

## Chapter 7 Other Spatial Prediction Techniques

### 7-1. General

*a.* In this chapter, some alternative approaches to spatial prediction are discussed. At the beginning of Chapter 2, the distinction between stochastic and nonstochastic techniques for spatial prediction was discussed. Kriging, the main subject of this ETL, is a stochastic technique because of the structure that is imposed in terms of an underlying random process (the regionalized variables) with joint probability distributions that obey certain assumptions. Kriging yields the predictor that is statistically optimal in the sense that it is the best linear unbiased predictor, given certain assumptions that are detailed in Chapter 2. There are other stochastic techniques that are less well-known than kriging in applications, such as Markov-random-field prediction and Bayesian nonparametric smoothing (see Cressie (1991)), but these will not be discussed here.

*b.* Several techniques that are often applied in a nonstochastic setting will be discussed. Techniques applied in such a setting are typically applied strictly empirically and not evaluated with respect to rigorous statistical criteria such as mean squared prediction error, although, as discussed in Chapter 2, such criteria may be applied in certain of the techniques such as simple average and trend analysis. It has been shown in this ETL that there are some compelling advantages for assuming some kind of stochastic setting. However, the simplicity of not having to postulate and justify the structure and assumptions inherent in stochastic analyses might be considered one advantage of nonstochastic techniques, and such an analysis may be perfectly adequate for certain problems. In addition to statistical optimality and simplicity, there are other considerations in selecting a spatial prediction technique, such as ease of computation, sensitivity to data errors, and whether the predictors are exact interpolators; that is, match the measurements exactly at the measurement locations  $x_1, x_2, \dots, x_n$ . The last property is one that needs to be

given careful consideration by the practitioner. Kriging, as it is usually applied, is an exact interpolator. Questions may be raised, however, about whether this is a desirable property if it is known that the measurements are contaminated with a considerable amount of measurement error. One advantage of stochastic methods in general is that existence of measurement error may be incorporated objectively, and, in fact, some kriging software packages (including STATPAC) have this feature, resulting in a surface that is not an exact interpolator. Several of the nonstochastic methods discussed in this section depend on a parameter that controls the deviation from exact interpolation. The ability to adjust such a parameter when using these techniques lends a degree of flexibility to the practitioner, but selecting the best value may not be straightforward and may involve considerable subjectivity on the part of the practitioner.

*c.* In most of the following techniques, the predictor of the process at location  $x_0$  takes the form of a linear combination of the measurements at locations  $x_i, i=1, 2, \dots, n$ . Using  $\tilde{Z}(x_0)$  to denote an arbitrary predictor (the notation distinguishes the predictors to be discussed in this section from the kriging predictor, which is denoted by  $\hat{Z}(x_0)$ ), the definition of  $\tilde{Z}(x_0)$  is

$$\tilde{Z}(x_0) = \sum_{i=1}^n w_i Z(x_i) \quad (7-1)$$

Although this form is the same form that is taken by the kriging predictor, the difference is in the way the coefficients  $w_i$  are computed.

### 7-2. Global Measure of Central Tendency (Simple Averaging)

*a.* The predictor for the process at any location  $x_0$  is the simple average of the measurements; that is, the weights  $w_i$  are all equal and are given by Cressie (1991)

$$w_i = \frac{1}{n} \quad (7-2)$$

This predictor represents the smoothest possible predictor surface. In using this predictor, a certain degree of spatial homogeneity is assumed. No attempt is made to incorporate any detectable patterns (or trends) in the mean or variance of the data as a function of location, and the fact that measurements made at points that are close to each other may be related is disregarded. Such a predictor has the advantage of being very simple to compute; it needs no estimation of a variogram or other model parameters. The disadvantage is that representing the spatial field by a single value ignores much of the relevant and interesting structure that may be very helpful in improving predictions.

*b.* As discussed in section 2-4, if applied in a stochastic setting, this predictor would be optimal (best linear unbiased) if there is no drift and if residuals are uncorrelated and have a common variance.

### 7-3. Simple Moving Average

*a.* Let  $h_{i0}$  be the distance of  $x_0$  from  $x_i$ , let  $h_{[i0]}$  be the ordered (from smallest to largest) distances, and fix  $1 \leq k \leq n$ . Then the weights  $w_i$  are (Cressie 1991)

$$w_i = \begin{cases} \frac{1}{k}, & h_{i0} \leq h_{[k0]} \\ 0, & h_{i0} > h_{[k0]} \end{cases} \quad (7-3)$$

Thus, this predictor is the average of the measurements at the  $k$  nearest locations from  $x_0$ .

