5. SPATIAL INTERPOLATION

5.1   *Introduction*

Spatial interpolation is the procedure for estimating the value of properties at unsampled sites within an area covered by existing observations (Waters, 1989). A wide variety of spatial interpolation techniques exist. You may already be familiar with some of these techniques, for example, contouring.   Others, such as Theissen polygons, Delaunay triangles, and spatial moving averages are likely to be new. Many of the more complex interpolation techniques such as kriging and Fourier analysis require a comprehensive understanding of spatial statistics, mathematics and, in some cases, involve the solving of linked simultaneous equations. Don't panic! The primary objective of this section is to explain the concepts which underpin the interpolation process. You will be introduced, in detail, only to those interpolation techniques which you are likely to find in commercially available GIS software. Where a more complex technique is introduced, additional references will be provided to guide you to a more detailed and practical discussion of the approach. A good overview to spatial interpolation can be found in Burrough and McDonnell (1998) and in Lam (1983).

5.2 *What do we mean by spatial interpolation?*

The definition by Waters (1989), provided in the first sentence of the section above, is a good starting point from which to explore what is meant by spatial interpolation. Another definition states the geographical nature of the problem more explicitly:

> "Given a set of spatial data either in the form of discrete points or for subareas, find the function that will best represent the whole surface and that will predict values at points or for other subareas."
> (Lam, 1983)

From a GIS perspective (and for those of you who have never come across the term before) it is useful to dissect these definitions and translate them into the language of attributes and entities introduced in **Spatial Data Modelling** and **DatabaseTheory**. Let's assume that you have a GIS which contains the location of a set of weather stations represented as a series of point entities. For each weather station a range of attribute information is recorded on a regular basis. One attribute observed for each station is the mean daily wind speed. Now, let us assume that a new airport is planned for the region. You would already have used a variety of the spatial operations outlined in Sections 2 and 3 of this module to identify a series of potential sites for the airport. The question that still remains is - which of these sites has the most favourable wind conditions? The problem is that none of the potential airport sites you have selected contain, or are adjacent to, a weather station. Even if one of the weather stations was bang in the middle of one of your preferred sites, the wind values available to you are for a point source and may not be representative of the wind conditions across the entire site.

With this airport example in mind, let us reconsider the definition of interpolation given by Waters (1989):

> **"Spatial interpolation is the procedure of estimating values of properties...."**

Let us replace the word 'properties' with 'attributes'. In our example the attribute we are interested in is mean daily wind speed

> **"at unsampled sites...."**

This indicates that we have a spatial reference which allows us to locate the position of our airport and also infers that we have a conceptual spatial model, in this case an area entity which represents the potential airport site

"within the area...."

This infers that we have a geographical region of interest or study in which we must perform our interpolation. The term 'area' implies that our area of study will have a set of boundaries. In certain cases these boundaries may be physical, for example, the coastline of Britain. On the other hand, our boundaries may be political, such as administrative boundaries. One of the problems with all interpolation techniques is what to do about boundaries because although we are rarely concerned about what happens outside our area of interest we should be concerned about how the properties of the 'outside world' influence our interpolation process. Unfortunately, in many cases we simply turn a blind eye. We will revisit the boundary issue in relation to individual interpolation techniques a little later in this section.

"covered by existing observations."

This implies that we have another set of spatial features (in the case of our example weather stations) for which we have a known set of attribute values, here mean daily wind speed.

Therefore, in the language of GIS, the problem is which spatial interpolation procedures are designed to address how to estimate the attribute values for one set of spatial entities from those for a second set of spatial entities (which may or may not be of the same type) for which a set of observed attribute values exist. Or, in the case of our example how to make the 'best guess' at what the wind speed is likely to be at each of our potential airport sites.

If we return to the definition by Lam (1983) two further concepts are introduced. These are **Function** and **Surface**. Lam's statement of the interpolation problem stresses that before we can estimate the values for our airport sites we must first derive a spatial surface from our known data points (weather stations). This surface should be defined by the nature of the relationship (function) which exists between our observed data points. In a spatial context this function is usually based on the observation that:

> " points *(or spatial entities)* close together in space
> are more likely to have similar values than points
> *(entities)* far apart"
> (Tobler's Law of Geography; Tobler, 1976).

Other spatial relationships can be envisaged, for example repeating sequences in which similar values are found for points only at regular intervals and not for points in between. However, the vast majority of phenomena are thought to conform to Tobler's 'law'.

This 'law' leaves open a wealth of possibilities for the exact nature of the function. If near points have similar values, then we need to consider just how dissimilar they can be and how dissimilarity is a function of distance. Our functions can be simple or complex, sensitive or insensitive to distance and other factors. The mapping of these functions produces surfaces in two dimensions. It is these surfaces which have to be 'fitted' to empirical observations for estimates to be made at sites where values are unknown. It is worth remembering that with such surfaces, as with map overlay, there are serious problems with error in the fitting and estimation processes.

Therefore, to summarise, the spatial interpolation process involves:-

1. Establishing the spatial entity type for which observed attribute data are available.

2. Establishing the spatial entity type for which you wish to estimate attribute values.

3. Establishing the scale of measurement of the observed information. For example, is it ordinal, ratio etc.? (Refer to Section 3, Box 2 of this module for a reminder of what these mean.).

4. Determining how much is known about the spatial behaviour of the observed attribute you are trying to interpolate.

5. Establishing the relationship between the sites for which you have observed attribute values and selecting an appropriate interpolation technique. In virtually every case this will result in choosing a mathematical function which fits a surface to the spatial entities for which you have data.

6. Extracting from your interpolated surface, if required, the 'best guess value' for an entity of the same type or different type at a new location where real world data are unavailable.

The interpolation process, described above, has also been summarised in Figure 5.1 below.

**Figure 5.1 : A conceptual overview of the spatial interpolation process**
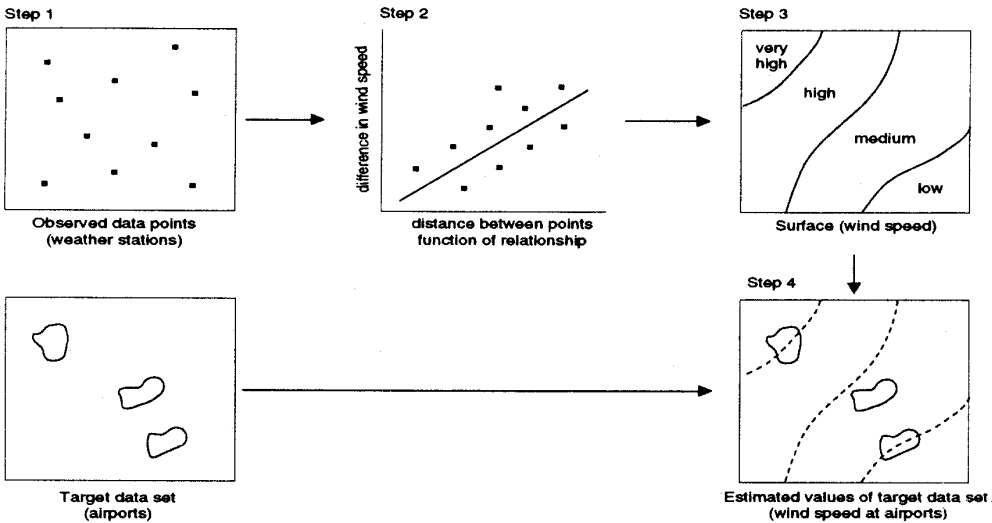
5.3. Why is spatial interpolation important in GIS?

