34
Spatial interpolation

L MITAS AND H MITASOVA

This chapter formulates the problem of spatial interpolation from scattered data as a method for prediction and representation of multivariate fields. The role and specific issues of interpolation for GIS applications are discussed and methods based on locality, geostatistical, and variational concepts are described. Properties of interpolation methods are illustrated by examples of 2-dimensional, 3-dimensional, and 4-dimensional interpolations of elevation, precipitation, and chemical concentrations data. Future directions focus on a robust data analysis with automatic choice of spatially variable interpolation parameters, and model- or process-based interpolation.

1 INTRODUCTION

Spatial and spatio-temporal distributions of both physical and socioeconomic phenomena can be approximated by functions depending on location in a multi-dimensional space, as multivariate scalar, vector, or tensor fields. Typical examples are elevations, climatic phenomena, soil properties, population densities, fluxes of matter, etc. While most of these phenomena are characterised by measured or digitised point data, often irregularly distributed in space and time, visualisation, analysis, and modelling within a GIS are usually based on a raster representation. Moreover, the phenomena can be measured using various methods (remote sensing, site sampling, etc.) leading to heterogeneous datasets with different digital representations and resolutions which need to be combined to create a single spatial model of the phenomenon under study.

Many interpolation and approximation methods were developed to predict values of spatial phenomena in unsampled locations (for reviews see Burrough 1986; Franke 1982a; Franke and Nielson 1991; Lam 1983; McCullagh 1988; Watson 1992; and for a discussion of Kriging and error, see Heuvelink, Chapter 14). In GIS applications, these methods have been designed to support transformations between different discrete and continuous representations of spatial and spatio-temporal fields, typically to transform irregular point or line data to raster representation, or to resample between different raster resolutions.

2 PROBLEM FORMULATION AND CRITERIA FOR SOLUTIONS

The general formulation of the spatial interpolation problem can be defined as follows:

Given the N values of a studied phenomenon z_j, $j = 1, \dots , N$ measured at discrete points $\mathrm{r}_j = (x_j^{[1]}, x_j^{[2]}, \dots , x_j^{[d]})$, $j = 1, \dots , N$ within a certain region of a d-dimensional space, find a d-variate function $F(\mathbf{r})$ which passes through the given points, that means, fulfils the condition

$$F(\mathbf{r}_j) = z_j, \quad j = 1, \dots ,N \tag{1}$$

Because there exist an infinite number of functions which fulfil this requirement, additional conditions have to be imposed, defining the character of various interpolation techniques. Typical examples are conditions based on geostatistical concepts (Kriging), locality (nearest neighbour and finite element methods), smoothness and tension (splines), or ad hoc functional forms (polynomials, multi-quadrics). Choice of the additional condition depends on the character of the modelled phenomenon and the type of application.

Finding appropriate interpolation methods for GIS applications poses several challenges. The modelled fields are usually very complex, data are spatially heterogeneous and often based on far from optimal sampling, and significant noise or discontinuities can be present (e.g. see De Floriani and Magillo, Chapter 38). In addition, datasets can be very large ($N \approx 10^3$–10^6), originating from various sources with different accuracies. Reliable interpolation tools, suitable for GIS applications, should therefore satisfy several important demands: accuracy and predictive power, robustness and flexibility in describing various types of phenomena, smoothing for noisy data, d-dimensional formulation, direct estimation of derivatives (gradients, curvatures), applicability to large datasets, computational efficiency, and ease of use.

Currently, it is difficult to find a method which fulfils all of the above-mentioned requirements for a wide range of georeferenced data. Therefore, the selection of an adequate method with appropriate parameters for a particular application is crucial. Different methods can produce quite different spatial representations (Plate 26 (a)–(f)) and in-depth knowledge of the phenomenon is needed to evaluate which one is the closest to reality. The use of an unsuitable method or inappropriate parameters can result in a distorted model of spatial distribution, leading to potentially wrong decisions based on misleading spatial information. An inappropriate interpolation can have even more profound impact if the result is used as an input for simulations, where a small error or distortion can cause models to produce false spatial patterns (Mitasova et al 1996), as illustrated in section 4.2.2 (Plate 27(b)): see also Heuvelink (Chapter 14) for a discussion of error propagation. Quantitative evaluation of interpolation predictive capabilities, for example by cross-validation, is often not sufficient for the selection of an appropriate interpolation method, as the preservation of geometrical properties is in some cases more important than actual accuracy (see Hutchinson and Gallant, Chapter 9). Advanced visualisation and analysis of slope, aspect, and curvature is helpful in detecting geometrical distortions (Brown et al 1995; Mitas et al 1997; Mitasova et al 1995; Nielson 1993; Wood and Fisher 1993).

3 METHODS

In recent years, GIS capabilities for spatial interpolation have improved by integration of advanced methods within GIS, as well as by linking GIS to systems designed for modelling, analysis, and visualisation of continuous fields. Because it is impossible to cover all or even most of the existing interpolation techniques, only methods which are often used in connection with GIS or have the potential to be widely used for GIS applications are included, and references are given to literature for more detailed descriptions.

3.1 Local neighbourhood approach

Local methods are based on the assumption that each point influences the resulting surface only up to a certain finite distance. Values at different unsampled points are computed by functions with different parameters, and the condition of continuity between these functions is defined only for some approaches. The method of point selection used for the computation of the interpolating function differs among the various methods and their concrete implementations.

3.1.1 Inverse distance weighted interpolation (IDW)

This is one of the simplest and most readily available methods. It is based on an assumption that the value at an unsampled point can be approximated as a weighted average of values at points within a certain cut-off distance, or from a given number m of the closest points (typically 10 to 30). Weights are usually inversely proportional to a power of distance (Burrough 1986; Watson 1992) which, at an unsampled location $\mathbf{r}$, leads to an estimator

$$F(\mathbf{r}) = \sum_{i=1}^{m} w_{\mathrm{i}}\, z(\mathbf{r}_i) = \frac{(\sum_{i=1}^{m} z(\mathbf{r}_i) \,/\, |\mathbf{r} - \mathbf{r}_i|^p}{\sum_{j=1}^{m} 1/|\,\mathbf{r} - \mathbf{r}_j|^p} \qquad (2)$$

where p is a parameter (typically p=2; for more details on the influence of this parameter see Watson 1992). While this basic method is easy to implement and is available in almost any GIS, it has some well-known shortcomings that limit its practical applications (Burrough 1986; Franke and Nielson 1991; Watson 1992). The method often does not reproduce the local shape implied by data and produces local extrema at the data points (Plate 26 (c)). A number of enhancements has been suggested, leading to a class of multivariate blended IDW surfaces and volumes (Franke and Nielson 1991; Tobler and Kennedy 1985; Watson 1992). However, most of these modifications are not implemented within GIS.

