Spatial Interpolation

N. S. Lam, Louisiana State University, Baton Rouge, LA, USA

Glossary

Areal Interpolation A mathematical procedure used to interpolate data values defined in one set of aggregated boundaries to another.
Kriging A spatial interpolation method named after D. G. Krige who introduced the method in the early 1970s.
Point Interpolation A mathematical procedure used to estimate values from a set of sample points with known values to locations with unknown values within a study area.
Pycnophylactic Interpolation An interpolation method introduced by W. R. Tobler in late 1970s for volume-preserving isopleth mapping.
Spatial Interpolation A mathematical procedure used to estimate values from locations with known values to locations with unknown values within a study area.
Variogram A diagram depicting the relationship between variance and distance for a set of sample points.

Introduction

Spatial interpolation is used to estimate values at locations whose values are unknown based on a sample of locations with known values. Before the 1980s, the spatial interpolation problem was confined mostly to point interpolation, where the sample data are in the form of point locations. Goodchild and Lam expanded the traditional notion of spatial interpolation by incorporating areal interpolation, where data defined in one set of areal boundaries (polygons) are used to estimate values of areas defined using another set of boundaries. Together, point and areal interpolations have since become a major tool for many scientific investigations. Hence, the spatial interpolation problem can be broadly defined as follows: given a set of spatial data either in the form of discrete points or aggregated subareas, the spatial interpolation problem is to find the function that best represents the whole surface so that prediction of values at other points or subareas within the study area whose values are unknown can be obtained.

Spatial interpolation is a crucial technique in many disciplines and has been utilized in a wide range of applications. Perhaps the most common and earliest use is in cartography, where point interpolation is employed to create a continuous topographic surface from sampled elevation points. The interpolated topographic surface is usually mapped in the form of contour lines or three-dimensional block diagrams. In climatology, spatial interpolation is an essential routine for estimating temperature and other climate parameters for the entire globe from a network of weather stations. Spatial interpolation techniques have often been used to estimate air pollution parameters such as SO_2, radon gas, and particulate matter (PM) concentrations using samples from monitor stations. In physical geography and geological applications, spatial interpolation is the key method for estimating subsurface geological strata and mining potentials. Similarly in human and economic geography, spatial interpolation is frequently applied to estimate population density, market potential, income potential, health parameters, voting behavior, and crime patterns.

With the development of geographical information systems (GIS) technology, the interpolated surface data can be stored in digital form called digital elevation models (DEM) or digital terrain models (DTM) (even though the variable of interest is not necessarily elevation). These digital data can further be used to calculate slopes, aspects, and volumes, and/or integrated with other data for more comprehensive analysis such as landscape planning, visibility assessment, route planning, and cut-and-fill engineering projects. This infusion of GIS and digital technology has opened up lots of application potential and makes the interpolation process even more indispensable and popular than before.

As in many other spatial estimation methods, spatial interpolation is subject to error, the magnitude of which depends on a number of factors. First and foremost, the mathematical function used in every spatial interpolation model to estimate unknown values is only an assumption; it may not be the best in approximating the underlying surface, thus resulting in error. Many interpolation methods have been developed, as well as refined and extended, in an attempt to find the best fit function that represents the underlying surface so as to increase interpolation accuracy. The second major source of error arises from the fact that spatial interpolation deals with spatial data that have x and y dimensions; the nature of the problem is thus different from nonspatial interpolation such as interpolation of time-series data. The spatial interpolation process will need to consider spatial proximity or neighborhood effects. The added dimension makes the interpolation process more complicated and subject to ambiguity, and factors other than the mathematical function itself, such as the grid size used during the interpolation process and the spatial distribution of samples, will have great impacts on the resultant interpolation error.