Spatial interpolation techniques have wide application in the design, construction and application of a GIS. This is because we are rarely in the fortunate position of having access to all the data required for every location in our study area. It is much more likely that we will have available partial, and very often patchy data. Figure 5.2 presents a conceptual view of the more common situations in which GIS designers will find themselves with respect to data availability (the figure makes reference to a point data set by way of example).
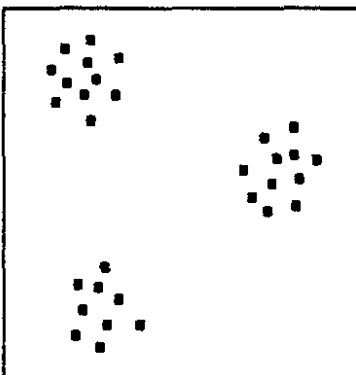
Figure 5.2 : Four different views of data availability in the real world

1	1	6	6	1	3	4	8	7	9
6	5	7	1	3	1	5	5	4	9
7	6	2	4	6	2	3	6	4	8
1	9	8	7	4	5	2	2	7	8
1	3	4	7	8	1	12	7	9	7
1	2	5	6	9	10	11	2	3	1
1	10	9	2	1	1	3	4	5	2
13	1	1	2	3	2	8	6	4	1
2	1	5	4	5	7	6	9	6	10
1	7	6	2	3	1	9	7	9	2

Utopia: Complete coverage,
an observed value at
every spatial location
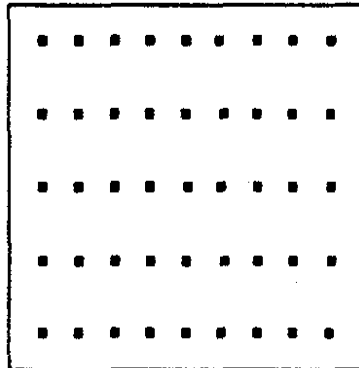
Stratified cover: An observed
value at regular intervals
across the study area

Patchy cover: Clusters of
observed values at
specific locations
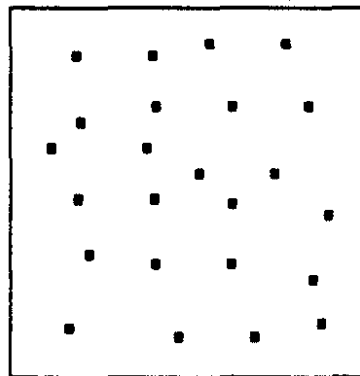
Random Cover: A set of
observed values at
random locations throughout
the study area

The basic role for interpolation in GIS is to fill in the missing data for those areas where real world observations are not available. In the case of Figure 5.2 this is the blank spaces or regions between data points. The important issue about interpolation, from a design perspective, is that any map layer we create through this process will contain

information which is only a representation of what we believe to be true. It is all too easy to forget this as we integrate and overlay data sets in the GIS environment. Therefore, understanding the quality of a map, or any data for that matter, derived from a set of GIS operations, relies on your appreciation of the limitations of the techniques that you have selected to generate the output data.

The field of spatial interpolation is an area of current research interest. Many of the techniques that are described below have been adopted from the one dimensional methods originally developed for the analysis of time series data. Research is being undertaken into the methods and problems of transferring these methods to two dimensional data with different scales, sampling distributions and temporal properties. At the present time few, if any, of the commercial GIS in the market place will have even 50% of the techniques described below as part of their GIS toolbox. Many of these missing techniques will gradually appear in the next generation of GIS, so keep your eyes and ears open!

Continuous Self Assessed Exercise 12

(time: 20 minutes)

Choose two examples of GIS applications with which you are familiar. These could be from your work, examples provided with the Diploma or any additional reading you may have done. Make a list of all the instances where you think it has been necessary to use an interpolation technique to generate a new data layer. Indicate what the original data layer consisted of (for example, vegetation sample points), what the new data layer produced was (for example, vegetation zones) and what the entity types involved were (for example, point to area).

The list of possible uses of interpolation techniques within GIS is vast. Waters (1989) provides the following list of potential uses:

- to provide contours for displaying data graphically
- to calculate some property of the surface at a given point
- to change the unit of comparison when using different data models in different layers
- to aid in the spatial decision making process both in physical and human geography and in related disciplines such as mineral prospecting and resource exploration

One of the most common methods used to classify the many different methods of interpolation is to group them according to the spatial entities on which they can operate; points, lines or areas. This approach will be used to structure the discussion which follows.

5.4. Point interpolation methods

Point interpolation methods may be categorised under a number of different headings:

- local or global
- exact or approximate
- gradual or abrupt
- deterministic or stochastic.

An individual interpolation method will fit several of these categories, for instance it may be local, exact, abrupt and deterministic (such as Theissen polygons). Before we can expect you to be able to recognise what type of interpolation methods you are using, we had better look at the ideas behind the terminology.

5.4.1 Local or global

The terms local and global describe how an interpolation technique uses the observed data collected for a set of geographically spread sampling points. Global interpolation techniques apply a single function to all the points in a study area. In such cases, as Waters (1989) indicates, a change in one observed value would affect the entire interpolation process. Global interpolators tend to produce a surface with few abrupt changes, because, in general, they employ the principle of averaging. This reduces the influence of extreme values. Such methods are most appropriate when the surface being modelled is known to have an overall trend (For example, a pollutant may be known to decrease in occurrence in proportion to distance from a source).
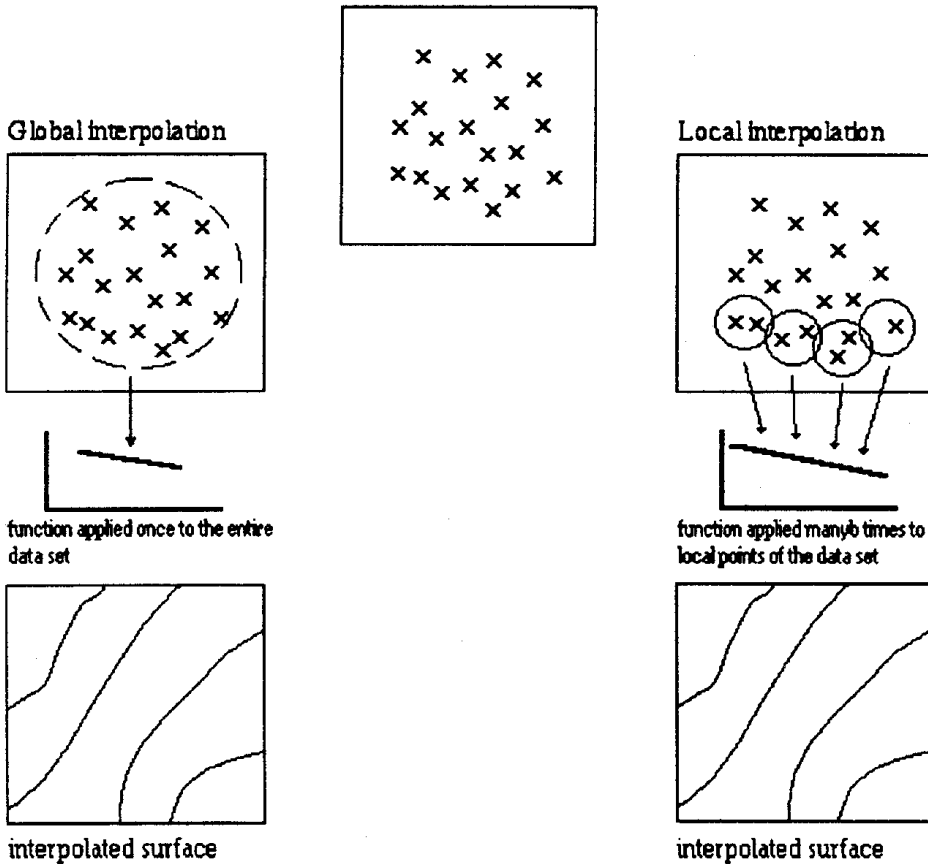
Where there is little or no knowledge about the overall trend in the surface you are trying to model then local interpolators are more appropriate. Local interpolation techniques apply the same function repeatedly to a small portion of the total set of sample points for which data have been observed. Then a surface is constructed by linking these regional observations together.