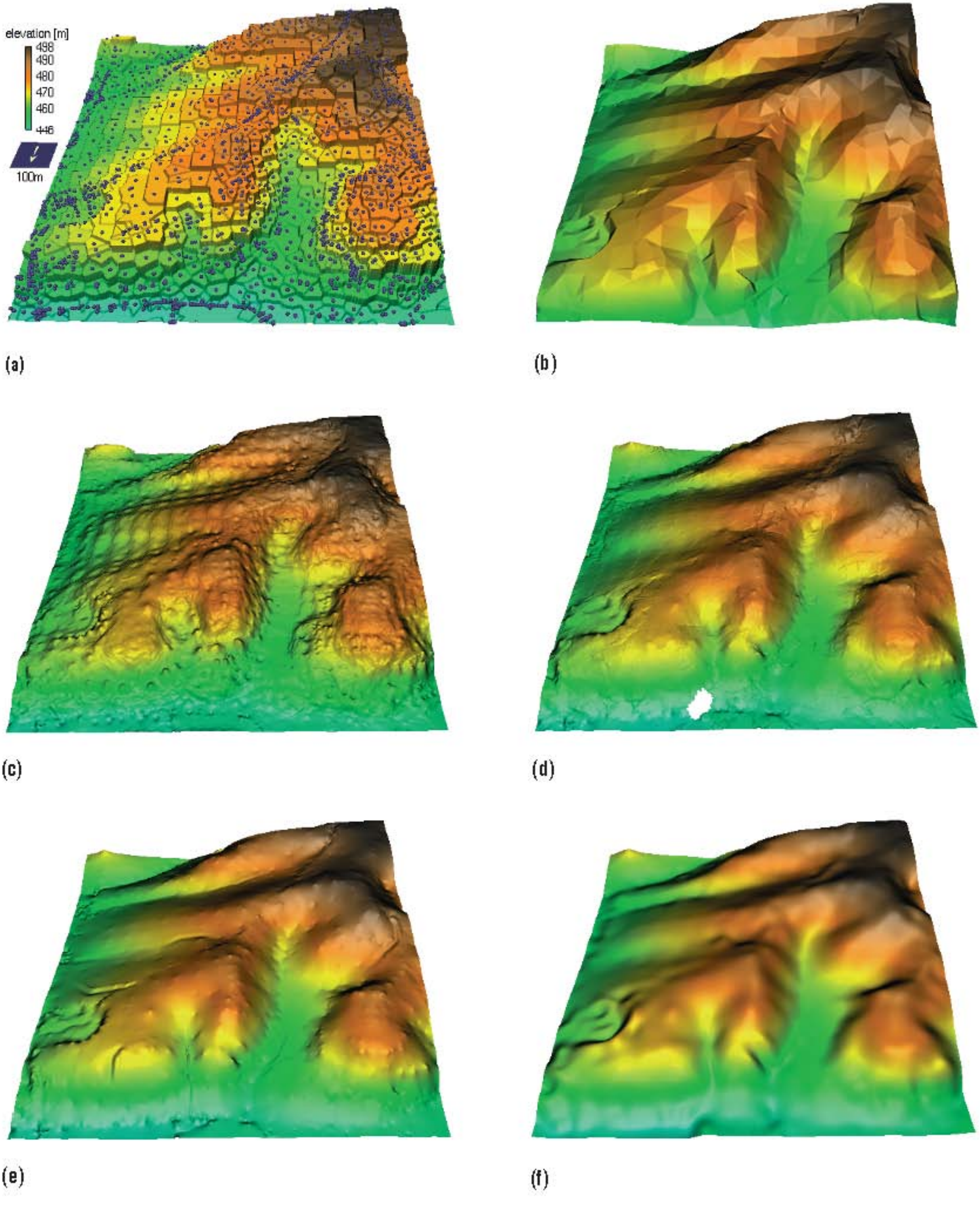


Plate 26

Interpolation of a DEM from scattered point data using methods available in GIS: (**a**) given data and Voronoi polygons; (**b**) TIN-based linear interpolation; (**c**) inverse distance weighting; (**d**) Kriging (spherical variogram); (**e**) spline with tension and stream enforcement; and (**f**) regularised spline with tension and smoothing.

3.1.2 Natural neighbour interpolation

This uses a weighted average of local data based on the concept of natural neighbour coordinates derived from Thiessen polygons (Boots, Chapter 36) for the bivariate, and Thiessen polyhedra for the trivariate case (Watson 1992). The value in an unsampled location is computed as a weighted average of the nearest neighbour values with weights dependent on areas or volumes rather than distances. The number of given points used for the computation at each unsampled point is variable, dependent on the spatial configuration of data points. Natural neighbour linear interpolation leads to a rubber-sheet character of the resulting surface. The addition of blended gradient information (derived from data points by local 'pre-interpolation') allows the surface to be made smooth everywhere with tautness, analogous to tension, tuned according to the character of the modelled phenomenon. The value of tautness is controlled by two empirically selected parameters which modify the shape of the blending function. The result is a surface with smoothly changing gradients and passing through data points, blended from natural neighbour local trends, with local tunable tautness, and with the capability to calculate derivatives and integrals. The method has been used typically for topographic, bathymetric, geophysical, and soil data (Laslett et al 1987; McCauley and Engel 1997; Watson and Philip 1987).

3.1.3 Interpolation based on a triangulated irregular network (TIN)

This uses a triangular tessellation of the given point data (Boots, Chapter 36) to derive a bivariate function for each triangle which is then used to estimate the values at unsampled locations. Linear interpolation uses planar facets fitted to each triangle (Akima 1978; Krcho 1970; Plate 26(b)). Non-linear blended functions (e.g. polynomials) use additional continuity conditions in first-order, or both first- and second-order derivatives (C^1, C^2), ensuring smooth connection of triangles and differentiability of the resulting surface (Akima 1978; McCullagh 1988; Nielson 1983; Renka and Cline 1984). Because of their local nature, the methods are usually fast, with an easy incorporation of discontinuities and structural features. Appropriate triangulation respecting the surface geometry is crucial (Hutchinson and Gallant, Chapter 9; Weibel and Dutton, Chapter 10; Weibel and Heller 1991). Extension to *d*-dimensional problems is more complex than for the distance-based methods (Nielson 1993).

While a TIN provides an effective representation of surfaces useful for various applications, such as dynamic visualisation and visibility analyses (De Floriani and Magillo, Chapter 38), interpolation based on a TIN, especially the simplest, most common linear version, belongs among the least accurate methods (Franke 1982a; Nielson 1993; Renka and Cline 1984).

3.1.4 Rectangle-based methods

These are analogons to a TIN and involve fitting blended polynomial functions to regular or irregular rectangles, such as Hermite, Bezier, or B-spline patches (Chui 1988; Watson 1992), often with locally tunable tension. These methods were developed for computer-aided design and computer graphics and are not very common in GIS applications.

3.2 The geostatistical approach

The principles of geostatistics and interpolation by Kriging are described in a large body of literature (e.g. Burrough 1986; Cressie 1993; Deutsch and Journel 1992; Isaaks and Srivastava 1989; Journel and Huijbregts 1978; Matheron 1971; Oliver and Webster 1990); therefore only the basic notions are outlined here.

Kriging is based on a concept of random functions: the surface or volume is assumed to be one realisation of a random function with a certain spatial covariance (Journel and Huijbregts 1978; Matheron 1971). Using the given data $z(\mathbf{r}_i)$ and an assumption of stationarity one can estimate a semivariogram $\gamma(\mathbf{h})$ defined as

$$\gamma(\mathbf{h}) = \frac{1}{2}\, Var[\{ z(\mathbf{r} + \mathbf{h}) - z(\mathbf{r})\}] \approx \frac{1}{(2N_h)} \sum_{(ij)}^{N_h} [z(\mathbf{r}_i) - z(\mathbf{r}_j)]^2 \qquad (3)$$

which is related to the spatial covariance $C(\mathbf{h})$ as

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \qquad (4)$$

where $C(0)$ is the semivariogram value at infinity (sill). The summation in Equation (3) runs over the number N_h of pairs of points which are separated by the vector $\mathbf{h}$ within a small tolerance $\triangle\mathbf{h}$ (size of a histogram bin). For isotropic data, the semivariogram can be simplified into a radial function dependent on $|\mathbf{h}|$. The Kriging literature provides a choice of functions which can be used as theoretical semivariograms (spherical, exponential, Gaussian, Bessel etc.: Cressie 1993). The parameters of these functions are then optimised for the best fit of the experimental semivariogram.

The interpolated surface is then constructed using statistical conditions of unbiasedness and minimum variance. In its dual form (Hutchinson and Gessler 1993; Matheron 1971) the universal Kriging interpolation function can be written as

$$F(\mathbf{r}) = T(\mathbf{r}) + \sum_{j=1}^{N} \lambda_j \, C(\mathbf{r} - \mathbf{r}_j) \tag{5}$$

where $T(\mathbf{r})$ represents its non-random component (drift) expressed as a linear combination of low-order monomials. The monomial and $\{\lambda_j\}$ coefficients are found by solving a system of linear equations (Hutchinson and Gessler 1993).

In general, Kriging predicts values at points and blocks in d-dimensional space and enables incorporation of anisotropy. Various extensions enhance its flexibility and range of applicability (Cressie 1993; Deutsch and Journel 1992). Co-Kriging includes information about correlations of two or more attributes to improve the quality of interpolation (Myers 1984), while disjunctive Kriging is used for applications where the probability that the measured values exceed a certain threshold is of interest (Rivoirard 1994). For cases in which the assumption of stationarity is deemed not to be valid, zonal Kriging can be used (Burrough 1986; Wingle and Poeter 1996). Approaches for spatio-temporal Kriging reflect the different behaviour of the modelled phenomenon in the time dimension. Time is treated either as an additional dimension with geometric or zonal anisotropy, or as a combination of the space and time correlation functions with a space-time stationarity hypothesis (Bogaert 1996; Rouhani and Myers 1990).