With the availability of spatial interpolation routines in GIS and other mathematical software, spatial interpolation nowadays is simple to use and has been frequently applied. However, many people perform spatial interpolation without realizing that there could be significant errors involved in the estimates. Moreover, some methods may yield more accurate results than others for different types of data surfaces. An understanding of the major methods in spatial interpolation and their error sources should help in improving the accuracy of the interpolated estimates, as well as raising the awareness of the errors involved in subsequent analysis that is based on the interpolated estimates.

In the following, a classification of spatial interpolation methods is described, which should help in putting the various methods in perspective. Then, major methods in each class are briefly explained. Factors affecting the interpolation accuracy are outlined, followed by some suggestions toward future research.

A Classification of Spatial Interpolation Methods

There are many methods for spatial interpolation, and a classification of these various methods according to some schemes is necessary to better understand the method itself and its relationship with the data. Using Lam's classification framework developed in 1980, spatial interpolation includes both point and areal interpolation (Figure 1). Point interpolation deals with data collectable at a point, such as temperature readings or elevation, whereas areal interpolation deals with data aggregated according to some boundary delineation, such as population counts by census tracts.

Point interpolation is further divided into 'exact' and 'approximate' methods according to whether they preserve the original sample point values. Exact methods are employed when there is high confidence of the accuracy of the original sample data, whereas approximate methods are preferred when the original sample data are thought to have some noise. Other classification schemes for point interpolation have been used in the literature. For example, some classify the methods according to the spatial extent into either 'global' or 'local' methods. Global methods utilize all sample points in determining value at a new point, whereas only nearby points are used in local methods. Another classification scheme distinguishes point interpolation methods according to whether it is a statistical technique or not. As such, Kriging is in the group of statistical technique, and the remainder of point interpolation methods is in another group. This article adopts the classification that categorizes methods into 'exact' or 'approximate', because the property of preserving the original sample point values on the estimated surface seems fundamental in analyzing accuracy and in examining the nature of interpolation methods.

Common examples of the methods of the 'exact' type include interpolating polynomials, most distance-weighting methods, Kriging, spline interpolation, and finite difference methods. Examples of the group of 'approximate' methods include power-series trend surface analysis, Fourier models, distance-weighted least squares, and least-squares

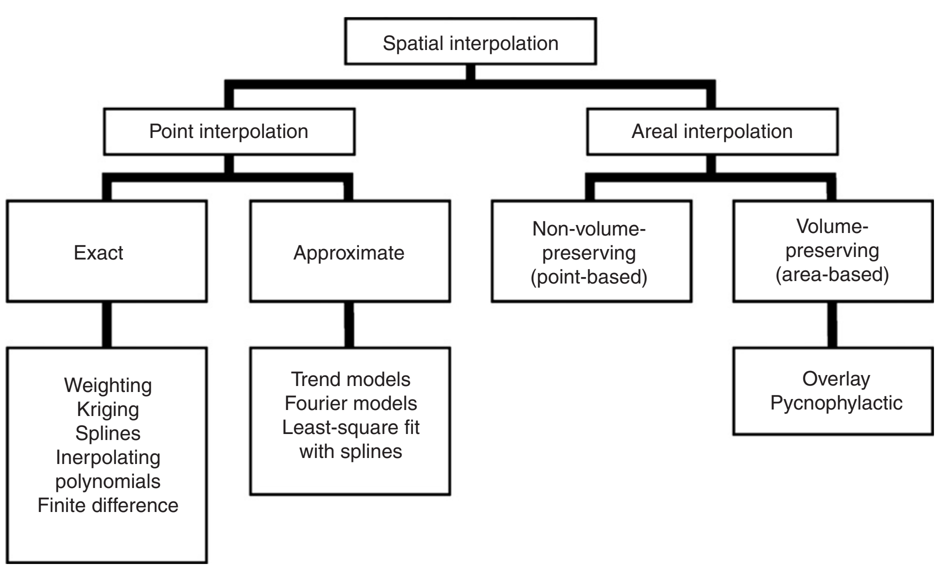

Figure 1 A classification of spatial interpolation methods. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

fitting with splines (Figure 1). These examples reflect the most generic version of the method itself. It is noted that there are many variations and extensions of the generic versions, with some methods being a combination of different basic methods. Moreover, other criteria can be used simultaneously to help understand a method. For example, an 'exact' method can also be global or local, statistical or nonstatistical.