*b.* If  $k$  is equal to  $n$ , this predictor is identical to the simple average, with weights as given in Equation 7-2. A choice of  $k$  smaller than  $n$  reflects an assumption that the predictor needs to incorporate more of the local fluctuation observed in the data, or, equivalently, that measurements at locations near  $x_0$  should be more informative than measurements at other locations in predicting  $z(x_0)$ ; the smaller  $k$  is, the more variable the predictor. If

$k = 1$ , the predictor is an exact interpolator and is constant on the Voronoi polygons (see section 7-5) induced by the measurement locations.

*c.* There are several variations of this predictor. In one such variation, a distance  $r$  may be fixed (rather than fixing  $k$ ) and averages over locations that are within distance  $r$  of  $x_0$  taken. Additionally, a moving-median may be used rather than a moving average. Sorting and testing distances can slow computations relative to obtaining the simple average, and use of medians rather than means leads to a more resistant (to outliers) predictor.

### 7-4. Inverse-Distance Squared Weighted Average

*a.* The weights  $w_i$  are (Journel and Huijbregts 1978)

$$w_i = \frac{\frac{1}{h_{i0}^2}}{\sum_{j=1}^n \frac{1}{h_{j0}^2}} \quad (7-4)$$

where again  $h_{i0}$  is the distance of  $x_0$  from  $x_i$ .

*b.* In the simple moving average, weights are the same, provided measurement locations are sufficiently close to the prediction location and are zero otherwise. For the inverse-distance squared method, weights are forced to decrease in a smoother manner as distance from the prediction location increases. This predictor again has the advantage of being easy to compute. Another feature of this predictor is that it is an exact interpolator. In addition, the exponent 2 of  $h_{i0}$  may be changed to any positive number, giving the user some flexibility in determining the rate of decrease of weights as a function of distance from  $x_0$ . Isaaks and Srivastava (1989, pp. 257-259) present an example illustrating the effects on weights of changing the exponent.

## 7-5. Triangulation

a. To compute this predictor, the region R is partitioned into what are referred to as Voronoi polygons  $V_1, V_2, \dots, V_n$ , with  $V_i$  being the set of locations closer to measurement location  $\underline{x}_i$  than to any other measurement location. If any two polygons,  $V_i$  and  $V_j$ , share a common boundary,  $\underline{x}_i$  and  $\underline{x}_j$  are joined with a straight line. The collection of all such lines defines what is known as the Delaunay triangulation. There will be one such triangle containing the prediction location  $\underline{x}_0$ ; the vertices of this triangle, which are measurement locations, are labelled  $\underline{x}_j, \underline{x}_k$ , and  $\underline{x}_l$ . The spatial prediction at  $\underline{x}_0$  will be the planar interpolant through the coordinates  $(\underline{x}_j, z(\underline{x}_j))$ ,  $(\underline{x}_k, z(\underline{x}_k))$ , and  $(\underline{x}_l, z(\underline{x}_l))$ . Joining  $\underline{x}_0$  and  $\underline{x}_j, \underline{x}_k$ , and  $\underline{x}_l$ , three subtriangles are formed. The weights  $w_i$  are (Cressie 1991)

$$w_i = \begin{cases} \frac{A_i}{A_j + A_k + A_l}, & i = j, k, \text{ or } l \\ 0, & \text{otherwise} \end{cases} \quad (7-5)$$

where  $A_i$  is the area of the subtriangle opposite vertex  $\underline{x}_i$ .

b. These definitions are illustrated in Figure 7-1. In this figure, the dashed lines depict the Voronoi polygons associated with points  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_6$ , and the solid lines define the Delaunay triangulation. Vertices of the triangle containing the prediction point  $\underline{x}_0$  are  $\underline{x}_1, \underline{x}_5$ , and  $\underline{x}_6$ , and dotted lines show the subtriangles defining the associated area  $A_1, A_5, A_6$ . For this example,  $j, k$ , and  $l$  in the general Equation 7-5 are 1, 5, and 6, so the weights assigned to points  $\underline{x}_1, \underline{x}_5$ , and  $\underline{x}_6$  are, respectively,

$$\begin{aligned} w_1 &= \frac{A_1}{A_1 + A_5 + A_6}, \\ w_5 &= \frac{A_5}{A_1 + A_5 + A_6}, \text{ and} \\ w_6 &= \frac{A_6}{A_1 + A_5 + A_6} \end{aligned} \quad (7-6)$$

It is seen that the weight assigned to a point is proportional to the area of the triangle opposite the point.

c. Computation of this predictor is slower than computation of those in sections 7-2, 7-3, and 7-4. The predictor is an exact interpolator, and the surface produced is continuous, but not differentiable at the edges of the triangulation.