Recent applications of geostatistics have de-emphasised the use of Kriging as an interpolation and mapping tool while shifting the focus towards models of uncertainty that depend on the data values in addition to the data configuration (Armstrong and Dowd 1993; Deutsch and Journel 1992; Englund 1993; Journel 1996; Rogowski 1996; Yarrington 1996). A stochastic technique of conditional simulation is used to generate alternative, equally probable realisations of a surface, reproducing both data and the estimated covariance. From such a set of statistical samples one can estimate a spatially-dependent picture of the uncertainty which is inherent in the data.

The main strengths of Kriging are in the statistical quality of its predictions (e.g. unbiasedness) and in the ability to predict the spatial distribution of uncertainty. It is often used in the mining and petroleum industries, geochemistry, geology, soil science and ecology, where its statistical properties are of great value (Burrough 1991; Cressie 1993; Isaaks and Srivastava 1989; Oliver and Webster 1990). It has been less successful for applications where local geometry and smoothness are the key issues and other methods prove to be competitive or even better (Deutsch and Journel 1992; Hardy 1990).

3.3 The variational approach

The variational approach to interpolation and approximation is based on the assumption that the interpolation function should pass through (or close to) the data points and, at the same time, should be as smooth as possible. These two requirements are combined into a single condition of minimising the sum of the deviations from the measured points and the smoothness seminorm of the spline function:

$$\sum_{j=1}^{N} |z_j - F(\mathbf{r}_j)|^2 \, w_j + w_0 I(F) = \mathit{minimum} \tag{6}$$

where w_j, w_0 are positive weights and $I(F)$ denotes the smoothness seminorm (Table 1). The solution of Equation (6) can be expressed as a sum of two components (Talmi and Gilat 1977; Wahba 1990):

$$F(\mathbf{r}) = T(\mathbf{r}) + \sum_{j=1}^{N} \lambda_j \, R(\mathbf{r}, \mathbf{r}_j) \tag{7}$$

where $T(\mathrm{r})$ is a 'trend' function and $R(\mathbf{r}, \mathbf{r}_j)$ is a basis function which has a form dependent on the choice of $I(F)$. A bivariate smoothness seminorm with squares of second derivatives (Table 1) leads to a *thin plate spline (TPS)* function (Duchon 1976; Harder and Desmarais 1972). The TPS function minimises the surface curvature and imitates a steel sheet forced to pass through the data points: the equilibrium shape of the sheet minimises the bending energy which is closely related to the surface curvature. There are at least two deficiencies of the TPS function: (1) the plate stiffness causes the function to overshoot in regions where data create large gradients; and (2) the second order derivatives diverge in the data points, causing difficulties in surface geometry analysis.

Table 1 Examples of bivariate spline functions, their corresponding smoothness seminorms and Euler–Lagrange equations.

Method	*I(F)*	*Euler–Lagrange Eq.*
Membrane	$\int [F_x^2 + F_y^2]\, d\mathbf{r}$	harmonic
Minimum curvature[a]	$\int [F_{xx}^2 + F_{yy}^2]\, d\mathbf{r}$	biharmonic modified
Thin plate spline[b]	$\int [F_{xx}^2 + F_{yy}^2 + 2F_{xy}^2]\, d\mathbf{r}$	biharmonic
Thin plate spline+tension[c]	$\int [\varphi^2 [F_x^2 + F_y^2] + [F_{xx}^2 + ...]\, d\mathbf{r}$	harmonic+biharmonic
Regular thin plate spline[c]	$\int [F_{xx}^2 + ...] + \tau^2[F_{xxx}^2 + ...]\, d\mathbf{r}$	biharmonic+6th-order
Regular spline with tension[d]	$\Sigma_{mn}\, c_{mn}(\varphi) \int [F_{x^n y^m}]^2\, d\mathbf{r}$	all even orders

[a] Briggs 1974, Duchon 1975, Hutchinson and Bischof 1983, Wahba 1990
[b] Franke 1985, Hutchinson 1989
[c] Mitas and Mitasova 1988
[d] Mitasova et al 1995; Mitas and Mitasova 1997

The problem of overshoots can be eliminated by adding the first order derivatives into the seminorm $I(F)$, leading to a *TPS with tension* (Franke 1985; Hutchinson 1989; Mitas and Mitasova 1988). Change of the tension tunes the surface from a stiff plate into an elastic membrane (Plate 30 (a), (b), and (c); Mitas et al 1997). In the limit of an infinite tension the surface resembles a rubber sheet with cusps at the data points. The analytical properties of TPS can be improved by adding higher order derivatives into $I(F)$, leading to a function with regular second- and possibly higher-order derivatives (Mitas and Mitasova 1988).

To synthesise the desired properties into a single function the *regularised spline with tension (RST)* was proposed and implemented within a GIS (Mitasova et al 1995). The RST function includes the sum of all derivatives up to infinity with rapidly decreasing weights. The resulting surface is of C^∞ class which means that it has regular derivatives of all orders (similar, for example, to a Gaussian) and therefore is suitable for differential analysis and calculations of curvatures (Mitasova and Hofierka 1993; Mitasova et al 1995). Other forms of smoothness seminorm are also possible (e.g. Mitas and Mitasova 1997; Wahba 1990). It is important to note that the splines described in this section are, in general, different from a rich class of piecewise polynomial splines (Chui 1988; Wahba 1990).

RST can be generalised to an arbitrary dimension and the corresponding d-variate formula for the basis function is given by

$$R_d(\mathbf{r}, \mathbf{r}_j) = R_d(|\mathbf{r} - \mathbf{r}_j|) = R_d(\mathbf{r}) = \rho^{-\delta}\gamma(\delta, \rho) - (\tfrac{1}{\delta}) \quad (8)$$

where $r = |\mathbf{r} - \mathbf{r}_j|$, $\delta = (d - 2) / 2$, and $\rho = (\varphi\, r / 2)^2$.

Further, φ is a generalised tension parameter, and $\gamma(\delta,\rho)$ is the incomplete gamma function, not to be confused with a semivariogram (Abramowitz and Stegun 1964). The somewhat less obvious case for $d = 2$ is given by

$$R_2(\mathbf{r}) = \lim_{d\to 0} [\, \rho^{-\delta}\gamma(\delta,\rho) - (\tfrac{1}{\delta})] = -[E_1(\rho) + \ln \rho + C_E] \quad (9)$$

where C_E=0.577215... is the Euler constant and $E_1(.)$ is the exponential integral function (Abramowitz and Stegun 1964).

As has been pointed out by several authors (Cressie 1993; Hutchinson and Gessler 1993; Matheron 1981; Wahba 1990), splines are formally equivalent to universal Kriging with the choice of the covariance function determined by the seminorm $I(F)$. Therefore, many of the geostatistical concepts can be exploited within the spline framework. However, the physical interpretation of splines makes their application easier and more intuitive. The 'thin plate with tension' analogy helps to understand the behaviour of the function also in higher dimensions where the interpolation function models an elastic medium with a tunable tension (Mitas et al 1997; see Plate 30). The RST control parameters such as the tension φ and smoothing weights $\{w_j\}$ proved to be useful and effective for preventing an introduction of artificial features such as waves along contours, artificial peaks, pits or overshoots, often found in the results of less general interpolation techniques, or in spline surfaces with tension set too high (for examples of waves, peaks, and pits: see Plate 26(e) and 29(b)), or too low (overshoots, see Plate 30(c) and (d)) (Mitas et al 1997; Mitasova and Mitas 1993; Mitasova et al 1995; Mitasova et al 1996; Watson 1992). The tension and smoothing parameters can be selected empirically, based on the knowledge of the modelled phenomenon, or automatically by minimisation of the predictive error estimated by a cross-validation procedure (Mitasova et al 1995). Moreover, the tension parameter φ can be generalised to a tensor which enables modelling of anisotropy (Mitas and Mitasova 1997; Mitasova and Mitas 1993).

The interpolation function given by Equation (7) requires solving a system of N linear equations. Therefore processing of large datasets ($N > 10^3$) becomes computationally intractable, as the computer time scales as N^3. Treatment of large datasets is enabled by implementation of an automatic segmentation procedure proposed in various forms (Franke 1982b; Hardy 1990; Mitas and Mitasova 1988; Mitasova and Mitas 1993; Mitasova et al 1995) with computational demands proportional to N. The segmented processing is based on the fact that splines have local behaviour, that is, the impact of data points which are far from a given location diminishes rapidly with increasing distance (Powell 1992). The segmentation uses a decomposition of the studied region into rectangular segments with variable size dependent on the density of data points (Plate 28), using 2^d-trees (Mitasova et al 1995). For a given segment, the interpolation is carried out using the data points within this segment and from its neighbourhood, selected automatically depending on their spatial distribution. For very low tension, this approach requires large neighbourhoods to achieve smooth connection of segments, which makes the segmentation method less efficient for flat, very smooth surfaces with strongly heterogeneous point distributions. Recently, a new, more robust version of RST has been developed which reduces the influence of higher order derivatives and the need for large segment neighbourhoods even for low values of tension (Mitas and Mitasova 1997). Segmentation has allowed users to apply RST to datasets with over a million data points and to interpolate multi-million grid sizes (e.g. Hargrove et al 1995).