For areal interpolation, two different approaches can be used. Following the terminology by Goodchild and Lam, the geographic areas for which data are available are called source zones and those for which data are needed are called target zones. The non-volume-preserving approach, which is an older approach, does not preserve the original source zone data values, whereas the newer approach, the volume-preserving approach, does. Again, similar to point interpolation, the preservation of original source zone values is considered an important property for accuracy assessment. It has been shown in subsequent research that those methods that can preserve the original source zone values are far more accurate than those that cannot. Examples of non-volume-preserving methods include a number of point-based interpolation methods, whereas examples of volume-preserving methods include polygon overlay and pycnophylactic interpolation. Again, these are basic versions of the methods, and variations and extensions of these two methods have been made to compensate the shortcomings existed in each method.

Point Interpolation

Grid versus TIN

Since sample points are most likely irregularly spaced, prior to point interpolation, a decision is needed to make on whether the interpolated values will be in the form of regular grid cells or a triangulated irregular network (TIN) (Figure 2). For the 'grid' approach, a mesh of grids will first superimpose on the study area. A point interpolation method is then applied to interpolate from irregularly spaced sample points to regularly spaced grid centers or intersections. The size of the grid is often arbitrarily determined, which has great impact on the accuracy of resultant grid estimates, as discussed below.

The TIN approach links the irregularly spaced sample points into a set of triangles. Interpolation will be directly made on any point along the edges of the triangles. This approach was frequently used to produce contour maps and shaded relief maps, as it is more direct and generally faster in computation. However, inconsistencies arise because there are many ways of making the triangulation of the same set of points. Furthermore, for different sets of points, different TINs will result. This makes the TIN approach less desirable

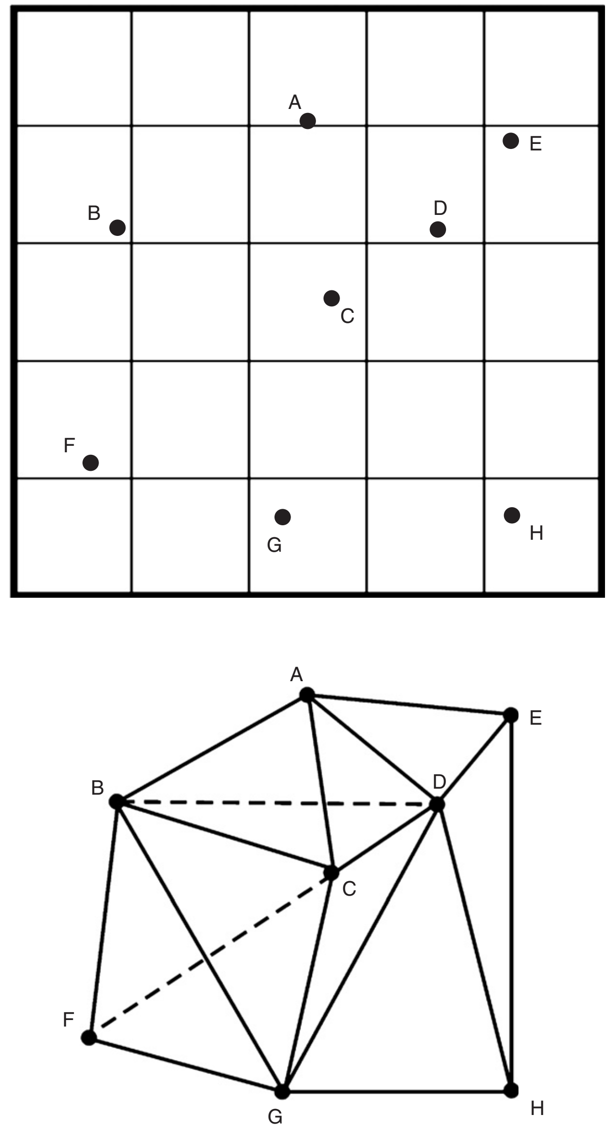

Figure 2 Grid vs. TIN configurations for eight sample points. Note that TIN configuration is not unique; alternative triangles can be connected between two points (e.g., dotted line $\overline{FC}$ instead of $\overline{BG}$). Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