## 7-6. Splines

a. In spline modeling, the measurements are interpolated using combinations of certain so-called basis functions. These basis functions are usually taken to be piecewise polynomials of a certain degree, say  $k$ , which is determined by the user. The coefficients of these polynomials are chosen so that the function values and the first  $k-1$  derivatives agree at the locations where they join. The larger  $k$  is, the smoother will be the prediction surface. Spline techniques are often applied in a non-stochastic framework; in such a context they represent a way of fitting a surface with certain smoothness properties to measurements at a set of locations with no explicit consideration of statistical optimality. There is, however, a considerable body of work in which this technique is applied in a stochastic setting. Splines may be used, for example, in nonparametric regression estimation problems (Wegman and Wright 1983).

b. A typical approach to formulating a spline problem is to pose it as an optimization problem. In one special case, it is assumed that the first two derivatives of the prediction surface exist, which is a way of imposing a certain degree of smoothness, and that the spline function minimizes

$$\frac{1}{n} \sum_{i=1}^n [z(\underline{x}_i) - \tilde{z}(\underline{x}_i)]^2 + \eta Q \quad (7-7)$$

where  $Q$  is a term that depends on the first two derivatives of the predictor surface. The parameter  $\eta$  is a nonnegative number that needs to be

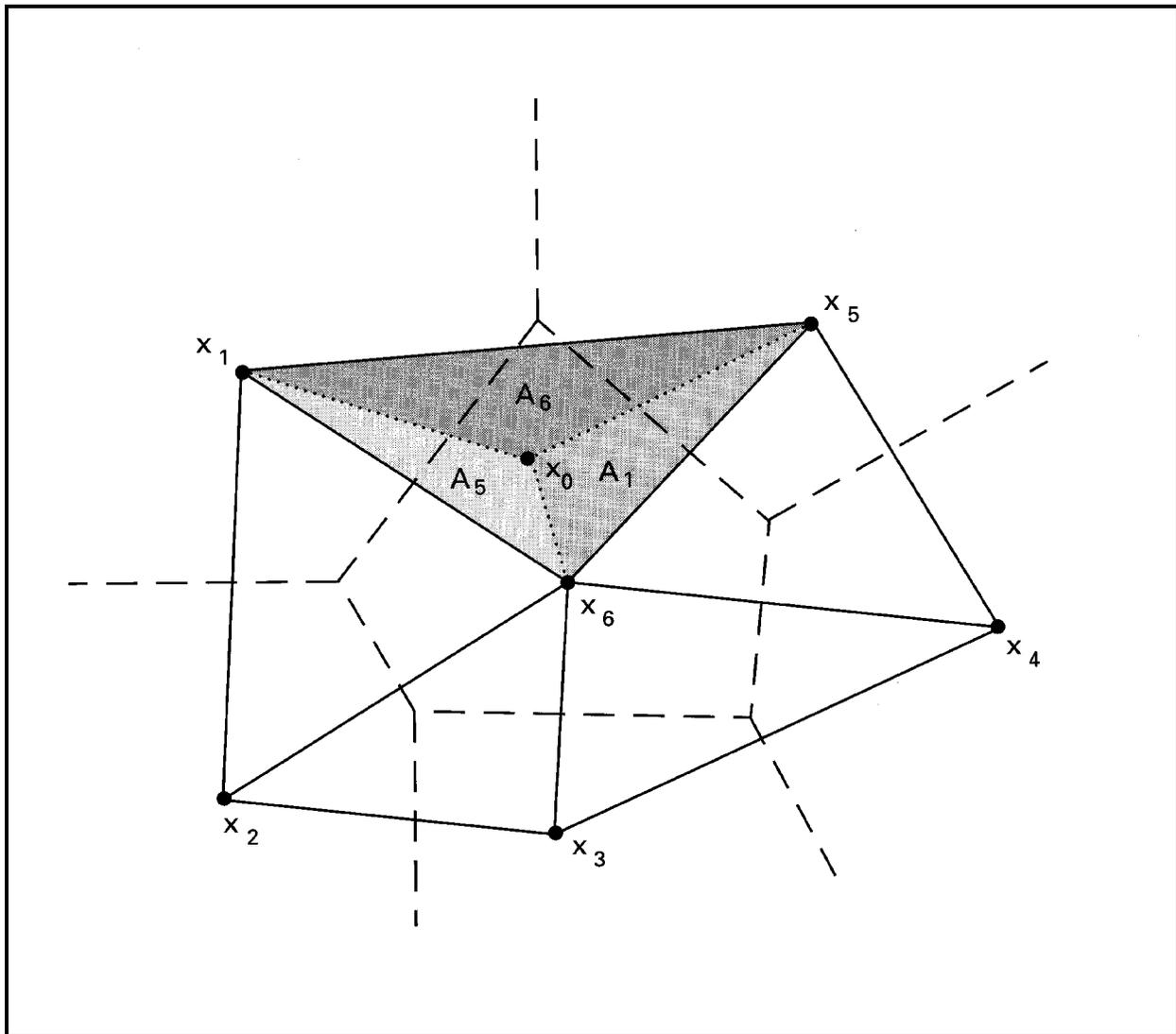


Figure 7-1. Diagram showing Voronoi polygons

specified by the user; the value of this parameter reflects the trade-off between goodness of fit to the data, measured by the first term, and smoothness, as measured by  $Q$ . If  $\eta$  is chosen to be 0, the spline is an exact interpolator and passes through all the data points. If  $\eta > 0$ , the spline is not an exact interpolator. (Splines that are not exact interpolators are referred to as smoothing splines.) There are a number of numerical procedures that may be used for fitting splines, but allowing the

smoothing parameter  $\eta$  to be  $> 0$  renders the computational problem more complex.

c. Under some conditions a solution to the optimization problem (Equation 7-7) may also be obtained by a kriging algorithm if the smoothing parameter  $\eta$  is taken to be equal to the variance of measurement error and if a special form is chosen for the covariance function. Therefore, in this situation, spline approximation is a special type of

kriging. However, the variogram that needs to be used in the kriging equations to make the kriging predictor equivalent to the spline predictor is determined by the basis functions selected for the spline. Because the type of basis functions used is subjective on the part of the user, the resulting equivalent variogram may not be representative of the true variogram of the data. Because kriging uses the data to indicate reasonable variogram choices, kriging has an important advantage over splines. Another advantage of placing the problem in the kriging framework is the interpretation of the smoothing parameter in terms of measurement errors. In many cases, an objective estimate of the magnitude of measurement error can be obtained. The connections between kriging and splines are discussed further by Wegman and Wright (1983), Watson (1984), and Cressie (1991).