Instead of using the explicit solution (7), the minimisation in Equation (6) can be carried out numerically by solving an Euler–Lagrange differential equation corresponding to a given functional (Briggs 1974), for example, by using a finite difference multi-grid iteration method (Hutchinson 1989).

The variational approach offers a wide range of possibilities to incorporate additional conditions such as value constraints, prescribed derivatives at the given or at arbitrary points, and integral constraints (Talmi and Gilat 1977; Wahba 1990). Incorporation of dependence on additional variables, analogous to co-Kriging, leads to partial splines (Hutchinson 1996; Wahba 1990). Numerical solution enables the incorporation of stream enforcement and other topographic features (Hutchinson 1989). Known faults or discontinuities can be handled through appropriate data structures using masking and several independent spline functions (Cox et al 1994). A similar approach can be used to handle regions with spatially variable tension, with blending along their borders, an approach analogous to zonal Kriging. Spatio-temporal interpolation is performed by employing an appropriate anisotropic tension in the temporal dimension (Mitas et al 1997; Mitasova et al 1995).

The spline methods are often used for terrain and bathymetry (Hutchinson and Gallant, Chapter 9; Hargrove et al 1995; Mitasova and Mitas 1993), climatic data (Hutchinson 1996; Hutchinson and Bischof 1983; Wahba 1990), chemical concentrations and soil properties (McCauley and Engel 1997; Mitasova et al 1995), and most recently also for image rectification (Fogel 1996).

While not obtained by a variational approach, similar in formulation and performance are *multiquadrics* (Fogel 1996; Foley 1987; Franke 1982a; Hardy 1990; Kansa and Carlson 1992; Nielson 1993) with $R_d(r) = (r^2 + b)^{1/2}$ or $R_d(r) = (r^2 + b)^{-1/2}$, offering high accuracy, differentiability, d-dimensional formulation, and, with segmentation, applicability to large datasets. Originally ad hoc multiquadrics were later put on more solid theoretical ground. Good performance of multiquadrics, especially in three dimensions, is not surprising, considering that for $d = 3$ in the limit of $b \rightarrow 0$ the basis functions $(r^2 + b)^{1/2}$ and $(r^2 + b)^{-1/2}$ are solutions of biharmonic and harmonic equations respectively (Hardy 1990).

3.4 Relationships and differences between the geostatistical and variational approaches

Theoretical and practical issues of the relationship between Kriging and splines have been discussed in several papers (e.g. Cressie 1993; Dubrule 1984; Hutchinson and Gessler 1993; Laslett et al 1987; Matheron 1981; Wahba 1990); therefore only a brief comment is presented here.

Kriging assumes that the spatial distribution of a geographical phenomenon can be modelled by a realisation of a random function and uses statistical techniques to analyse the data (drift, covariance) and statistical criteria (unbiasedness, minimum variance) for predictions. However, subjective decisions are necessary (Journel 1996) such as judgement about stationarity, choice of function for theoretical variogram, etc. In addition, often the data simply lack information about important features of the modelled phenomenon, such as surface analytical properties or physically acceptable local geometries. As mentioned earlier, Kriging is the most successful for phenomena with a very strong random component or for estimation of statistical characteristics (uncertainty).

Splines rely on a physical model with flexibility provided by change of elastic properties of the interpolation function. Often, physical phenomena result from processes which minimise energy, with a typical example of terrain with its balance between gravitation force, soil cohesion, and impact of climate. For these cases, splines have proven to be rather successful. Moreover, splines provide enough flexibility for local geometry analysis which is often used as input to various process-based models.

However, most of the surfaces or volumes are neither stochastic nor elastic media, but are the result of a host of natural (e.g. fluxes, diffusion) or socioeconomic processes. Therefore, each of the mentioned methods has a limited realm of applicability and, depending on the knowledge and experience of the user, proper choice of the method and its parameters can significantly improve the final results. This will be illustrated to some extent in section 4.2 with applications.

3.5 Application-specific methods

There is a large class of methods specially designed for certain applications which use one of the above-mentioned general principles, but they are modified to meet some application-specific conditions. These methods are too numerous to mention, so only a few examples with references related to GIS applications have been selected.

Area to surface interpolations are designed to transform the data assigned to areas (polygons) to a continuous surface, represented by a high-resolution raster. This task is common in socioeconomic applications, for example for transformation of population density data from census units to a raster, while preserving the value for an area (mass preservation condition), and ensuring smooth transition between the area units (Martin, Chapter 6; Dyn and Wahba 1982; Goodchild and Lam 1980; Goodchild et al 1993; Moxey and Allanson 1994; Tobler 1979, 1996).

Voronoi polygons (Boots, Chapter 36) are sometimes used for transformation of qualitative point data to polygons or a raster when the condition of continuity is not appropriate, resulting in a surface with zero gradients and faults (see Plate 26(a)).

Interpolations on sphere are modifications of the methods described in sections 3.1–3.3 for data given in spherical (latitude/longitude) coordinates. The interpolation functions are dependent on angle rather than on distance (Nielson 1993; Tobler 1996; Wahba 1990). These methods are used for applications covering large areas, such as continental and global Earth or other planets datasets.

Contour/isoline data interpolations are modifications of mostly local methods specifically designed to handle isoline data (Auerbach and Schaeben 1990; Weibel and Heller 1991). Modification of splines for contour data (Hutchinson 1989) supports incorporation of topographic features, such as streams and ridges, to improve the quality of the resulting Digital Elevation Model (DEM: Hutchinson and Gallant, Chapter 9).

Raster data resampling and smoothing can be performed by modifications of methods described in sections 3.1–3.3, with increased efficiency achieved by taking advantage of data regularity. Numerous simpler methods are also available, such as bilinear and local least square polynomial functions.

4 GIS APPLICATIONS OF SPATIAL INTERPOLATION

Spatial interpolation is an important component of almost any GIS. While the basic bivariate methods are common, implementation of multivariate tools is restricted to the most advanced systems, because of the lack of data structures and supporting tools for multivariate and temporal data processing and analysis (Peuquet, Chapter 8). In spite of recent advances in the development of methods and algorithms and an exponential increase in computational power, spatial interpolation, especially for large and complex datasets, can still be an iterative, time-consuming task, requiring an adequate knowledge of underlying methods and their implementation.

4.1 Integration of spatial interpolation within a GIS

Depending on application, spatial interpolation can be performed at three levels of integration with a GIS: (1) within a more general program/command; (2) as a specialised command; or (3) using linkage to specialised software.

Interpolation integrated at a 'sub-command' level can be found in many GIS application programs, such as computation of slope and aspect, automatic raster resampling, flow-tracing, hydrological modelling, etc. Mostly simple and fast local interpolations such as IDW, bilinear, or local polynomial methods are used in this case. The interpolation is fully automatic, hidden from the user, and while it is sufficient for most applications, it can result in artefacts in surfaces if an improper method is implemented.

Interpolations integrated at a command level serve as data transformation functions. A limited set of basic and some advanced methods have been integrated within GIS, most often the simpler versions of IDW, TIN, Kriging and splines. Compromises in numerical efficiency, accuracy, and robustness are common and upgrades to improved modifications of methods are slow, especially for commercial systems. Therefore, it is necessary to evaluate the results carefully and, if possible, to use more than one independent interpolation procedure.

Although interpolations performed by specialised software linked to a GIS provide the most advanced and flexible tools, a time-consuming import/export of data, or inconvenient work in a different software environment might be necessary. This approach still can be preferable, especially when data are complex and high accuracy is required. The advanced surface and volume modelling systems with strong interpolation capabilities often also provide some basic spatial data processing, analysis, and graphical tools, thus evolving into specialised GIS.