Figure 5.3 illustrates the distinction between local and global methods of interpolation. A useful analogy is to think of global interpolators as a telescope where everything you wish to observe is present in the field of view and local interpolators as a microscope where you build up a complete picture by observing each part in turn.

One question debated in the literature is when does a local interpolation technique cease to be local and become global? This is an interesting question since, in theory, the neighbourhood of points used in a local interpolation procedure could be extended to cover the entire study area. For our purposes we will consider a local interpolator to be

any technique which makes more than a single pass over the data before constructing a surface. Examples of local interpolation techniques we will review in the following sections include Theissen polygons, Delaunay triangles, moving averages and kriging. Global interpolators to be considered include line threading, trend surface analysis and Fourier analysis.

**Figure 5.3 : Global and local interpolators**

Global interpolation

Local interpolation

function applied once to the entire data set

function applied manyb times to local points of the data set

interpolated surface

interpolated surface

5.4.2 Exact or approximate

Exact interpolation techniques must honour all the points for which observed data are available. They are most appropriate when there is a high degree of certainty attached to measurements. Altitude, for example, is usually known with a high degree of certainty. You would not want to use an interpolation technique which did not honour an observed altitude value but instead replaced it with a smoother approximation. Approximate interpolators, on the other hand, do not have to honour all the points for which observed data are available. These techniques are most appropriate when there is some uncertainty about the ability to reproduce the observed value at a given sample point. Figure 5.4 shows a surface which has been fitted to a set of points using both techniques. Note how the surface produced by the approximate interpolator does not honour all of the data points.

Figure 5.4 : Two surfaces produced by exact and approximate interpolators
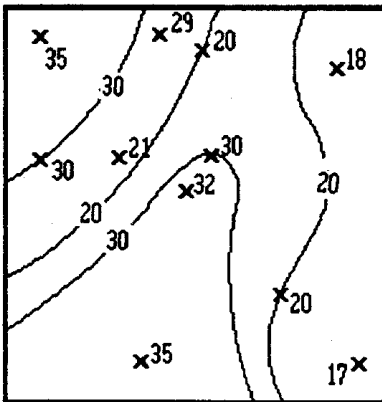
An exact surface which honours the observed values
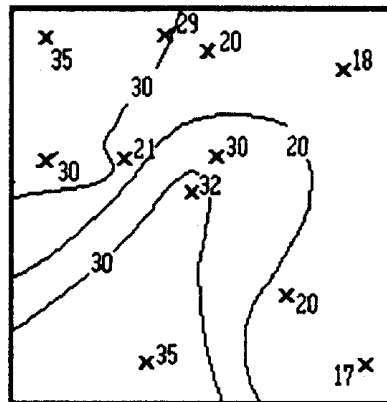
An approximate surface which does not honour all the observed values

Examples of exact interpolation techniques discussed later include; line threading, Theissen polygons and kriging. Examples of approximate interpolators include spatial moving averages, trend surface analysis and a variety of the filtering operations used in remote sensing.

5.4.3. Gradual or abrupt

Gradual or abrupt interpolators are distinguished by the continuity of the surface they produce. Gradual interpolation techniques will produce a smooth surface with gradual changes occurring between observed data points. Abrupt interpolators will produce a surface which is stepped in appearance. An example of an abrupt interpolation method is buffering which you looked at in an earlier section. Another technique to be introduced later is Theissen polygons. This defines discrete territories associated with specific points. Waters (1989) suggests that it may be necessary to include barriers into a gradual interpolation process in which case an abrupt interpolator is appropriate. In terrain modelling, for example, you may wish to represent the presence of a cliff or in climatic modelling the presence of a weather front. For such features there has to be a definable step function which can be used with a gradual surface.

5.4.4. Deterministic or stochastic

The most desirable situation to be in when faced with interpolating new values from a known data set is where there is sufficient knowledge about the phenomena to allow its behaviour to be described by a mathematical function. This is known as deterministic modelling. Figure 5.5 shows how knowledge of the laws of physics and measurements of actual velocity would allow the calculation, from seven sample points, of the trajectory of a bouncing ball.

Figure 5.5 : A deterministic model of the trajectory of a bouncing ball
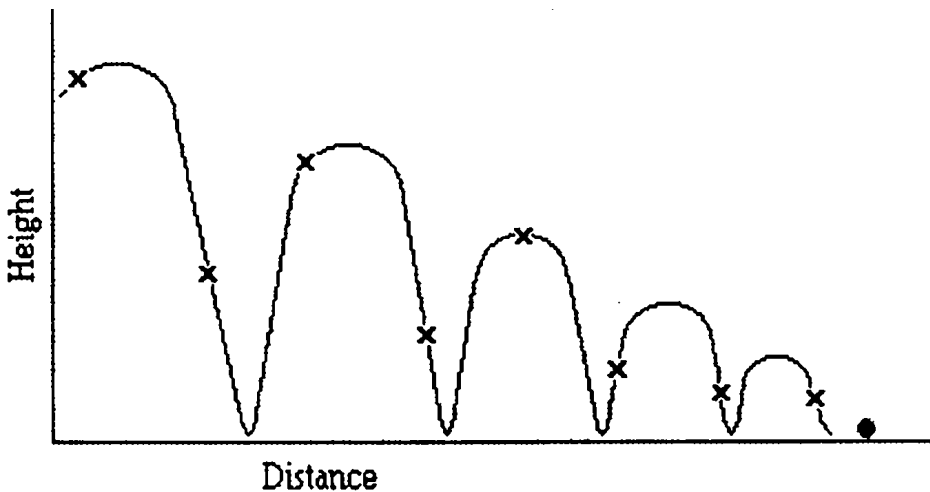
In addition, deterministic modelling allows extrapolation beyond the range of our sample data set. However, this is only possible if the context of the data values is understood. Unfortunately, few geographical phenomena are understood in sufficient detail or obey such precise rules to permit a deterministic approach to interpolation or extrapolation. There is a great deal of uncertainty about what happens at unsampled locations. To handle this problem stochastic models are frequently used. Stochastic models incorporate the concept of randomness suggesting that the interpolated surface, point line or area values are only one of an infinite number that might have been produced from the known data points. Figure 5.6 shows a series of random trajectories for our bouncing ball, all of which could in theory occur if the behaviour of our ball was assumed to be random. Kriging, trend surface analysis and Fourier analysis are all techniques which assume some degree of random behaviour in the observed data.

