at each point, it is possible to perform a test of goodness-of-fit of a trend-surface model (Draper and Smith, 1981; Legendre & McArdle, 1997). By comparing the observed error mean square after fitting the trend surface to the error mean square estimated by the among-replicate within-location variation, one can test if the model fits the data properly. The among-replicate within-location variation is computed from the deviations from the means at the various locations; it is actually the

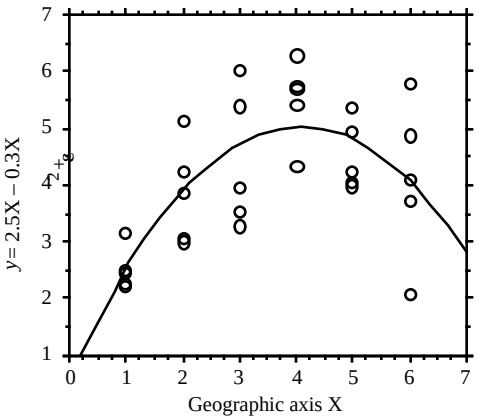

Figure 13.15 Artificial data representing sampling along a geographic axis X with 5 replicates at each site; $n = 30$. The F test of goodness-of-fit indicates that the trend-surface equation $\hat{y} = 0.562 + 2.184X - 0.267X^2$ ($R^2 = 0.4899$) fits the data properly.

residual mean square of an ANOVA among locations. These error mean squares are not much different if the trend surface goes through the expected values at the various locations, so that the F ratio of the two mean squares is not significant. If, on the contrary, the fitted surface does not follow the major features of the variation among locations, the deviations of the data from the fitted trend-surface values are likely to be larger than expected from our knowledge of the sampling error; the F statistic is then significantly larger than 1, indicating that the trend surface is misrepresenting the variation among locations.

Numerical example. Consider the artificial data in Fig. 13.15. Variable X represents a geographic axis along which sampling has taken place at 6 sites with replication. Variable y was constructed using equation $y = 2.5X - 0.3X^2 + \varepsilon$, where ε is a random standard normal deviate [N(0, 1)]. A quadratic trend-surface model of X was fitted to the data. The residual mean square, or "error mean square after fitting the trend surface", was $MS_1 = 0.84909$ ($\nu = 27$). An analysis of variance was conducted on y using the grouping into 6 sites as the classification criterion. The residual mean square obtained from the ANOVA was $MS_2 = 0.87199$ ($\nu = 24$). The ratio of these two mean squares gave an F statistic:

$$F = \frac{MS_1}{MS_2} = \frac{0.84909}{0.87199} = 0.97374$$

which was tested against $F_{\alpha=0.05\,(27,\ 24)} = 1.959$. The F statistic was not significantly different from 1 ($p = 0.5308$), which indicated that the model fitted the data properly.

The trend-surface analysis was recomputed using a linear model of X. The model obtained was $\hat{y} = 3.052 + 0.316X$ ($R^2 = 0.1941$). MS_1 in this case was 1.29358 ($\nu = 28$). The F ratio

$MS_1/MS_2 = 1.29358/0.87199 = 1.48348$. The reference value was $F_{0.05(28, 24)} = 1.952$. The probability associated with the F ratio, $p = 0.1651$, indicated that this model still fitted the data, which were constructed to contain a linear term (2.5X in the construction equation) as well as a quadratic trend (term $-0.3\,X^2$), but the fit was poorer than with the quadratic polynomial model which was capable of accounting for both the linear and quadratic trends.

This numerical example shows that trend-surface analysis may be applied to data collected along a transect; the "trend surface" is one-dimensional in that case. The numerical example at the end of Subsection 10.3.4 is another example of a trend-surface analysis of a dependent variable, salinity, with respect to a single geographic axis (Fig. 10.9). Trend-surface analysis may also be used to model data in three-dimensional geographic space (geographic coordinates X, Y and Z, where Z is either altitude or depth) or with one of the dimensions representing time. Section 13.5 will show how the analysis may be extended to a multivariate dependent data matrix **Y**.

Haining (1987) described alternative methods for estimating the parameters of a trend-surface model when the residuals are spatially autocorrelated; in that case, least-squares estimation of the parameters is inefficient and standard errors as well as tests of significance are biased. Haining's methods allow one to recognize three components of spatial variation corresponding to the site, local, and regional scales.