because of the difficulty to compare from one surface to another. Integrating the interpolated surface data with other types of data in a GIS setting using TIN configuration would be impossible. The grid approach enables comparison of different surfaces by using the same grid size; hence, it is often preferred for point interpolation.

Exact Methods – Distance-Weighting

Distance-weighting method is probably the most commonly used point interpolation method, as it is flexible, easy to understand, and easy to use. Given a set of N data points, the distance-weighting method is to assign more weight to nearby points than to distance points. This follows the logic of spatial autocorrelation, spatial dependency, and the Tobler's law of geography, where things close by are more related than things located farther apart. The general mathematical expression for computing the value at a grid center (x, y) is

$$f(x,y) = \frac{\left[\sum_{i=1}^{N} w(d_i) z_i\right]}{\left[\sum_{i=1}^{N} w(d_i)\right]} \quad [1]$$

where $w(d)$ is the weighting function, z_i is the data value at point i, and d_i is the distance from point i to (x, y).

The main variation for this method is the weighting function. Unfortunately, the choice of a weighting function is often arbitrarily made, and it may or may not represent the underlying surface well. Some weighting functions will place far more weights for the points nearby and very little weights for the points farther, resulting in a more varying surface, such as the family of inverse distance functions (e.g., $w = d^{-2}$). Other weighting functions may place a less drastically reduced weight for distant points, such as a linear weighting function (e.g., $w = a - bd$), thus resulting in a smoother surface. It is important to note that the weighting method can be 'exact' or 'approximate', though most of the functions used in geographical applications lead to 'exact' results. For those weighting functions where $w(0) = \infty$, such as $w = d^{-1}$, the weighting method will give the exact value of the original sample points. On the other hand, for a negative exponential weighting function, such as $w = 2e^{-d^2}$, the method will only approximate the original values. Moreover, the distance-weighting method can be made either a 'local' method, where only nearby data points are used to estimate the value at the grid center, or a 'global' method, where all data points in the study area are utilized to estimate a grid value.

The simplicity and flexibility of the distance-weighting method make it a very popular method for point interpolation. However, in addition to the problem of selecting the appropriate weighting function, the distance-weighting method is subject to two additional disadvantages. First, the method is easily affected by uneven distributions of data points, since an equal weight will be assigned to each of the points even if it is in a cluster (Figure 3). This problem has long been recognized and can be handled by additional algorithmic manipulation, such as averaging all the points or selecting a single point to represent the cluster. Second, as a weighted average method, it has a smoothing effect, and the resultant interpolated values will be bounded by the maximum and minimum values of the samples. This means that if the original sample points do not capture the underlying true maximum and minimum surface values (e.g., mountain top and valley bottom), then the interpolated surface will never reach these maxima or minima.