Figure 5.6 : Two possible random trajectories for the bouncing ball
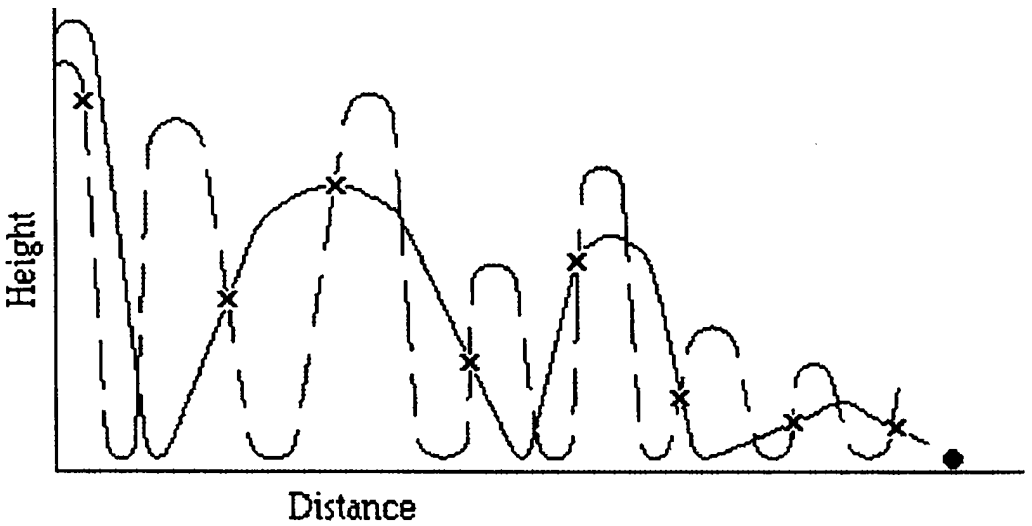
The remainder of this discussion of interpolation techniques will focus on methods that you are most likely to find in GIS. The discussion will focus on the interpolation of attribute data associated with point, line and area features. For the discussion of point interpolation methods, a clear distinction has been made between exact and approximate methods.

Continuous Self Assessed Exercise 13

(time: 30 minutes)

In the above discussion we have introduced a range of concepts, terms and techniques which we are certain will be new to many of you.

So far we have:-

>   local interpolator
>   global interpolator
>   exact interpolator
>   approximate interpolator
>   gradual interpolator
>   abrupt interpolator
>   deterministic interpolator
>   stochastic interpolator
>   Theissen polygons
>   kriging
>   Fourier analysis
>   trend surface analysis
>   moving average
>   line threading

Before you proceed we suggest that you cross off the list those that you understand. Keep the list close to hand as you proceed through the following sections and cross off terms and concepts as they are explained.

5.5 Exact point interpolation methods

5.5.1 Line threading or eye balling

This is the original, and still considered by some to be the best, method of interpolating a surface from a set of observed data. The process involves threading a line, by eye, between data points with the objective of enclosing points of similar value. In most cases, the person performing the line threading will use their local knowledge about the geographical area and their scientific knowledge about the phenomena they are studying to enhance the interpolation process. For example, in geological mapping, the presence or absence of a fault line may be known to the geologist but not evident in the point sampled data set. The geologist may be able to build this expert knowledge into the interpolation process, where the non-specialist would be unable to do so.

Three basic limitations of the manual line threading approach have led to the computerisation of the method. These are:

- the problem of handling a large number of points,
- the subjective nature of the procedure, and
- the time consuming nature of the process.

Many automated line threading methods now exist which perform the task with varying degrees of intelligence. Over the last few years knowledge engineering techniques have been used to capture the knowledge from 'local experts' and build expert systems for this type of interpolation.

Continuous Self Assessed Exercise 14

(time: 45 minutes)

Use the manual line threading or eye balling technique to interpolate a surface for the point data set in Figure 5.7. Use an interval of 10 units to produce a contour map. Don't throw this diagram away as you will be using it again later. To check how good your own eye balling is compare your map with the figure under the 'SAE Answers' button. Do not worry if your interpolated surface does not exactly match ours. The line threading approach is very subjective and though the main trends of the lines you have threaded should match with ours it is certain that the exact location of the lines will differ.

Figure 5.7 : Observed values for a set of irregularly spaced points
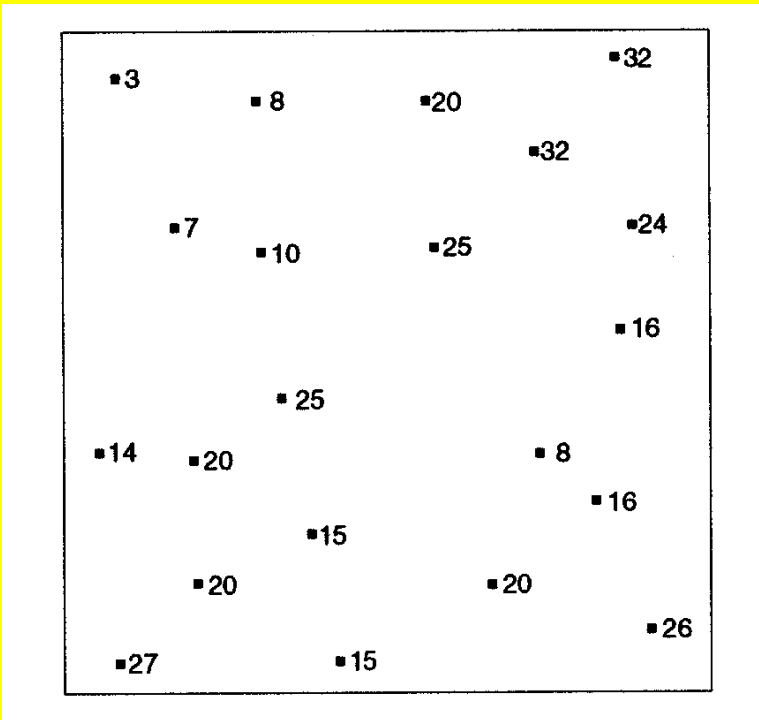
5.5.2 Theissen polygons

Theissen polygons are an exact method of interpolation that assumes the unknown values of the points on a surface to be equal to the value of the nearest known point. This can be considered as a local interpolator because the characteristics of the global data set have no influence on the interpolation process. The Theissen polygon technique was originally used for computing areal estimates from rainfall data. The approach is graphically represented in Figure 5.8 and involves spreading the territory associated with a point until the boundary associated with the territory of a neighbouring point is encountered. If you like you can imagine a bubble being blown outwards from each point simultaneously until they meet. Where they meet the boundaries of the interpolated areas are drawn. If the data points are irregularly spaced a surface made up of a mosaic of irregular polygons will result. The attribute value of each point is then adopted by its' associated polygon.