Ecological application 13.2

A survey was conducted at 200 locations within a fairly homogeneous 12.5 ha rectangular sandflat area in Manukau Harbour, New Zealand, to identify factors that control the spatial distributions of the two dominant bivalves, *Macomona liliana* Iredale and *Austrovenus stutchburyi* (Gray), and to look for evidence of adult-juvenile interactions within and between species. Results are reported in Legendre *et al.* (1997). Most of the broad-scale spatial structure detected in the bivalve counts (two species, several size classes) was explained by the physical and biological variables. Results of principal component analysis and spatial regression modelling suggested that different factors controlled the spatial distributions of adults and juveniles. Larger size classes of both species displayed significant spatial structures, with physical variables explaining some but not all of this variation; the spatial patterns of the two species differed, though. Smaller organisms were less strongly spatially structured; virtually all of their spatial structure was explained by physical variables.

Highly significant trend-surface equations were found for all bivalve species and size classes (log-transformed data), indicating that the spatial distributions of the organisms were not random, but highly organised at the scale of the study site. The trend-surface models for smaller animals had much smaller coefficients of determination (10-20%) than for larger animals (30-55%). The best models, i.e. the models with the highest coefficients of determination (R^2), were for the *Macomona* > 15 mm and *Austrovenus* > 10 mm. The coefficients of determination were consistently higher for *Austrovenus* than for *Macomona*, despite the fact that *Macomona* were usually far more numerous than *Austrovenus*. A map illustrating the trend-surface equation is presented for the largest *Macomona* size class (Fig. 13.16); the field counts are also given for comparison.

(a)

35 41 37 48 52 68 64 51 52 19 50 48 46 50 40 33 27 29 32 39
22 47 46 33 51 71 54 71 44 56 63 50 30 49 46 41 43 40 29 32
45 66 58 60 27 59 50 52 41 40 54 39 41 46 47 19 24 30 20 28
41 56 33 60 59 60 53 52 26 39 31 33 29 54 37 18 36 42 27 29
48 36 38 34 41 55 42 67 53 57 34 34 42 33 32 46 33 18 27 32
41 32 41 35 48 52 40 56 40 44 41 43 44 41 43 32 22 34 30 32
17 29 50 67 35 52 28 44 45 28 30 35 30 28 35 37 20 28 50 31
27 43 38 44 29 44 35 39 44 38 40 24 37 43 35 31 38 48 50 37
27 42 36 36 29 46 39 38 29 37 38 35 35 38 25 38 39 54 50 44
52 41 32 41 42 28 47 22 24 17 37 31 38 29 41 38 36 46 38 52

(b)

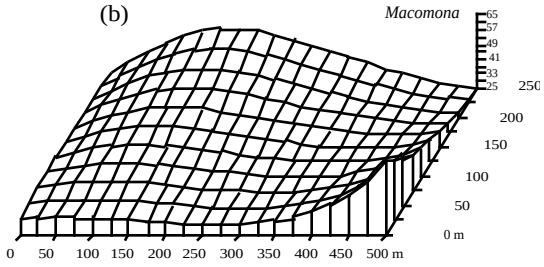

Figure 13.16 *Macomona* > 15 mm at 200 sites in Manukau Harbour, New Zealand, on 22 January 1994. (a) Actual counts at sampling sites in 200 regular grid cells; in the field, sites were not perfectly equispaced. (b) Map of the trend-surface equation explaining 32% of the spatial variation in the data. The values estimated from the trend-surface equation (log-transformed data) were back-transformed to raw counts before plotting. Modified from Legendre *et al*. (1997).

2 — Interpolated maps

In this family of methods, the value of the variable at a point location on a map is estimated by local interpolation, using only the observations available in the vicinity of the point of interest. In this respect, interpolation mapping differs from trend surface analysis (Subsection 1), where estimates of the variable at given locations were not obtained by interpolation, as in the present Subsection, but through a statistical model calibrated over the entire study area. Fig. 13.17 illustrates the principle of interpolation mapping. A regular grid of nodes (Fig. 13.17c) is defined over the area that contains the study sites $Ø_i$ (Fig. 13.17a, b). Interpolation assigns a value to each point of that grid. This is the single most important step in mapping. Following that, results may be represented in the form of contours (Fig. 13.17d) with or without colours or shades, or three-dimensional constructs such as Fig. 13.16b.

Assigning a value to each grid node may be done in different ways. Different interpolation methods may produce maps that look different; this is also the case when using different parameters with a same method (e.g. different exponents in inverse-distance weighting).