4.2 Examples of applications

To illustrate the properties of selected interpolation methods as well as their typical GIS applications, a few representative examples and corresponding references are presented. Data were processed by GRASS, ARC/INFO and S-PLUS, and the illustrations were created by SG3d, SG4d and Nviz visualisation programs (Brown et al 1995).

4.2.1 Bivariate interpolation of elevation data

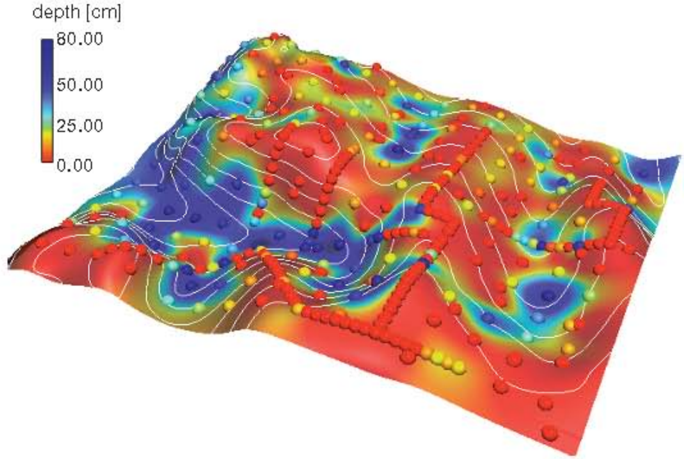
The character of interpolation methods from the simple to the more elaborate ones available in most GIS is illustrated by a common task of interpolation from scattered point elevation data to a raster with 2 m resolution, for an area of approximately 1 km^2. Voronoi polygons (Plate 26(a)), producing a surface with discontinuities, are used to illustrate a relatively homogeneous spatial distribution of the original data. Linear TIN-based interpolation (Plate 26(b)) produces a surface with a typical triangular structure and inadequate description of smaller valleys (triangles creating 'dam' structures across valleys). Application of a non-linear TIN-based method to this dataset resulted in unacceptable overshoots within the triangles. Results of IDW (Plate 26(c)) show a typical pattern with extrema in given points and artificial roughness biased towards the data points. Results of Kriging (Plate 26(d)) and a TPS with tension and stream enforcement (Plate 26(e)) represent a significant improvement; however, subtle discontinuities in Kriging (Plate 26(d)) and small cusps in the data points for both methods are visible at this resolution, although the artefacts are within the data accuracy. The results can be further improved by properly tuning the tension and smoothing, as illustrated by the application of the RST method (Plate 26(f)).

4.2.2 The role of interpolation in modelling

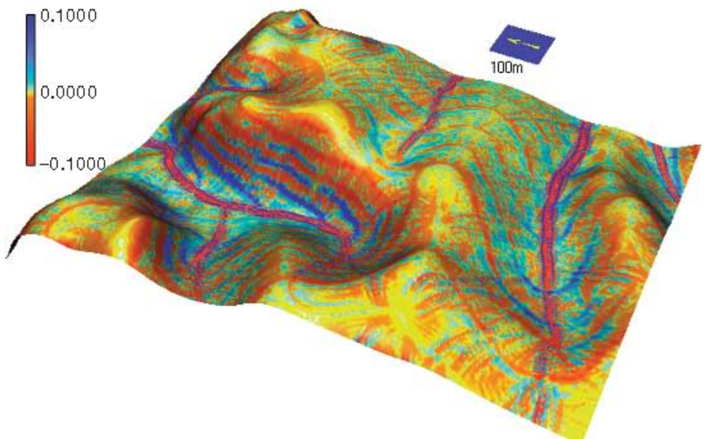
To highlight the impact of subtle interpolation artefacts on the results of a model, an erosion/deposition model (Mitasova et al 1996) was used with DEMs computed from contours (Plate 27(a) white lines) using splines with tension and stream enforcement (Plate 27(b)) and by the RST (Plate 27(c)), and compared with observed depositional areas. Small distortions in the DEM computed by splines with tension set too high lead to an artificial pattern of erosion/deposition biased towards the given contours (Plate 27(b)), because the model is dependent on the highly sensitive surface curvatures.

4.2.3 Interpolation of a large dataset

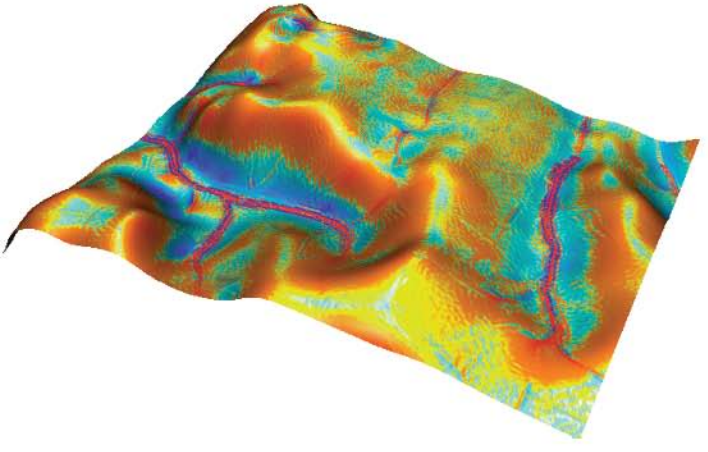
RST method was used to interpolate a 50 m DEM (400x700 grid) from approximately half a million



(a)



(b)



(c)

Plate 27
Influences of interpolation on the results of an erosion/deposition model: (**a**) observed depth of colluvial deposits in the sampling points and as interpolated raster shown in colour; (**b**) artificial erosion/deposition pattern predicted from a DEM interpolated from contours by spline with tension set too high; and (**c**) realistic erosion/deposition pattern predicted using a DEM interpolated from contours by regularised spline with tension and smoothing.

data points digitised from a contour map of Santa Cruz island in California (Plate 28). A segmentation procedure (Plate 28 inset) was used to make the RST method applicable to a large dataset. The example also illustrates that splines can realistically represent rough surfaces in spite of the smoothness condition, if the roughness is sufficiently described by the input data (Plate 28 inset).

4.2.4 Bivariate and trivariate interpolation of precipitation

Multi-dimensional interpolation is also a valuable tool for incorporating the influence of an additional variable into interpolation, for example for evaluation of precipitation with the influence of topography. Plates 29(a) and (b) illustrate the difference between bivariate spline interpolation of annual precipitation in tropical South America, and interpolation with the influence of topography. The importance of incorporating the terrain data is visible especially in the mountainous areas, where the barrier effect of the Andes is very well represented in the results of trivariate interpolation (Plate 29(b)).

4.2.5 Multi-variate interpolation of scattered spatio-temporal data

Capabilities of RST to model spatial and spatio-temporal distributions of phenomena measured in points scattered in 3-dimensional space and time are illustrated by interpolation of nitrogen concentrations in Chesapeake Bay and their change over the year. The analogy between the tensions for bivariate and trivariate interpolation is illustrated by surfaces and volumes interpolated with increasing tension, changing the surface from a stiff medium with overshoots, through adequate tension, to a highly elastic medium with extrema in data points (Plates 31(a)–(f)). Spatio-temporal interpolation was performed by a quadvariate RST function with time as a fourth independent variable. Anisotropic tension in the third (depth) and fourth (time) dimensions was used to ensure a stable solution, and the appropriate tension and smoothing parameters were found by minimising the cross-validation error. The resulting time series of 3-dimensional rasters (Mitasova et al 1995) was then animated to present the dynamic character of the modelled phenomenon (Plate 30 presents a snapshot; Mitas et al 1997 provide animation).

4.3 Future directions

The following paragraphs identify the tasks considered to be the most important in the future development of spatial interpolation techniques.

4.3.1 Robustness and fully automatic method/parameters selection

For widespread, routine use of GIS by users with little expertise in spatial data processing, fully automatic selection of interpolation methods and their parameters based on the robust analysis of given data or *a priori* information about the modelled phenomenon is crucial. With the fast development of communication technologies and accessibility of a variety of data in different formats, this is becoming one of the most urgent tasks for practical applications.

4.3.2 Increase in accuracy and realism

Improvements in accuracy and realism can be expected by employing spatially-variable adaptive interpolation (Hutchinson 1996; Kansa and Carlson 1992) and by further developments in model-based interpolation. More information can be extracted from field data by using process-based models to extrapolate scattered field observations over both time and space (McLaughlin et al 1993). When field data are combined with model predictions the resulting estimates are able to capture the unique characteristics of a specific area while respecting the general physical principles that control the process influencing the spatial distribution of the studied phenomenon. This can be accomplished by using a stochastic/deterministic model of a process together with the concepts of Bayesian estimation theory.