Figure 5.8 : Construction of Theissen polygons for a network of five points

Neighbouring points identified and territory markers established at the half way point along the line connecting pairs of points

Bisecting lines drawn through territory markers and a Theissens polygon mosaic established by linking the bisecting lines
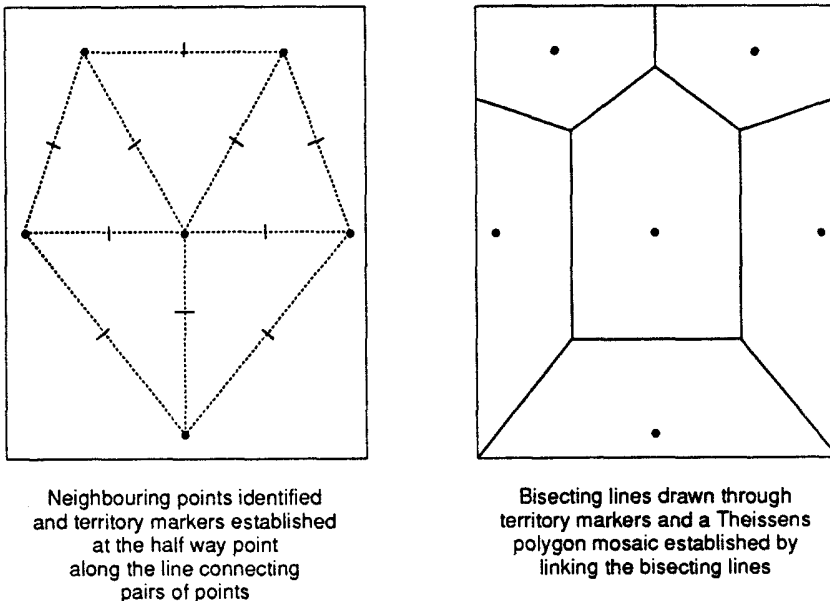
The method of Theissen polygons is a robust technique and will always produce the same surface from the same set of data points. This is a disadvantage in that the technique is unintelligent and unable to respond to external knowledge about factors which may influence the values recorded at the observed data points. One of the most common uses for this technique is for the generation of area territories from a set of points. For example, in the UK the smallest level of census administrative regions is the Enumeration District (ED). For each ED a geographical centroid (an (x,y) coordinate location) is recorded. The precise boundary of each ED could be captured by digitising the appropriate polygon boundary. However, an approximate boundary can be generated by using the Theissen polygon method to interpolate the area of influence associated with each ED centroid. The area polygons produced will not accurately represent the true boundary of the administrative area but may well be sufficient for the graphical presentation of data.

What your comparison of the two maps should reveal is that both maps have a rough similarity. The greatest discrepancy occurs at the boundaries between class intervals and at the edge of the study area.

5.5.3 The Triangulated Irregular Network (TIN)

You have already been introduced to the Triangulated Irregular Network (TIN) Model in **Spatial Data Modelling**. Developed in the early 1970's, the TIN is a simple way to construct a surface from a set of known points. It is a particularly useful technique for irregularly

spaced points. The TIN approach is an exact interpolation method. In the TIN model, the known data points are connected by lines to form a series of triangles. Figure 5.9 shows a TIN model for a set of 10 points. Because the value at each node of the triangle is known and the distance between nodes can be calculated, a simple linear equation can be used to calculate an interpolated value for any position within the boundary of the TIN.

Figure 5.9 : A simple TIN for a set of 10 points

To construct contour isolines from a TIN model you must first choose a contour interval. Then, identify all the arcs which would be dissected by a contour of a given value. To do this the computer must compare the values of the points which make up the nodes of a particular triangle. For example, in Figure 5.9 above, 9 out of the 19 arcs could be dissected by a value of 100. The next stage would be for the

computer to calculate the (x, y) coordinates of the point of intersection. Finally, the intersections are joined (see Figure 5.10). As you can see from Figure 5.10 one of the problems with producing contours from the TIN model is the angular character of the isolines produced. To overcome this some TIN models represent the surface in each triangle using a non-linear mathematical function chosen to ensure that the isoline changes continuously, not abruptly, at the edge of the triangle. The TIN model is also used to generate digital terrain models, which we consider in the next section of this module.

Figure 5.10 : Contouring a TIN
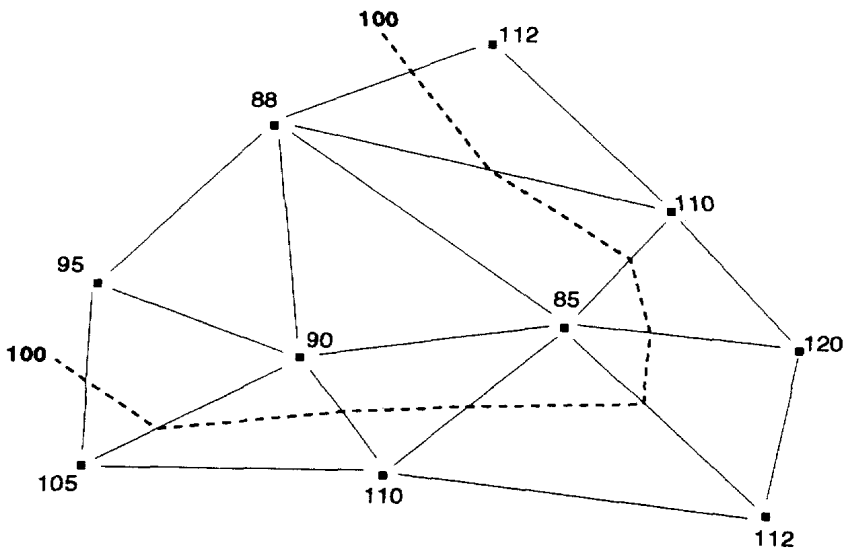
Source: After NCGIA (1989)

What should be apparent is that the isolines produced from the TIN method are much more angular than those produced by threading a line through the data points by hand. In addition, in the TIN model it is usual to interpolate a surface only within what is know as the 'hull' of

the data set. The hull is defined as the area enclosed by the arcs joining the sampling points which are on the periphery of the data set.

5.5.4 Kriging

Kriging, also known as the "Theory of regionalized variables", was developed by G. Matherson and D. G. Krigge as an optimal method of interpolation for use in the mining industry. Burrough (1986) notes that the origins of the method lie in the recognition that the spatial variation of many geographical properties is too irregular to be modelled by a smooth mathematical function.

The basis of kriging lies in estimating the average rate at which the difference between values at points change with distance between points. This is the most complex of all the exact interpolation methods and until recently was absent from virtually all commercially available GIS software. It has found wide application in the earth and environmental sciences, such as geology, where it is essential to both honour the known data points and take into consideration local variability. The complex nature of the technique and the prerequisites in terms of statistical knowledge which are necessary to understand its application are too detailed for us to pursue here. If you are interested in learning more about kriging a comprehensive and readable account is provided in Oliver and Webster (1990).

5.6 Approximate point interpolation methods

Approximate interpolation techniques are applicable when there is uncertainty associated with reproducing the measured values for the observed data points but where it is known that there are global trends overlain by local fluctuations (Waters, 1989). Approximate interpolation techniques, therefore, employ smoothing techniques which reduce the effects of local variability on the interpolated surface.

An important characteristic of all approximate point interpolation techniques is that the surface being interpolated does not have to pass through the known data points.

5.6.1 Spatial moving average

The most common method used in GIS is the Spatial Moving Average. This process involves calculating a new value for each known location based upon a range of values associated with neighbouring points. The new value will usually be either an average or weighted average of all the points within a predefined neighbourhood. In some textbooks you will find that they use the term 'moving window'.

Applying a moving average involves making a number of decisions about the shape, size and character of the neighbourhood about a point. The most common shape for a neighbourhood is a circle, since points in all directions have an equal chance of falling within the radius of a given point. However, in the raster world a square or rectangular window of cells is often used. The size of the neighbourhood is determined by the user and is usually based upon assumptions about the distance over which local variability in the data is important.