The most simple rule would be to give, to each node of the grid, the value of the observation which is the closest to it. The end result is a division of the map into Voronoï polygons (Subsection 13.3.1) displaying a “zone of influence” drawn around each observation. Another simple solution consists in dividing the map into Delaunay triangles (Subsection 13.3.1). There is an observed value y_i at each site $Ø_i$. A triangular portion of plane, adjusted to the points $Ø_i$ that form the apices (corners) of a Delaunay triangle, provides interpolated values for all points lying within the triangle. Maps obtained using these solutions are shown in Chapter 11 of Isaaks & Srivastava (1989).

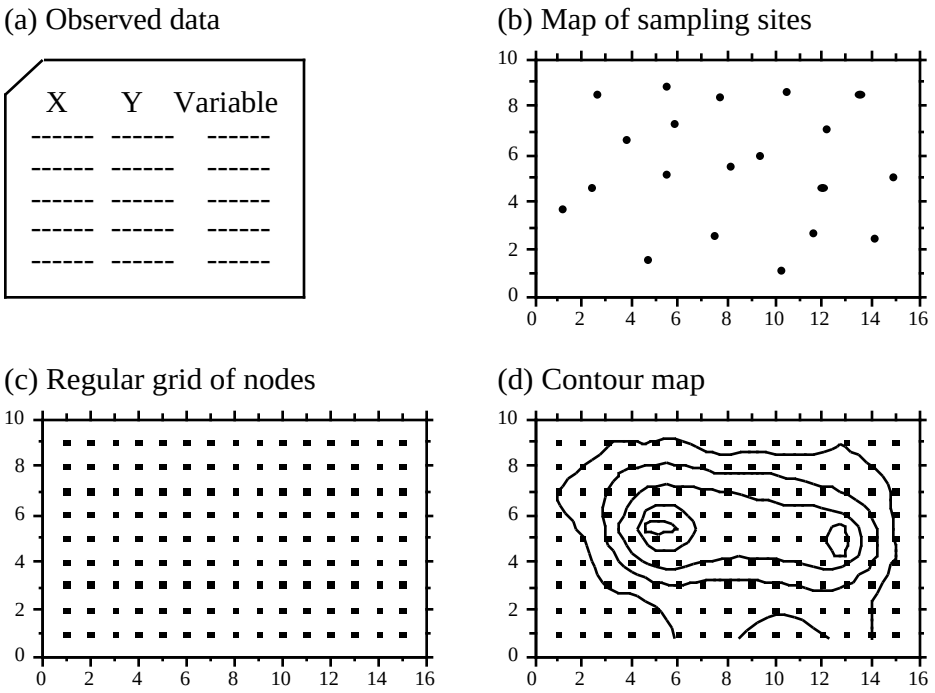

Figure 13.17 Summary of the interpolation procedure.

Alternatively, one may draw a "search circle" (or an ellipsoid for anisotropic data) around each grid node (Fig. 13.18). The radius of the circle may be determined in either of two ways. (1) One may fix a minimum number of observed points that must be included in the interpolation for each grid node; or (2) one may use the "distance of influence of the process" found by correlogram or variogram analysis (Section 13.1). The estimation procedure is repeated for each node of the grid. Several methods of interpolation may be used.

- Mean — Consider all the observed study sites found within the circle; assign the mean of these values to the grid node. This method does not produce smooth maps; discontinuities in neighbouring grid node values occur as observed points move in or out of the search circle.

- Inverse-distance weighting — Consider the observation sites found within the circle and calculate a weighted mean value, using the formula:

$$\hat{y}_{\text{Node}} = \sum_i w_i y_i \qquad \textbf{(13.19)}$$

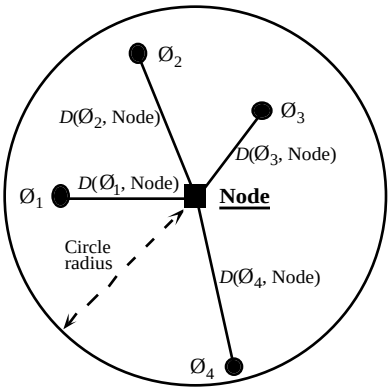

Figure 13.18 To estimate the value at a grid node (square), draw a search circle around it and consider the observed points ($Ø_i$) found within the circle. Observed points are separated from the node by distances $D(Ø_i, \text{Node})$.