4.3.3 Synthesis of data from various sources

One of the important developments in geosciences is the increasing availability of data generated by various sources (e.g. local measurements, GPS, satellites, radar) which have diverse character from the point of view of resolution, accuracy, distribution etc. This requires novel approaches to data processing and synthesis so that the extraction of information from all sources of data is properly weighted and optimised (Goodchild and Longley, Chapter 40).

4.3.4 Multi-scale modelling

Currently, new types of simulation methods are being developed which span several spatial or spatio-temporal scales. Such approaches provide new challenges for interpolation focused on the design of a versatile and robust approach applicable across the range of scales. Recent progress in wavelet techniques offers one possibility as scale flexibility is one of the fundamental properties of wavelet construction. However, its potential for general multivariate applications remains to be investigated.

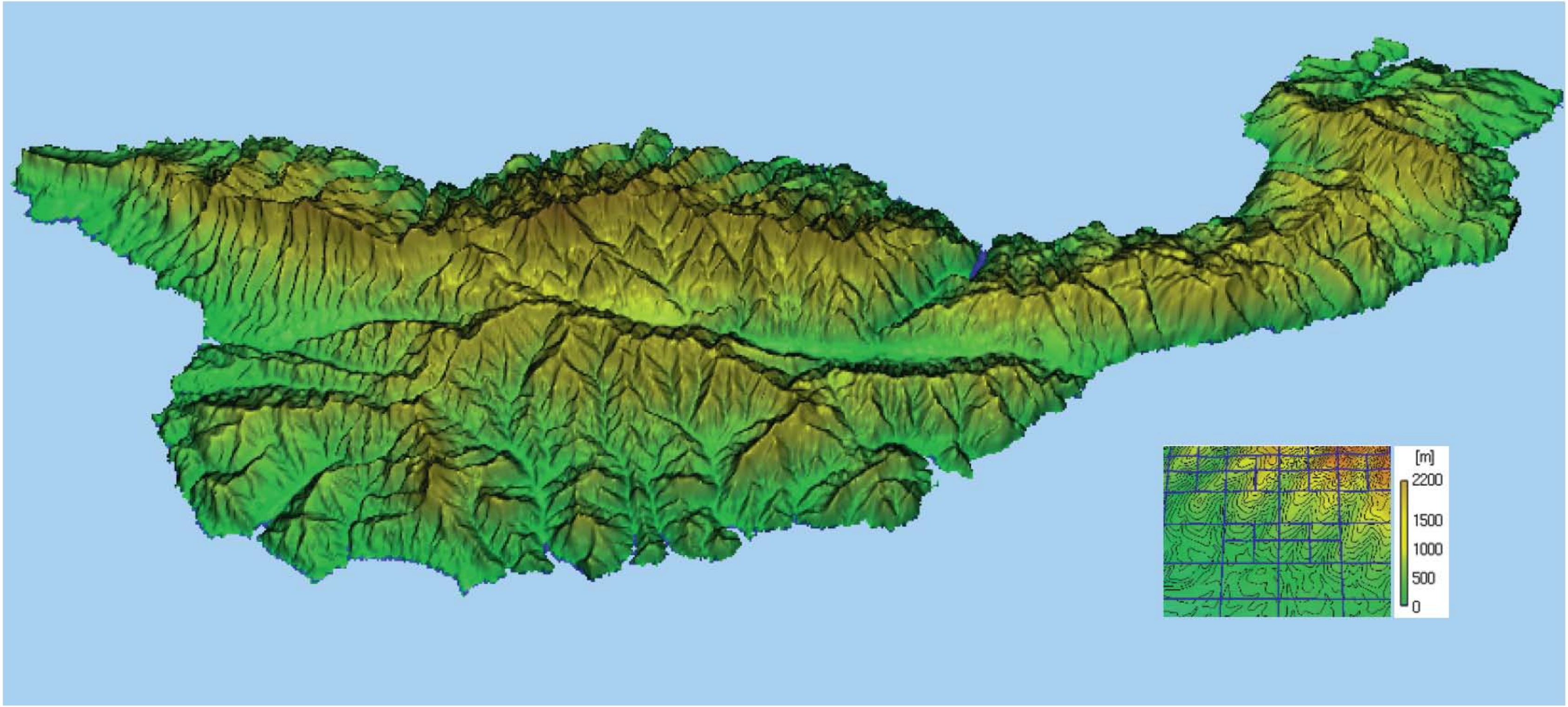
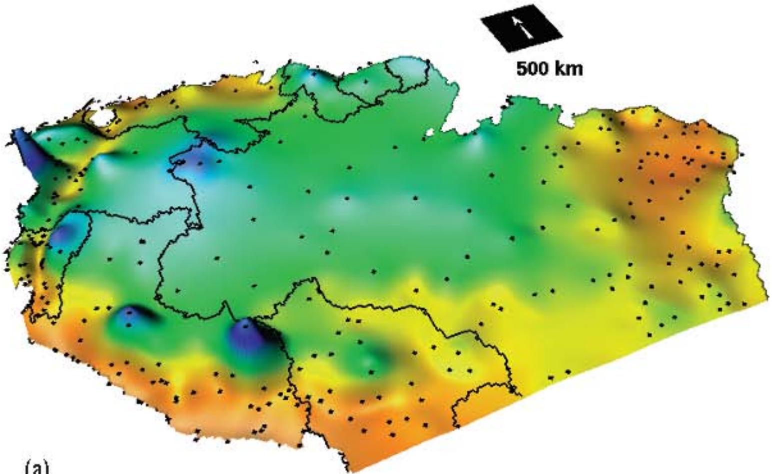
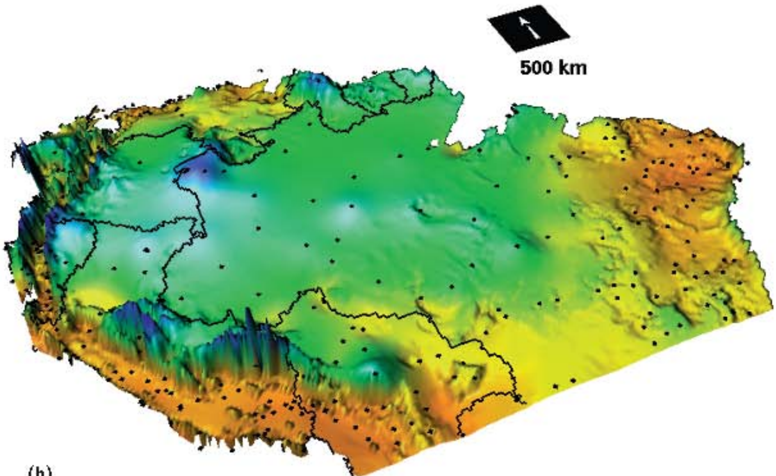


Plate 28

Interpolation of a 50-m resolution DEM (700 × 300) from a large dataset with complex topography using a regularised spline with tension.
Inset shows the principle of segmented processing based on quadtrees and the detail of given contours.



(a)



(b)

Plate 29
Interpolation of annual precipitation in tropical South America using: (**a**) bivariate regularised spline with tension and smoothing; (**b**) trivariate interpolation with incorporation of the influence of topography.

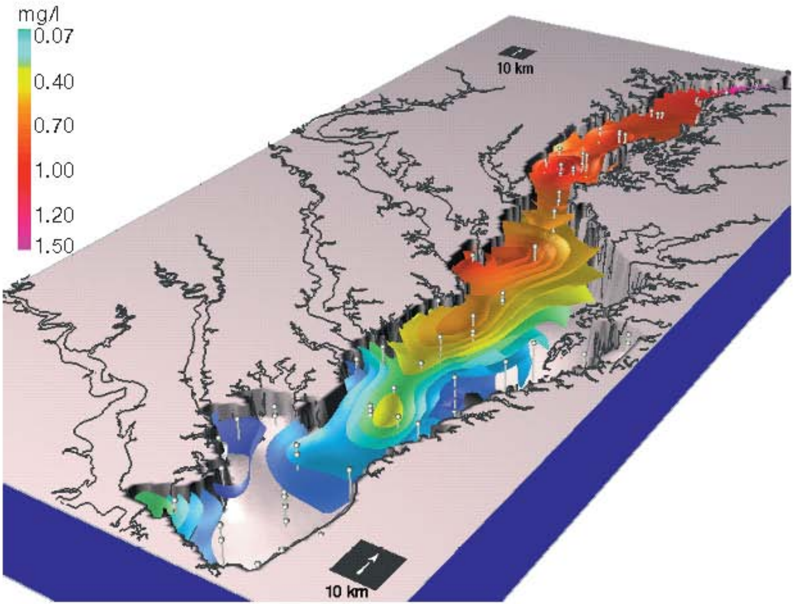


Plate 30
Snapshot from a result of spatio-temporal interpolation (3-D + time) of nitrogen concentrations in Chesapeake Bay (January 1991). For the complete animated result see *http://www.cecer.army.mil/grass/viz/ches.html*.