One weighting function which may be used to enhance the technique is the distance of points away from the centre of the neighbourhood. For example, a point which is only 100m away may be considered to be twice as important as another point 200m away. This is described as a linear weighting. It is also possible to have a non linear weight where points closer to the centre of neighbourhood are considered to be one or two orders of magnitude more important than those towards the edge of the neighbourhood. The number of values used in the calculation is also important. Burrough (1986), for example, suggests a spatial moving average should use between 4 and 12 data points with the norm being between 6 and 8. Including too many points creates a

very smooth surface, ultimately resulting in a single mean value for the entire study area. Conversely, too few points can result in individual points with extreme values dominating the interpolated surface.

Once a mathematical model has been constructed to describe the character of the neighbourhood the next stage is to calculate a new value for each of the known data points. Figure 5.11 summarises this process for a neighbourhood in which each point is equally weighted.

Figure 5.11 : Calculating a spatial moving average in the raster vector world

$$\frac{3 + 7 + 5}{3} = 5$$

$$\frac{3 + 7 + 5}{3} = 5$$

 The basic problem with spatial moving average techniques is the number of variations that exist. It is, therefore, highly probable that the spatial moving average functions available in one proprietary GIS will produce significantly different results from those in another GIS. This is because different systems will allow you to select different distance decay models, select the number of points used in the model, and apply different procedures for handling those points which fall on the edge of a neighbourhood. In addition, in the raster world, it is more common for the neighbourhood to be defined as a group of cells (usually 9) compared with the circular neighbourhood used in the vector world. These problems do not arise if you have a complete understanding of

the phenomena which you are trying to interpolate because you can design a 'perfect' moving window.

The spatial moving average approach is best suited to those situations where:-

1 - observations are not considered to be exact representations of a true surface, but rather singular observations within a statistical sample
2 - where the measurement techniques used to measure values at discrete locations on the true surface are known to be subject to error
3 - where it is known that the true surface exhibits a general pattern within which there is local variability

Examples include demographic data, survey or questionnaire responses and pollution monitoring observations. A spatial moving average will produce a surface which suppresses individual values and reveals broad structures in data. The nature of those structures depends to some extent on the filter size and the weightings, which you determine based on known objectives of the analysis.

5.6.2 Trend surface analysis

To introduce the concept of trend surface analysis let us take a look at what it says next to the appropriate command from the IDRISI help:

> 'TREND calculates linear, quadratic and cubic trend surface polynominal equations for spatial data sets and interpolates surfaces based on these equations'.

If this sounds like a statistical nightmare do not worry, the ideas are simple. All we have time to do here is introduce the basic principles. In addition, because the statistical assumptions of the model are rarely met in practice the technique is not a very robust interpolation method. Therefore, it is much safer to use one of the simpler methods described elsewhere in this section. For those of you interested in getting to grips with trend surface analysis in detail Unwin (1981) provides an excellent introduction which includes a worked example based around a set of five points.

For those of you who have done some basic statistics then the concepts which underpin trend surface analysis should be immediately apparent. If you have not done any statistics then the following discussion should help. The trend surface method fits a surface to the known observation points by a method known as least squares. The shape of the surface is determined by the mathematical equation, or polynomial, which is used to describe it. A polynomial surface has some very important characteristics, most importantly it changes smoothly and predictably from one map edge to the other. The polynomial equation may be linear, quadratic or even cubic. A linear equation would be used to describe a tilted plane. A quadratic equation would define a simple hill or hollow and a cubic equation would define a more complex surface (Figure 5.11). These concepts should be familiar to anyone who has calculated regression equations ('lines of best fit').
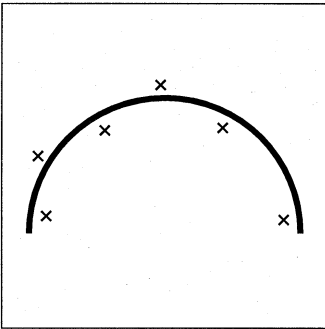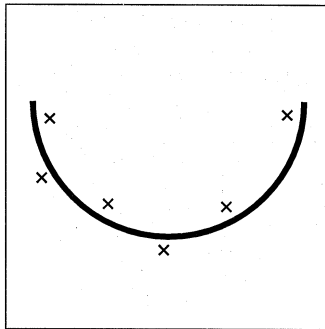
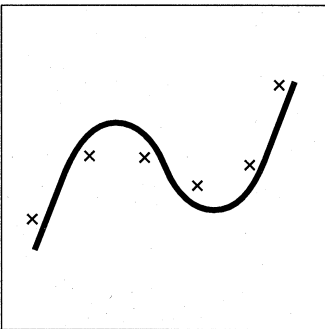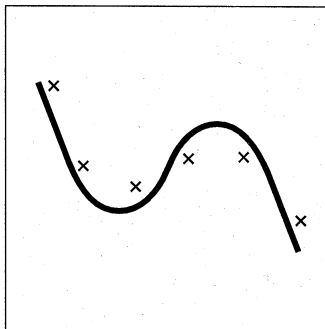**Figure 5.11**
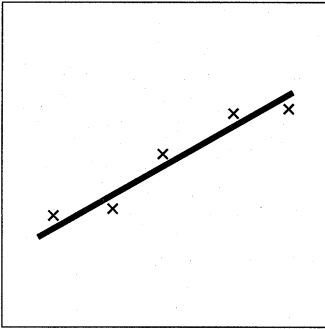
Linear
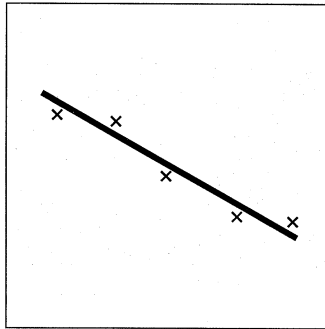
Linear

Quadratic

Quadratic

Cubic

Cubic

To fit a simple tilted plane to a surface of observed points all you need to do is solve a simple linear regression equation in which the spatial coordinates (x, y) are the independent variables, and the z value, the attribute of interest, the dependent variable. For those of you who are not familiar with regression what this statistical procedure does is fit what is often termed a ' best fit line' through observed data values. It is easiest to understand by way of a two dimensional example. Graph A in Figure 5.12 is known as a scatter plot and shows how height varies with distance along a transect. If we were certain that these height values were exact then we might join the points together as in Graph B to interpolate the height value at unsampled locations.

However, if we are uncertain of the quality of our data, or we are more interested in the general trend, then it might be safer to place a line of best fit through our points (see Graph C).Conceptually, all trend surface analysis does is fit a similar 'best-fit surface' to the known data points based on a mathematical definition of fit, i.e. least squares. Trend surfaces are, therefore, smoothing functions, which do not necessarily pass through the original data points.

The main limitation of trend surface analysis is that it makes adjustments for all points simultaneously. I would suggest that these techniques are more appropriate for exploratory data analysis than for interpolation.

5.6.3 Fourier analysis

Fourier analysis is another global interpolation technique which uses applied geostatistical methods to approximate a surface by overlaying a series of sine and cosine waves. The technique was developed for signal analysis and has been widely applied in climatic change modelling. At present it has received little attention as a tool for interpolating geographical phenomena. Burrough (1986) suggests this

is probably because it is best suited to data sets which exhibit periodicity (such as ocean waves and sand dunes) which are rare in most geographical phenomenon. Therefore, because of the lack of practical applications using this technique and the complex mathematical nature of the approach we will only mention it here. If your inquisitive nature gets the better of you then both Burrough (1986) and Waters (1989) suggest Davis (1986) as a readable introduction.