where y_i is the value observed at point $Ø_i$ and weight w_i is the inverse of the distance (D) from point $Ø_i$ to the grid node to be estimated. The inverse distances, to some power k, are scaled by the sum of the weights for all points $Ø_i$ in the estimation, so as to produce values that are consistent with the values observed at points $Ø_i$ (unbiasedness condition):

$$w_i = \left(\frac{1}{D(Ø_i, \text{Node})^k} \right) \Big/ \sum_i \frac{1}{D(Ø_i, \text{Node})^k} \qquad \textbf{(13.20)}$$

A commonly-used exponent is $k = 2$. This corresponds, for instance, to the decrease in energy of waves dispersing across a two-dimensional surface. The greater the value of k, the less influence distant data points have on the value assigned to the grid node. This method produces smooth values over the grid of nodes. The range of estimated values is smaller than the range of observed data so that, contrary to trend-surface analysis (Fig. 13.13b), inverse-distance weighting does not produce meaningless values in the parts of the map beyond the area that was actually sampled. When the observation sites $Ø_i$ do not form a regular or nearly regular grid, however, this interpolation method may generate features in maps that have little to do with reality. As a consequence, inverse-distance weighting is not recommended in that situation.

• Weighted polynomial fitting — In this method, a trend-surface equation (Subsection 13.2.1) is adjusted to the observed data points within the search circle, weighting each observation $Ø_i$ by the inverse of its distance (using some appropriate power k) to the grid node to be estimated. A first or second-order polynomial equation is usually used. z_{Node} is taken to be the value estimated by the polynomial equation for the coordinates of the grid node. This method suffers from the same problem as inverse distance weighting with respect to observation sites $Ø_i$ that do not form a regular or nearly regular grid of points.

Kriging

• Kriging — This is the mapping tool in the toolbox of geostatisticians. The method was named by Matheron after the South African geostatistician D. G. Krige, who was the first to develop formal solutions to the problem of estimating ore reserves from sampling (core) data (Krige, 1952, 1966). Geostatistics was developed by Matheron (1962, 1965, 1970, 1971, 1973) and co-workers at the *Centre de morphologie mathématique* of the *École des Mines de Paris*. Geostatistics comprises the estimation of variograms (Subsection 13.1.6), kriging, validation methods for kriging estimates, and simulations methods for geographically distributed ("regionalized") data. Major textbooks have been written by former students of Matheron: David (1977) and Journel & Huijbregts (1978). Other useful references are Clark (1979), Rendu (1981), Verly *et al.* (1984), Armstrong (1989), Isaaks & Srivastava (1989), and Cressie (1991). Applications to environmental sciences and ecology have been discussed by Gilbert & Simpson (1985), Robertson (1987), Armstrong *et al.* (1989), Legendre & Fortin (1989), Soares *et al.* (1992), and Rossi *et al.* (1992). Geostatistical methods can be implemented using the software library of Deutsch & Journel (1992).

As in inverse-distance weighting (eq. 13.19), the estimated value for any grid node is computed as:

$$\hat{y}_{\text{Node}} = \sum_i w_i y_i$$

The chief difference with inverse-distance weighting is that, in kriging, the weights w_i applied to the points $Ø_i$ used in the estimation are not standardized inverses of the distances to some power k. Instead, the weights are based upon the covariances (semi-variances, eq. 13.9 and 13.10) read on a variogram model (Subsection 13.1.6). They are found by linear estimation, using the equation:

$$\mathbf{C} \cdot \mathbf{w} = \mathbf{d}$$

$$\begin{bmatrix} c_{11} & \cdots & c_{1n} & 1 \\ \cdot & \cdots & \cdot & 1 \\ \cdot & \cdots & \cdot & 1 \\ c_{n1} & \cdots & c_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} d_1 \\ \cdot \\ \cdot \\ d_n \\ 1 \end{bmatrix} \qquad \textbf{(13.21)}$$

where **C** is the covariance matrix among the n points $Ø_i$ used in the estimation, i.e. the semi-variances corresponding to the distances separating the various pair of points, as read on the variogram model; **w** is the vector of weights to be estimated (with the constraint that the sum of weights must be 1); and **d** is a vector containing the covariances between the various points $Ø_i$ and the grid node to be estimated. This is where a variogram model becomes essential; it provides the weighting function for the entire map and is used to construct matrix **C** and vector **d** for each grid node to be estimated. λ is a Lagrange parameter (as in Section 4.4) introduced to minimize the

variance of the estimates under the constraint $\Sigma w_i = 1$ (unbiasedness condition). The solution to this linear system is obtained by matrix inversion (Section 2.8):

$$\mathbf{w} = \mathbf{C}^{-1}\,\mathbf{d} \qquad \textbf{(13.22)}$$

Vector **d** plays a role similar to the weights in inverse-distance weighting since the covariances in vector **d** decrease with distance. Using covariances, the weights are statistical in nature instead of geometrical.