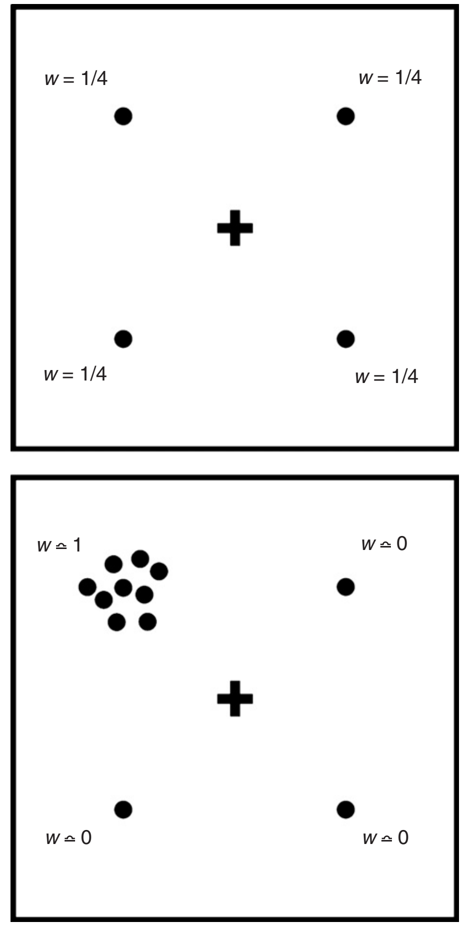

Figure 3 Effects of clustered data points on distance-weighting methods. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

Exact Methods – Kriging

The term Kriging is named after D. G. Krige, who introduced the use of the technique for mining applications in the early 1970s. Since then, Kriging has been extended from simple Kriging to other variations such as universal Kriging, block Kriging, co-Kriging, and others. Applications of Kriging have been increasing, especially now that the technique is available in popular GIS and statistical software packages. Kriging is a statistical technique, which is different from the rest of the point interpolation methods. It is the main method in the field of geostatistics.

In its simplest form, Kriging assumes that the surface to be interpolated is a statistical surface that has a certain degree of stationarity. The structure of the surface, also termed regionalized variable in Kriging literature, can be represented by its variogram, which is a function

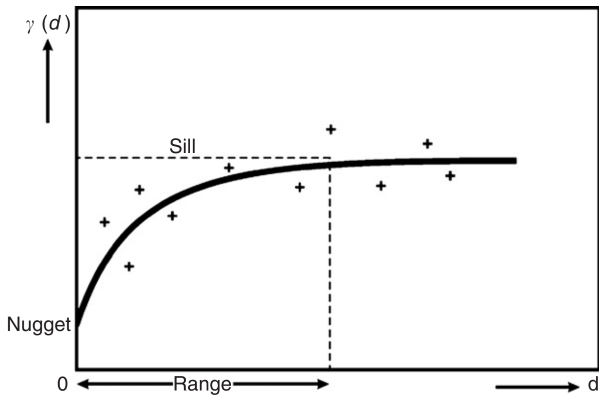

Figure 4 A hypothetical variogram function estimated from observed data points. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

depicting the relationship between the variance of the sample points and their distance (Figure 4). Mathematically, the variogram (2γ) or semivariogram (γ) is defined by

$$\gamma = \frac{1}{2}N\sum_{i=1}^{N}\left[z_{(x_i+d)} - z_{(x_i)}\right]^2 \qquad [2]$$

where d is the distance between two sample points. This function (variance) is expected to increase with increasing distance between two points. But just like many geographical phenomena, the increase is expected to taper off at a distance 'range', reaching a level called 'sill'. Sometimes, there is a small nugget effect, where samples taken from the same location (zero distance) may not have exactly the same value.

Based on the empirical estimated variogram function, points that have unknown values can be estimated as a weighted average of the sample point values according to their distances to these points. The system of linear equations to solve for the weights λ_j for an unknown point A is

$$\sum_{j=1}^{n}\lambda_j\gamma(d_{ij}) + u = r(d_{\mathrm{iA}}) \quad \text{for all } i = 1, n \qquad [3]$$

$$\sum_{i}^{n}\lambda_i = 1 \qquad [4]$$

where $\gamma(d_{ij})$ is the semivariogram function value for the distance between sample point i and j, $\gamma(d_{\mathrm{iA}})$ is the semivariogram function value between point A and sample point i, and μ is the Lagrange multiplier. Once the λ's are found, the Kriging estimate of an unknown point A is

$$z_{\mathrm{A}}^{*} = \sum \lambda_i z_{(x_i)} \qquad [5]$$

This is an estimate that is derived using two criteria: (1) that it is an unbiased estimate; and (2) that it minimizes the expected variance. Hence, the Kriging estimate obtained is also called the best linear unbiased estimate (BLUE). The corresponding error provided for each estimate is the Kriging error. For points that belong to the set of samples, Kriging returns the original data values; hence, it is an 'exact' interpolation method.