Figure 5.12 : The best fit concept
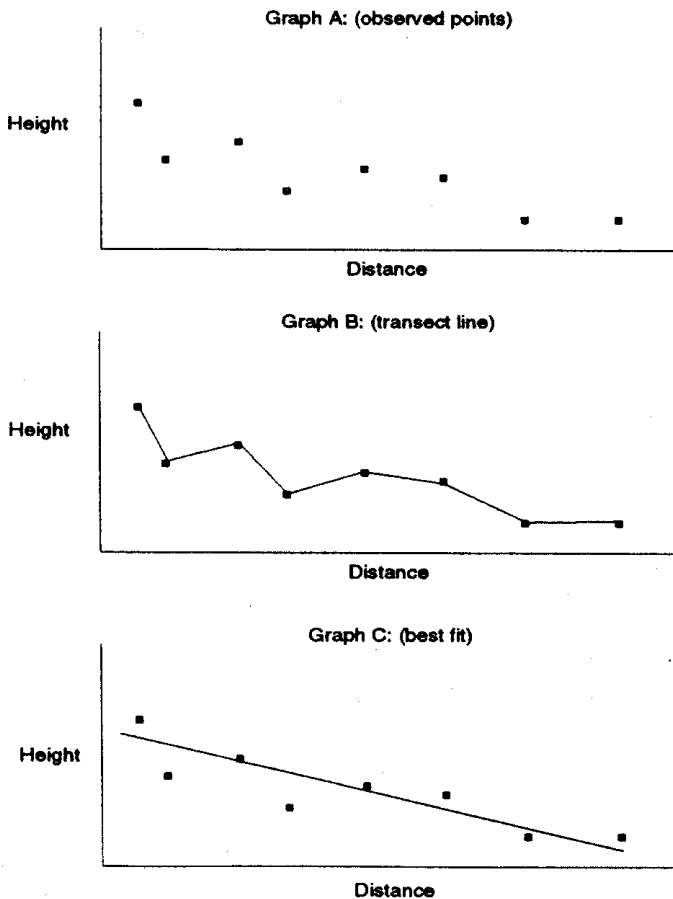
5.7 Line interpolation methods

You have already examined the most common method used to interpolate observed data associated with a linear feature. The technique known as buffering or corridor analysis was described in Section 2.4.1 of this module. Buffering is an exact, abrupt and deterministic interpolation technique which produces a surface to which an attribute can be applied which is some derivative of that associated with the linear feature.

Buffering is an exact interpolation technique because the procedure honours the observed values associated with the linear entity from which it has been derived. It is abrupt because there are exact boundaries between the surface territories associated with the line features used to produce the buffer (though these will be less marked if a graded buffering approach is used to generate a surface). The technique is deterministic in that it assumes a linear feature will have a known fixed influence on its neighbourhood. The interpolated surface and subsequent values produced using buffering hold true only when the character of the linear feature is known to be the dominant feature in the area of interest.

Another common interpolation operation found in many GIS is estimating point values from isolines. The most common instance of this application is the conversion of height contour lines into an elevation grid for the purpose of constructing a digital terrain module. Since digital terrain models are now an integral part of most GIS we will leave our discussion on the interpolation of point data from linear features until later in this module.

5.8 Area based interpolation methods

Most socio-economic and environmental data are reported by reference to an areal unit. Examples include: population density by country; the number of employed or unemployed by administrative district; crime statistics by police beat and water quality measures by water districts. Within the UK alone there are some 70 different areal units which are used for reporting socio-economic data. Area based interpolation techniques are necessary when we wish to allocate the attributes associated with one set of area features to another. For example, in the UK we might wish to report crime statistics for police beat at electoral ward level. As Laurini and Thompson (1992) indicate, assigning a value is easy if either:-

1. the area units are an identical match, or
2. the source areas are a nested subset of the target area.

It is when the areas do not match that interpolation is necessary and problematic.

Waters (1989) summarises the areal interpolation problem as follows:

> "Given a set of data mapped on one set of source zones determine the values of the data for a different set of target zones."

Burrough (1986), Waters (1989) and Lam (1983) identify two different approaches to solving the problem:-

- non-volume preserving
- volume preserving

which will be considered in turn.

5.8.1 *Non-volume preserving*

The non-volume preserving method is best described by way of example. The following example is based loosely around that described by Waters (1989). Consider the two maps in Figure 5.13. Map A shows the total population for four administrative districts on the planet Mars. Map B shows the flood corridor which has been established for one of the Martian built canals on the planet. What is the population density within the flood zone?

Figure 5.13 : Non-volume areal interpolation method

Map A: Administrative boundaries with total Martian population counts

A 200	B 400
D 600	C 800

Map B: Flood hazard map for the canal

canal — — — — — — — —   flood hazard zone
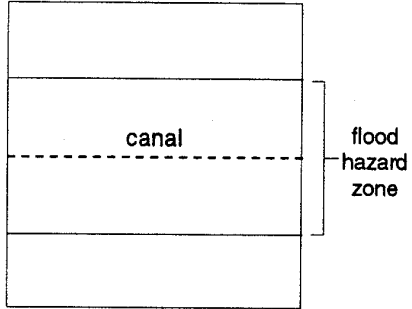
Map C: Population density counts

Area each cell 100 M units

■2	■4
■6	■8

Map D: Interpolated population density grid

2	3	4	4
3	4	5	5
4	5	7	6
6	6	7	7

Map E: Estimated total population for grid

50	75	100	100
75	100	125	125
100	125	175	150
150	150	175	175

Map F: Total Martian population counts for flood zone

325
975
650

The first step is to calculate the population density for each district by dividing population by area. Assuming we know the area of each district to be 100 m units. This gives us a population density of 2, 4, 6 and 8 Martians/m unit for zones A, B, C and D respectively.

The next stage is to identify a centroid for each region (this is the point at the geometric centre of the area) and then assign to this location the population density value associated with its enclosing area. This would give us Map C in Figure 5.13. The next stage would be to interpolate a gridded population density surface using any of the methods described earlier. Map D provides one possible surface which may result from such a process.

Each grid cell value can then be adjusted to a population by multiplying the estimated density by the cell area (see Map E). Now, overlay the target map. In this case our flood corridor goes on top of the interpolated grid and then each grid value can be assigned to the flood and non-flood zones. From this map an estimated value of the total number of Martians at risk in the flood zone can be calculated (see Map F).

<div style="background-color: yellow; border: 2px solid black;">

Continuous Self Assessed Exercise 17

(time: 20 minutes)

Taking a careful look at the above example can you identify the weaknesses in this approach? Hint: how does the total population across all zones compare between Map A and Map F?

</div>

The most important point which we hope you will have identified from the exercise above is that the total population of the whole region is not preserved. In our example, the total population has dropped from 2000 in Figure 5.13A to 1950 in Figure 5.13F. Another criticism of this method is that the choice of the centroid as the point location on which to base an interpolation technique may be a poor assumption. In our example there may well in fact be nobody living at this central location and all the Martians might live at the edge of zone A. A further problem is that this approach suffers from all the inadequacies associated with point interpolation methods as described earlier.