Kriging takes into account the grouping of observed points $Ø_i$ on the map. When two points $Ø_i$ are close to each other, the value of the corresponding coefficient c_{ij} in matrix **C** is high; this contributes to lowering their respective weights w_i. In this way, the redundancy of information introduced by dense groups of sampling sites is taken into account.

When anisotropy is present, kriging can use two, four, or more variogram models computed for different geographic directions and combine their estimates when calculating the covariances in matrix **C** and vector **d**. In the same way, when estimation is performed for sampling sites in a volume, a separate variogram can be used to describe the vertical spatial variation. Kriging is the best interpolation method for data that are not on a regular grid or display anisotropy. The price to pay is increased mathematical complexity during interpolation.

Among the interpolation methods, kriging is the only one that provides a measure of the error variance for each value estimated at a grid node. For each grid node, the error variance, called *ordinary kriging variance* (s^2_{OK}), is calculated as follows (Isaaks & Srivastava, 1989), using vectors **w** and **d** from eq. 13.21:

$$s^2_{OK} = \mathrm{Var}\,[y_i] - \mathbf{w}'\mathbf{d} \qquad \textbf{(13.23)}$$

where $\mathrm{Var}[y_i]$ is the maximum-likelihood estimate of the variance of the observed values y_i (eq. 13.14). Equation 13.23 shows that s^2_{OK} only depends on the variogram model and the local density of points, and not on the values observed at points $Ø_i$. The ordinary kriging variance may be used to construct confidence intervals around the grid node estimates at some significance level α, using eq. 13.4. It may also be mapped directly. Regions of the map with large values s^2_{OK} indicate that more observations should be made because sampling intensity was too low.

Kriging, as described above, provides point estimates at grid nodes. Each estimate actually applies to a "point" whose size is the same as the grain of the observed data. The geostatistical literature also describes how *block kriging* may be used to obtain estimates for blocks (i.e. surfaces or volumes) of various sizes. Blocks may be small, or cover the whole map if one wishes to estimate a resource over a whole area. As mentioned in the introductory remarks of the present Section, additive variables only may be used in block kriging. Block kriging programs always assume that the variable is *intensive*, e.g. the concentration of organisms (Subsection 1.4.2). For *extensive*

variables, such as the number of individual trees, one must multiply the block estimate by the ratio (block size / grain size of the original data).

3 — Measures of fit

Different measures of fit may be used to determine how well an interpolated map represents the observed data. With most methods, some measure may be constructed of the closeness of the estimated (i.e. interpolated) values $\hat{y}_i$ to the values y_i observed at sites $Ø_i$. Four easy-to-use measures are:

- The mean absolute error: $MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i|$
- The mean squared error: $MSE = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2$
- The Euclidean distance: $D_1 = \sqrt{\sum_i (y_i - \hat{y}_i)^2}$
- The correlation coefficient (r) between values y_i and $\hat{y}_i$ (eq. 4.7). In the case of a trend-surface model, the square of this correlation coefficient is the coefficient of determination of the model.

In the case of kriging, the above measures of fit cannot be used because the estimated and observed values are equal, at all observed sites $Ø_i$. The technique of cross-validation may be used instead (Isaaks & Srivastava, 1989, Chapter 15). One observation, say $Ø_1$, is removed from the data set and its value is estimated using the remaining points $Ø_2$ to $Ø_n$. The procedure is repeated for $Ø_2$, $Ø_3$, …, $Ø_n$. One of the measures of fit described above may be used to measure the closeness of the estimated to the observed values. If replicated observations are available at each sampling site (a situation which does not often occur), the test of goodness-of-fit described in Subsection 1 can be used with all interpolation methods.

13.3 Patches and boundaries

Multivariate data may be condensed into spatially-constrained clusters. These may be displayed on maps, using different colours or shades. The present Section explains how clustering algorithms can be constrained to produce groups of spatially contiguous sites; study of the boundaries between homogeneous zones is also discussed. Prior to clustering, one must state unambiguously which sites are neighbours in space; the most common solutions to this are presented in Subsection 1.