Although Kriging is a weighted average method, it has several major advantages over the distance-weighting method. First, unlike the distance-weighting method where the weighting function is arbitrarily determined, Kriging uses the variogram estimated from the empirical data. The more the sample points, the more accurate the empirically derived variogram function, hence the more accurate the estimates. Second, as a statistical technique, Kriging provides an error estimate and confidence interval for each unknown point. This error information reflects the density and spatial distribution of sample points and the degree of spatial autocorrelation within the surface, and therefore is very useful in analyzing the reliability of the Kriged estimates. This proves to be very useful for future improvement of the estimates by showing where more samples are needed. Finally, the Kriged estimates are not bounded by the sample data minimum and maximum.

On the other hand, if the assumption of stationarity is not met and if the variogram function is inaccurate, then the advantages of Kriging will diminish. Subsequent empirical studies have shown that Kriging will perform the best with a large number of sample points. Moreover, even with a small number of data points, Kriging will still return with more accurate estimates than distance-weighting methods. Hence, Kriging is considered to be the preferred point interpolation method over the remainder, and it should be used whenever possible.

Approximate Method – Trend Models

This method is to find the best fit polynomial function that assumes values at the data points approximately but not generally equal to the observed values. Thus, there will be an 'error' or residual at every data point. The general equation is

$$f(x,y) = \sum_{i,j=0}^{m} a_{ij}x^i y^j + e \qquad [6]$$

where a_{ij} are the coefficients to be estimated from observed data point values, e is the residual, and m is the degree of the polynomial, which is less than the total number of points, N. These methods are called trend surface models since they are often used to simplify the surface into a major trend with associated residuals. Once the coefficients are estimated, they can then be used to determine values at other points. This method is highly affected by the extreme values and uneven distribution of data points. Moreover, the goodness-of-fit of the polynomial (i.e., the r^2 value) is not a true estimation of the error, but rather an underestimation (Figure 5).

Factors Affecting Point Interpolation Accuracy

To summarize, there are four major factors that affect the accuracy of point interpolation estimates using the grid approach, and these four factors are also interrelated. First, the size of the grid cell affects the resultant estimates. Unfortunately, there remains no general rule on the optimal size of the grid and how it relates to the data density. A small grid size may lead to a surface that is artificially complicated, whereas a large grid size may not be sufficient to portray the real details. Second, the number of data points affects the estimates. Generally, the larger the number of points, the more accurate the estimates. However, this is not always true, as some points are more 'informative' (e.g., peaks, pits) than others. This is further complicated by the third factor, which is the spatial distribution of data points. Clustered sample points may not add much information, compared with sample points that are more evenly spread out. In fact, clustered sample points may bias the estimate, as shown in the distance-weighting method. Finally, the interpolation method itself, such as the weighting functions used, will seriously affect the accuracy of the estimates.

Areal Interpolation

Areal interpolation, coined by Goodchild and Lam in 1980, was regarded as a variant of the general spatial interpolation problem. As GIS applications increase, the need to integrate various types of data increases. Areal interpolation has been used to transform data defined in one set of boundaries to another, such as census data to political boundaries. It has also been used to transform time-series data where boundaries change over time to enable analysis. Lately, remote-sensing images as well as other data layers have also been utilized to improve the accuracy of the interpolated estimates. The following describes the two most generic approaches using the classification scheme, non-volume-preserving and volume-preserving.