### 7-7. Trend-Surface Analysis

*a.* Trend-surface analysis is the process of fitting a function, such as that in Equation 2-43, using least squares to determine the coefficients that yield the best fit. Computationally, trend-surface analysis is equivalent to universal kriging with an assumption that the  $Z^*(x_i)$  are uncorrelated. Thus, there is no need to estimate a variogram, and readily available regression packages may be used for estimating the coefficients. As in universal kriging, polynomial surfaces are the most commonly used.

*b.* When trend surfaces are applied in a stochastic setting, the resulting predictor will be optimal if deviations from the surface are uncorrelated and have a common variance.

### 7-8. Simulation

*a.* Consider again a regionalized random variable  $Z(\underline{x})$ , where  $\underline{x}$  is a location in a two-dimensional study region  $R$ . Kriging is an interpolation algorithm that yields spatial predictions  $\hat{Z}(\underline{x})$  that are best, or optimal, in the sense that has been discussed at some length in this ETL. The

mean-squared prediction error is smallest among all predictors that are linear in the measurements. This optimality property is local, in that the mean-squared error of predictions at unsampled locations considered one at a time is minimized, without specific regard to preservation of global spatial features. If, however, the actual realization  $z(\underline{x})$  could be compared to the kriged prediction surface based on  $n$  measured values, the kriged surface would be much smoother than the actual surface, especially in regions of sparser sampling. Thus, the kriged surface will be a good and realistic representation of reality in the sense that the  $n$  measured values are honored, but it will be less realistic with respect to global properties, such as overall variability.

*b.* The purpose of simulation is to produce one or more spatial surfaces (realizations) that are more realistic in preserving global properties than the surface produced by interpolation algorithms, such as kriging. These realizations are produced by using numbers that are drawn randomly (Monte Carlo) to impart variability to the simulated surface, making the simulated surface more realistic in preserving the overall appearance of the actual surface. Generally speaking, simulation uses the idea that the true value of a random surface may be expressed as the sum of a predicted value (which is obtained by kriging) plus a random error, which varies spatially and depends on the random numbers drawn. Generally a number of independent realizations will be generated, and these realizations will be taken to be equally probable representations of reality.