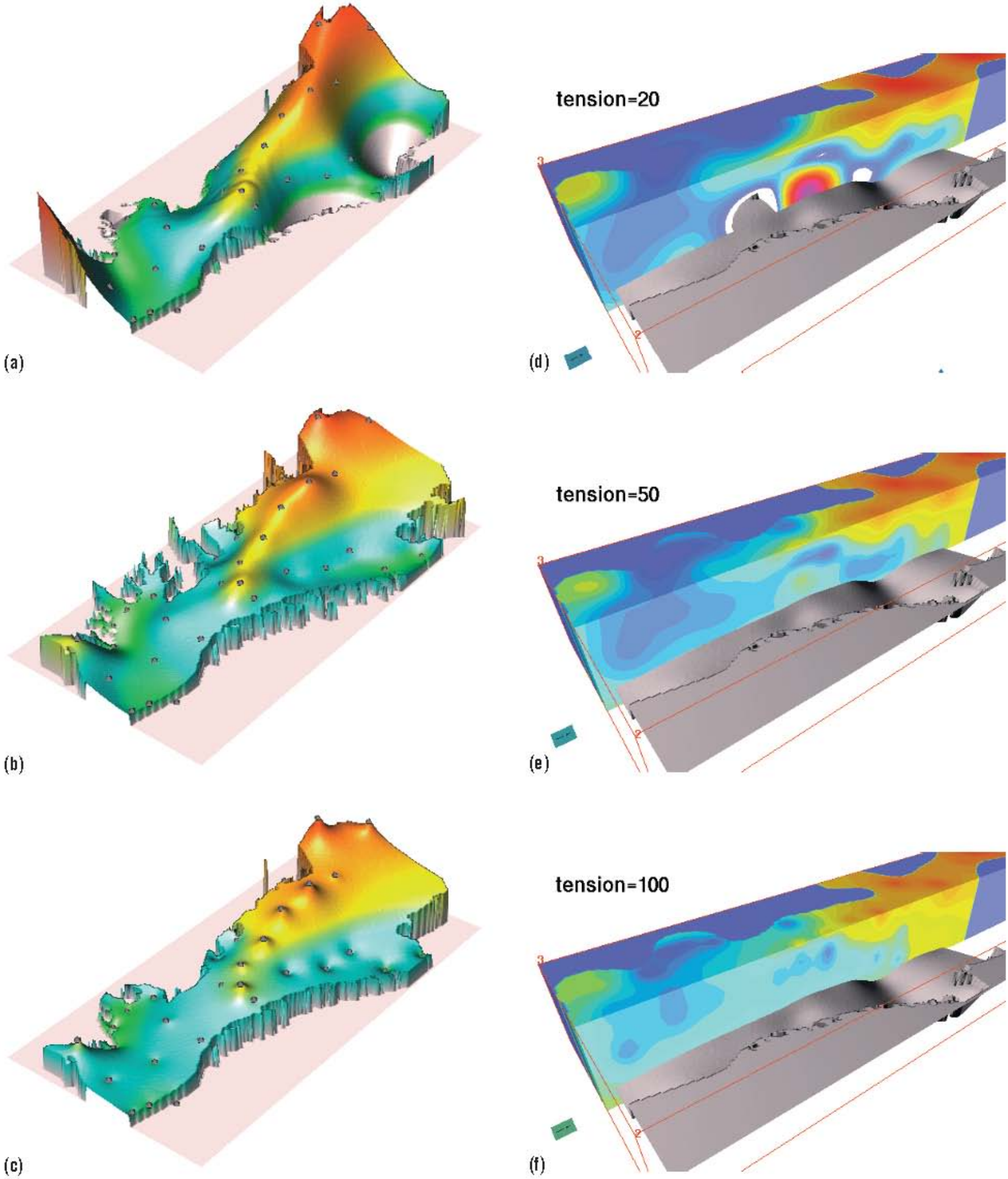


Plate 31
Influence of tension on the surface (**a–c**) and volume (**d–f**) models of nitrogen concentrations in the middle section of Chesapeake Bay. Low tension (**a, d**) leads to overshoots; high tension (**c, f**) creates extrema in data points. Appropriate tension (**b, e**) was found by minimising the cross-validation error.

4.3.5 Multi-dimensional representation

Full integration of support for multi-dimensional data, including data structures, analytical, and visualisation tools will stimulate multivariate applications. Although methods have been presented already which are fully capable of treatment of multi-dimensional data, the current GIS computational infrastructure does not effectively support wide application of multi-dimensional and spatio-temporal modelling.

4.3.6 Computational efficiency

High-accuracy interpolation of large datasets is computationally very intensive, and increase in performance is important for both large cutting-edge applications, as well as for routine use. Further development of algorithms and the use of parallel architectures will be one of the options for speeding up calculations.

5 CONCLUSION

This chapter has presented a review of scattered-data spatial interpolation methods which are relevant for GIS applications. It is obvious that there has been substantial development over the past decade from the points of view of accuracy, multivariate frameworks, robustness, variety of applications, and size of problems tackled.

However, the conclusions outlined by Burrough (1986) are still valid: 'It is unwise to throw one's data into the first available interpolation technique without carefully considering how the results will be affected by the assumptions inherent in the method. A good GIS should include a range of interpolation techniques that allow the user to choose the most appropriate method for the job at hand.' Computers will take over a large part of this nontrivial task, but many problems remain to be resolved.

References

Abramowitz M, Stegun I A 1964 *Handbook of mathematical functions*. New York, Dover: 297–300, 228–31

Akima H 1978 A method of bivariate interpolation and smooth surface fitting for irregularly spaced data points. Algorithm 526. *ACM Transactions on Mathematical Software* 4: 148–64

Armstrong M P, Dowd P A (eds) 1993 *Geostatistical simulations*. Quantitative Geology and Geostatistics 7. Dordrecht, Kluwer

Auerbach S, Schaeben H 1990 Surface representation reproducing given digitised contour lines. *Mathematical Geology* 22: 723–42

Bogaert P 1996 Comparison of Kriging techniques in a space–time context. *Mathematical Geology* 28: 73–83

Briggs I C 1974 Machine contouring using minimum curvature. *Geophysics* 39: 39–48

Brown W M, Astley M, Baker T, Mitasova H 1995 GRASS as an integrated GIS and visualization environment for spatio-temporal modeling. In *Proceedings AutoCarto 12, Charlotte*. ASPRS/ASM: 89–99

Burrough P A 1986 *Principles of geographical information systems for land resources assessment*. Oxford, Clarendon Press

Burrough P A 1991a Soil information systems. In Goodchild M F, Maguire D J, Rhind D W (eds) *Geographical information systems: principles and applications*. Harlow, Longman/New York, John Wiley & Sons Inc. Vol. 2: 153–64

Chui C 1988 *Multivariate splines*. Philadelphia, Society for Industrial and Applied Mathematics

Cox S J D, Kohn B P, Gleadow A J W 1994 Toward a fission track tectonic image of Australia: model-based interpolation in the Snowy Mountains using a GIS. Presented at 12th Australian Geological Convention, Perth, W Australia. *http://www.ned.dem.csiro.au/AGCRC/papers/snowys/12agc.cox.html*

Cressie N A C 1993 *Statistics for spatial data*. New York, John Wiley & Sons Inc.