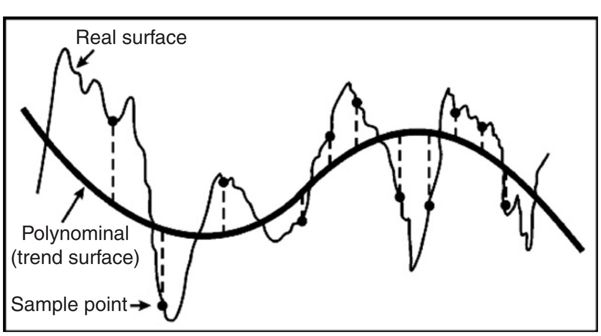

Figure 5 An example of underestimation of true variance at unknown point. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

Non-Volume-Preserving Methods

Consider the areas for which data are available as source zones, and the areas for which data are needed as target zones (Figure 6), the non-volume-preserving approach generally proceeds by overlaying a grid mesh on the study area and assigning a control point to each source zone. A point interpolation is applied to interpolate the values at each grid. Then, the estimates of the grid points

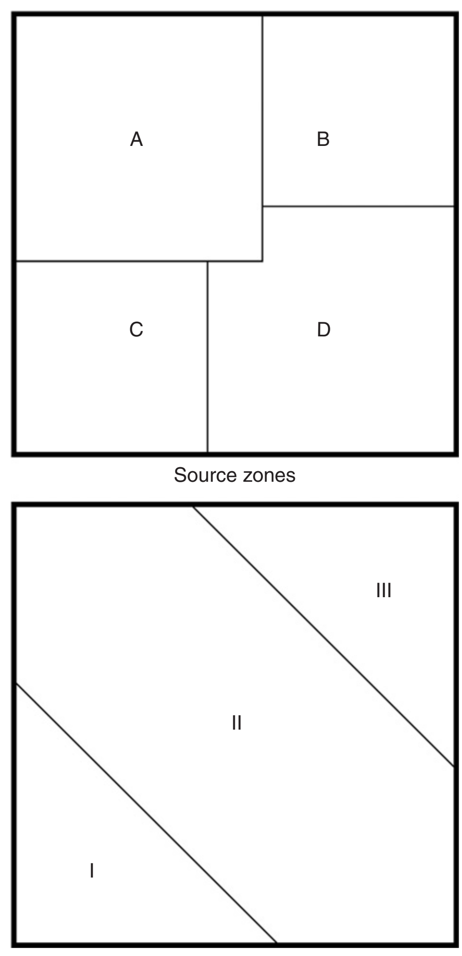

Figure 6 Hypothetical source and target zones. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

are tabulated and averaged according to the new target zone boundaries, yielding the final target zone estimates.

In this approach, the major variations between the numerous methods are the different point interpolation models used in assigning values to grid points. This approach suffers the same sources of error as point interpolation methods. It does not guarantee that the original source zone values are preserved; hence, it is called non-volume-preserving or point-based approach.

Volume-Preserving Methods

In this approach, the original source zone values are preserved. Also, the zone itself is now used as the unit of operation instead of the arbitrarily assigned control point, and no point interpolation is involved. Two basic methods utilizing this approach can be distinguished.

The 'overlay' method, also called area-weighting method, superimposes the target zones on the source zones. The values of the target zones are then estimated as a weighted sum or average based on the size of overlapping areas. For example, if target zone II has 70% overlapping area with source zone A, then 70% of the population in source zone A will be assigned to target zone II. The equation is

$$V_t = \sum_s U_s a_{ts} / \sigma_s \qquad [7]$$

where V_t is the value for target zone t, U_s is the source zone value, a_{ts} is the area of overlaps, and σ_s is the total area of source zone s. There will be minor variations of the formula, depending on whether the variable represents counts or density.