5.8.2 Volume preserving

Volume preserving procedures involve overlaying the target zones on top of the source zones and then determining the proportion of the target zone which falls within each source zone. The total value of the attribute for each source zone is then given to the target zone according to the areal proportions. Figure 5.14 summarises this approach for estimating the total Martian population in our flood zone. The main problem is that the method assumes uniform density of the attribute within each zone. An enhanced version of this approach is the Pycnophylactic method developed by Tobler in 1979.

Figure 5.14 : Volume preserving areal interpolation methods

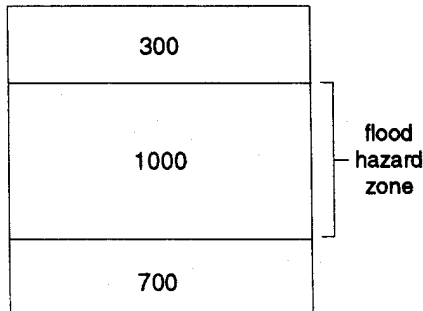Map A: Administrative boundaries with total Martian population counts

Map B: Flood hazard map for the canal

A 200	B 400
D 600	C 800

canal ---- flood hazard zone
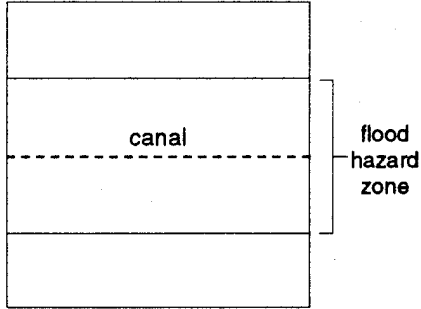
Map C: Area ratio population counts

100	200
100	200
300	400
300	400

Map D: Estimated total population counts for flood zone

300
1000
700

flood hazard zone

This particular interpolation procedure raises an interesting question. What do you do about the cells whose neighbours are external to the study area and therefore unknown. Do you:-

- assume the population is equal to zero, or
- assume the population density is equal to the last neighbour?
- simply ignore them as we did above?

There is no simple answer to this question. In this case it is necessary to rely on your own judgement or external information to help you assess the situation and determine the appropriate action. Here another interesting problem arises, since when we are performing this

operation by hand it is possible for us to assign boundary values as we see fit. For example, if we know there is a desert to the top of our study area we can build zero values into the cell averaging procedure. However, if this process was being performed by a computer algorithm this information would have to come from another source? What other source? Artificial intelligence or another map layer perhaps? At the present time most commercial GIS have no simple answer to this problem.

5.9 Discussion

In this section we have provided an overview of the objectives, importance and applications of a range of point, line and area interpolation techniques. As with so many other topics in this diploma, we could quite easily have written a whole module on this topic. However, our purpose is to provide you with a flavour of the subject and a rational for why, where and when you might use some of the more common methods.

You probably now feel that selecting and applying the appropriate interpolation technique to a given GIS problem is much more complex than you had first considered. Unfortunately, there are few hard and fast rules that can be applied to selecting the appropriate interpolation procedure. Where possible we have tried to identify and isolate these in the above discussion (these are also summarised in Table 4 which is reproduced from Burrough (1986)).

Perhaps the best piece of advice we can give is that you should try to ensure that you understand the data you are using and that you are aware of its characteristics and limitations. Remember though that even the most rigorous procedure is still only providing you with a 'best guess' - the only exact method is to go out there and sample each and every location!

Table 4 : Methods of interpolation (Burrough, 1986)

Method	Deterministic/ stochastic	Local/ global	Transition abrupt/ gradual	Exact interpolator	Limitations of the procedure	Best for	Output data structure	Computing load	Assumptions of interpolation model
'Eyeball'	Subjective/ deterministic	Global	Abrupt	No	Non-reproducible, Subjective	Field data, aerial photo interpretation	Polygons	None	Intuitive understanding of spatial processes; homogeneity within boundaries
Edge-finding algorithm	Deterministic	Global	Abrupt	No	Often requires shapes to be defined and stored; better for man-made features than for natural landscapes	Raster images from remote sensors	Raster	Moderate	Homogeneity within boundaries
Proximal (Theissen poly.)	Deterministic	Local	Abrupt	Yes	One data point per cell; no error estimates possible; tesselation pattern depends on data point distribution	Normal data from point patterns	Polygons	Light/ Moderate	Nearest neighbour' gives best information
Fourier series	Stochastic	Global	Gradual	No	Edge effects, outliers, complex polynomials do not necessarily have meaning; errors are rarely spatially independant	Demonstrating broad features and removing them prior to other methods of interpolation	Points on a raster	Light/ Moderate	Multiple regression-phenomenological explanation of trend surface; independent Gaussian errors
Trend surface	Stochastic	Global	Gradual	No	Not applicable to data; lacking periodicity	Periodic features such as sand dunes, ripple marks or gilgai, or man-made features	Points on a raster	Moderate	Strict periodicity in phenomenon of interest
B-splines	Deterministic	Local	Gradual	Yes	No estimates of errors, masks all uncertainties in surface	Very smooth surfaces	Points on a raster	Light/ Moderate	Absolute smoothness of variation
Moving average	Deterministic	Local	Gradual	No unless constrained	Results depend on configuration of data points and size of window, sample versions assume isotropy; no error estimates unless retrospectively calculated	Quick contour plots of moderately smooth data	Points on a raster	Moderate	Continuous, differentiable surface is appropriate
Optimal interpolation (kriging)	Stochastic	Local	Gradual	Yes	Practical and theoretical problems of non-stationarity in data; large computing costs for mapping	Situations where the most detailed estimates and their errors are required	Points on a raster	Heavy (very heavy for universal kriging	Intrinsic hypothersis (homogeneity of first differences) average local values can be represented by a continuous surface

If you are intrigued and enthralled by the complexities surrounding spatial interpolation procedures and wish to delve further then the bibliography at the end of this module should provide you with an appropriate starting point. If you have access to a GIS other than the one you are using as part of this course an interesting exercise would be to undertake a review of the spatial interpolation procedures which it contains. At the present time it will probably make use of one or a few of the methods described in this section. The more analytically intensive techniques are often absent from proprietary GIS because there has been no demand for them. It may be that until somebody is able to develop a usable interface to these more sophisticated procedures it is likely they will remain in the domain of the spatial scientists.

So, after all that, why have we spent a large part of this module discussing interpolation techniques ? The answer lies in the fact that even if you aren't required to interpolate data in your job then it is still highly likely that you will be using data that have been interpolated by someone else. If this section has achieved nothing else then it should have warned you of the need to understand the methods which have been applied to those data.

Before completing the final exercise I would like to draw your attention to the difference between **interpolation** and **extrapolation**. All of the interpolation techniques that we have described have been developed to estimate values at locations within the area covered by the source data. Estimating values beyond those boundaries (extrapolating) is more difficult and prone to greater uncertainty because there is less `control' on the interpolation technique. You will also appreciate that not all techniques can be used for extrapolation. Strictly (statistically) speaking none of them should be used but we live in a world where sometimes we are asked to perform analysis against our better judgement.

We will return to some of these issues in the **Data Quality** module when we look at error propagation.

Periodic Self Assessed Exercise 5

(time: 30 minutes)

To help you summarise this section try to define and describe what is meant by the following terms in the context of interpolation:-

1. Local/Global
2. Deterministic
3. Theissen polygons
4. Volume preserving area interpolation
5. Exact/Approximate
7. TIN
8. Eye balling
9. Neighbourhood