Deutsch C V, Journel A G 1992 *GSLIB geostatistical software library and user's guide*. New York, Oxford University Press

Dubrule O 1984 Comparing splines and Kriging. *Computers and Geosciences* 10: 327–38

Duchon J 1976 Interpolation dès fonctions de deux variables suivant le principe de la flexion des plaques minces. *RAIRO Annales Numériques* 10: 5–12

Dyn N, Wahba G 1982 On the estimation of functions of several variables from aggregated data. *SIAM Journal of Mathematical Analysis* 13: 134–52

Englund E J 1993 Spatial simulation: environmental applications. In Goodchild M F, Parks B O, Steyaert L T (eds) *Environmental modeling with GIS.* New York, Oxford University Press: 432–7

Fogel D N 1996 Image rectification with radial basis functions: Application to RS/GIS data integration. *Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling. Santa Fe, 21–25 January.* Santa Barbara, NCGIA. CD and WWW

Foley T 1987 Interpolation and approximation of 3-dimensional and 4-dimensional scattered data. *Computers and Mathematics with Applications* 13: 711–40

Franke R 1982a Scattered data interpolation: tests of some methods. *Mathematics of Computation* 38: 181–200

Franke R 1982b Smooth interpolation of scattered data by local thin plate splines. *Computers and Mathematics with Applications* 8: 273–81

Franke R 1985 Thin plate spline with tension. *Computer Aided Geometrical Design* 2: 87–95

Franke R, Nielson G 1991 Scattered data interpolation and applications: a tutorial and survey. In Hagen H, Roller D (eds) *Geometric modelling: methods and applications.* Berlin, Springer: 131–60

Goodchild M F, Anselin L, Deichmann U 1993 A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25: 383–97

Goodchild M F, Lam N S-N 1980 Areal interpolation: a variant of the traditional spatial problem. *Geoprocessing* 1: 297–312

Harder R, Desmarais R 1972 Interpolation using surface splines. *Journal of Aircraft* 9: 1989–91

Hardy R L 1990 Theory and applications of the multiquadric-biharmonic method. *Computers and Mathematics with Applications* 19: 163–208

Hargrove W, Levine D, Hoffman F 1995 Clinch River environmental restoration program. *http://www.esd.ornl.gov/programs/CRERP/*

Hutchinson M F 1989 A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* 106: 211–32

Hutchinson M F 1996b Thin plate spline interpolation of mean rainfall: getting the temporal statistics correct. In Goodchild M F, Steyaert L T, Parks B O, Johnston C A, Maidment D R, Crane M P, Glendinning S (eds) *GIS and environmental modelling: progress and research issues*. Fort Collins, GIS World Inc.: 85–91

Hutchinson M F, Bischof R J 1983 A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales. *Australian Meteorological Magazine* 31: 179–84

Hutchinson M F, Gessler P E 1993 Splines – more than just a smooth interpolator. *Geoderma* 62: 45–67

Isaaks E H, Srivastava R M 1989 *An introduction to applied geostatistics*. Oxford, Oxford University Press

Journel A G 1996 Modelling uncertainty and spatial dependence: stochastic imaging. *International Journal of Geographical Information Systems* 10: 517–22

Journel A G, Huijbregts C J 1978 *Mining geostatistics*. London, Academic Press

Kansa E J, Carlson R E 1992 Improved accuracy of multiquadric interpolation using variable shape parameters. *Computers and Mathematics with Applications* 24: 99–120

Krcho J 1973 Morphometric analysis of relief on the basis of geometric aspect of field theory. *Acta Geographica Universitae Comenianae, Geographica Physica* Vol. 1. Bratislava, SPN

Lam N S-N 1983 Spatial interpolation methods: a review. *The American Cartographer* 10: 129–49

Laslett G M, McBratney A B, Pahl P J, Hutchinson M F 1987 Comparison of several spatial prediction methods for soil pH. *Journal of Soils Science* 38: 32–50

Matheron G 1971 The theory of regionalised variables and its applications. *Les Cahiers du Centre de Morphologie Mathematique de Fontainebleu* 5. Paris

Matheron G 1981 Splines and Kriging: their formal equivalence. *Syracuse University Geological Contributions*: 77–95

McCauley J D, Engel B A 1997 Approximation of noisy bivariate traverse data for precision mapping. *Transactions of the American Society of Agricultural Engineers* 40: 237–45

McCullagh M J 1988 Terrain and surface modelling systems: theory and practice. *Photogrammetric Record* 12: 747–79

McLaughlin D, Reid L B, Li S G, Hyman J 1993 A stochastic method for characterising groundwater contamination. *Groundwater* 31: 237–49

Mitas L, Mitasova H 1988 General variational approach to the interpolation problem. *Computers and Mathematics with Applications* 16: 983–92

Mitas L, Mitasova H 1997 *Multivariate approximation by regularized spline with tension*. Urbana, National Center for Supercomputing Applications: 1–5

Mitas L, Mitasova H, Brown W M 1997 Role of dynamic cartography in simulations of landscape processes based on multi-variate fields. *Computers and Geosciences* 23: 437–46 includes CD and *http://www.elsevier.nl/locate/cgvis*

Mitasova H, Hofierka J 1993 Interpolation by regularised spline with tension: II. Application to terrain modeling and surface geometry analysis. *Mathematical Geology* 25: 657–69

Mitasova H, Hofierka J, Zlocha M, Iverson L R 1996 Modelling topographic potential for erosion and deposition using GIS. *International Journal of Geographical Information Systems* 10: 629–41

Mitasova H, Mitas L 1993 Interpolation by regularised spline with tension: I. Theory and implementation. *Mathematical Geology* 25: 641–55

Mitasova H, Mitas L, Brown W M, Gerdes D P, Kosinovsky I, Baker T 1995 Modelling spatially and temporally distributed phenomena: new methods and tools for GRASS GIS. *International Journal of Geographical Information Systems* 9: 433–46

Moxey A, Allanson P 1994 Areal interpolation of spatially extensive variables: a comparison of alternative techniques. *International Journal of Geographical Information Systems* 8: 479–87

Myers D 1984 Co-Kriging – new developments. In Verly G, David M, Journel A, Marechal A (eds) *Geostatistics for natural resource characterisation* Part 1. Dordrecht, Reidel: 295–305

Nielson, G 1983 A method for interpolating scattered data based upon a minimum norm network. *Mathematics of Computation* 40: 253–71

Nielson G M 1993 Scattered data modeling. *IEEE Computer Graphics and Applications* 13: 60–70

Oliver M A, Webster R 1990 Kriging: a method of interpolation for GIS. *International Journal of Geographical Information Systems* 4: 313–32

Powell M J D 1992 Tabulation of thin plate splines on a very fine 2-dimensional grid. In Braess D, Schumaker L L (eds) *Numerical methods of approximation theory* 9: 221–44

Renka R J, Cline A K 1984 A triangle-based C^1 interpolation method. *Rocky Mountain Journal of Mathematics* 14: 223–37

Rivoirard J 1994 *Introduction to disjunctive Kriging and non-linear geostatistics*. Oxford, Clarendon Press

Rogowski A S 1996 Quantifying soil variability in GIS applications: II Spatial distribution of soil properties. *International Journal of Geographical Information Systems* 10: 455–75

Rouhani S, Myers D E 1990 Problems in space–time Kriging of geohydrological data. *Mathematical Geology* 22: 611–23

Talmi A, Gilat G 1977 Method for smooth approximation of data. *Journal of Computational Physics* 23: 93–123

Tobler W R 1979b Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association* 74: 519–36

Tobler W R 1996 Converting administrative data to a continuous field on a sphere. *Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, 21–25 January*. Santa Barbara, NCGIA. CD and WWW

Tobler W R, Kennedy S 1985 Smooth multi-dimensional interpolation. *Geographical Analysis* 17: 251–7

Wahba G 1990 *Spline models for observational data*. CNMS-NSF Regional conference series in applied mathematics 59. Philadelphia, SIAM

Watson D F, Philip G M 1987 Neighborhood based interpolation. *Geobyte* 2: 12–16

Watson D F 1992 *Contouring: a guide to the analysis and display of spatial data*. Oxford, Pergamon

Weibel R, Heller M 1991 Digital terrain modelling. In Goodchild M F, Maguire D J, Rhind D W (eds) *Geographical information systems: principles and applications*. Harlow, Longman/New York, John Wiley & Sons Inc. Vol. 2: 269–97

Wingle W L, Poeter E P 1996 Site characterization, visualization, and uncertainty assessment using zonal Kriging and conditional simulation (poster). *Global Exploration and Geotechnology, AAPG Annual Meeting San Diego, May*

Wood J D, Fisher P J 1993 Assessing interpolation accuracy in elevation models. *IEEE Computer Graphics and Applications* 13: 48–56

Yarrington L 1996 Sequential Gaussian simulation. Fernald Geostatistical Techniques, Sandia National Laboratories. *http://www.nwer.sandia.gov/fernald/techniq.html*