*c.* A simulation algorithm is said to be conditional if the resulting realizations agree with the measurements at measurement locations  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ . If the underlying process  $Z(\underline{x})$  is assumed to be Gaussian (or if a transformation may be found that makes the process Gaussian), the most common method of conditional simulation is known as sequential Gaussian simulation (Deutsch and Journel (1992), pp. 141-143). Another, more complicated, Gaussian simulation method that is particularly useful for three-dimensional simulations because of its computational efficiency is the

turning-bands method (Deutsch and Journel 1992, Journel and Huijbregts 1978).

*d.* In sequential Gaussian simulation a set of grid points for which simulated values are desired is defined and the points are addressed sequentially from location to location along a predetermined path. At each location, a specified set of neighboring conditioning data is retained, including the original data and simulated grid-location values at previously traversed grid locations along the path. Then, a random number is generated from a Gaussian distribution with conditional mean and variance determined using a kriging algorithm, and the value of the random number determines the simulated process at this location. The conditional Gaussian distribution used in simulation is identical to the conditional distribution discussed in section 2-6*b*. An idea of the computational requirements can be obtained from the fact that a kriging algorithm needs to be applied for each simulation location. For multiple realizations, if the path connecting the grid points is kept the same, the kriging equations need to be solved for only the first simulation. However, implementation of this procedure needs to take into consideration the assumptions concerning the existence of drift; the details of such an implementation are beyond the scope of this ETL.

*e.* A sequential algorithm like this may also be applied in the context of indicator kriging (see section 2-6*c*). At each grid point along the path, a (Bernouli) random variable taking on only two possible values, 0 or 1, is generated, with the relative probability of these two values being determined by indicator kriging applied, as in the previous paragraph, to the original observed indicator data and the previously simulated indicator values.

*f.* To get an idea of how simulation results might be used in a risk-assessment setting, assume again that the underlying process is Gaussian and that 1,000 conditional realizations have been generated. If a single grid point  $x_0$  (which is not a measurement point) is considered, then the simulation has produced 1,000 values at  $x_0$ , which,

when analyzed in histogram form, approximates the probability distribution of potential measurements at that location. If an interval with exactly 25 (2.5 percent) of the values less than the lower end and 25 of the values larger than the upper end were constructed, the interval would almost correspond, as expected, to the 95-percent prediction interval to  $\hat{Z}(x_0) - 1.96\sigma_K(x_0)$  to  $\hat{Z}(x_0) + 1.96\sigma_K(x_0)$  discussed in section 2-6*b*. Thus, for this single location, the simulation has not produced much more information than kriging alone would have produced. The real value of simulation, however, is that realizations not just at a single location, but at all of the grid locations jointly, are obtained. These realizations can be used to calculate probabilities associated with any number of spatial locations together. For example, the probability that the largest (maximum) contaminant value over a certain subregion is greater than a particular concentration might be assessed. (If the word "largest" here were replaced with "average," then block kriging could be used to obtain the answer.)

*g.* A central point that needs to be emphasized is that simulation is especially useful when probabilities associated with complicated, usually nonlinear, functions of the regionalized variables over a region need to be analyzed. The maximum function mentioned in the preceding paragraph is one simple example. For another example, consider the problem of determining placement of groundwater monitoring wells to detect and monitor groundwater contamination emanating from a potential point source. Given an existing set of hydraulic-head data, kriging might be applied and flow lines determined from resulting hydraulic-head gradients. Intersection of the flow line from the point source with the regional boundary then might be used to determine monitoring well placement. Conditional simulation would be useful to determine uncertainty associated with location of well placement or to give an indication of how many monitoring wells might be appropriate. In this case, the variable of interest, well location, is a complicated function of hydraulic heads so this is a problem for which simulation is well-suited. The reader may refer to Easley, Borgman, and Weber

(1991) for a more detailed discussion of this type of application.

*h.* The complicated functions of interest in groundwater studies often involve physically based groundwater flow models. Conditional simulation may be used, for example, to generate a suite of

hydraulic-conductivity realizations to be used as input to a model that produces as output a set of corresponding hydraulic-head realizations. Weber, Easley, and Englund (1991) discuss how groundwater modeling might be used with conditional simulation to study the monitoring-well-placement problem discussed in the preceding paragraph.