The 'pycnophylactic' method was originally developed by Tobler in 1979 for isopleth mapping. Lam later expanded its uses in areal interpolation. The method starts by superimposing a grid mesh on the polygons and assigning the mean source zone value to each grid within the zone. Then, a number of iterations will be carried out to modify the grid values according to two criteria: smoothing and volume-preserving. An example of the smoothing condition is to change each grid cell value into the average of its four neighbors. The volume-preserving condition is enforced by summing all the grid values in each source zone and then decreasing/increasing all grid values within a zone if the total is above/below the observed source zone value. The final step is to reaggregate the grids according to the target zone boundaries to obtain the target zone estimates.

Factors Affecting Areal Interpolation Accuracy

The quality of areal interpolation estimates depends largely on how the source and target zones are defined and aggregated and the characteristics of the partitioned surface. Judging from the theoretical and subsequent empirical studies, the volume-preserving approach is far more accurate than the non-volume-preserving approach; hence, the former should be used whenever possible. Within the volume-preserving approach, the overlay method assumes spatial homogeneity within each source zone, the more homogeneous the source zones, the more accurate the estimates. On the other hand, the pycnophylactic method assumes a varying smooth surface. Accuracy of the pycnophylactic estimates hence depends on whether this assumption of the underlying surface is true or not.

Subsequent improvements and extensions of these two basic versions of areal interpolation hence focus on improving the knowledge of the source zones by using additional layers of information such as remote-sensing images. These images can be used to identify different land cover areas within source zones so that the area-weighted estimates are more accurate, a procedure that is similar to an old cartographic practice called dasymetric mapping.

Future Directions

Spatial interpolation encompasses a class of methods that are crucial to producing estimates for the analysis of spatial phenomena. Many controversial studies have based on spatial interpolation estimates to derive conclusions without realizing the interpolation error involved in these estimates. A case in point would be the current debate of global warming. While there are many solid scientific studies that document gradual increase in global temperature, the exact magnitude of increase through time and space will need to be carefully interpolated and calibrated using a better interpolation methods such as Kriging or its extension. Another case in point would be for estimating the air pollution PM concentration level through interpolation; the resultant estimates are often used to link with human health, and the conclusions of such studies are often controversial and subject to scrutiny.

Although there have already been numerous studies on interpolation error, it is suggested that future research should focus on investigating the causes and consequences of interpolation error. The emphasis of future studies should be more on developing realistic strategy to manage interpolation error, with an understanding that spatial interpolation error will exist no matter what. The task we are facing is to find the best strategy to mitigate and minimize such error, so that decisions and conclusions basing on the estimates can still be made with a minimal degree of uncertainty.

See also: Kriging and Variogram Models; Overlay (in GIS); Spatial Data Models.

Further Reading

Abraham, J. S. and Comrie, A. C. (2004). Real-time ozone mapping using a regression–interpolation hybrid approach, applied to Tucson, Arizona. *Journal of the Air and Waste Management Association* 54(8), 914–925.

Fisher, P. F. and Langford, M. (1996). Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *The Professional Geographer* 48(3), 299–309.

Fisher, P. F. and Tate, N. J. (2006). Causes and consequences of error in digital elevation models. *Progress in Physical Geography* 30(4), 467–489.

Flowerdew, R., Green, M. and Kehris, E. (1991). Using areal interpolation methods in geographic information systems. *Papers in Regional Science* 70, 303–335.

Goodchild, M. F. and Lam, N. S.-N. (1980). Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing* 1, 297–312.

Goodchild, M. F., Anselin, L. and Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25, 383–397.

Kennedy, E. and Kennedy, L. W. (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology* 20(3), 243–262.

Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

Lam, N. S.-N. (1981). The reliability of spatial interpolation models. *Modeling and Simulation* 12(3), 869–876.

Lam, N. S.-N. (1983). Spatial interpolation methods: A review. *The American Cartographer* 10(2), 129–149.

Liao, D., Peuquet, D. J., Duan, Y. *et al*. (2006). GIS approaches for the estimation of residential-level ambient PM concentrations. *Environmental Health Perspectives* 114(9), 1374–1380.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association* 74(367), 519–536.