7

SPATIAL INTERPOLATION

James Conolly

Introduction

To estimate the value of a phenomenon at an unsampled location from samples of surrounding data requires *interpolation*. This contrasts with *extrapolation* which is a process of estimating values beyond the extent of a sample. Both approaches require a model to provide the estimate, which is based on a prediction function. This might be a heuristic model – visually estimating or just educated guessing – but if the goal of prediction is to estimate values at multiple unsampled locations in space, then a statistical approach provides a more robust solution. Often, surface interpolation is used to generate a continuous raster-based surface model from a scatter of discrete point samples. The utility of spatial interpolation means that it is a commonly applied tool used across a wide range of disciplines with interests in visualizing and predicting spatial patterns and processes. In archaeology, it provides opportunities for the visualization and prediction of a wide variety of geographically variable phenomenon, such as artifact intensities (densities), topographic features, or even more complex models of processes such as the space-time dynamics of cultural change. Conversely, spatial extrapolation is more prone to error and has more limited value, but might be useful in cases where a clear trend such as declining artifact densities needs to be estimated beyond a survey zone. For more information on extrapolation, see Miller, Turner, Smithwick, Dent, and Stanley (2004) and Peters, Herrick, Urban, Gardner, and Breshears (2004) for a discussion of issues and applications.

The basics of spatial interpolation are relatively easy to define, but there are several distinct types of methods – Li and Heap (2008) review over forty – and they vary in their assumptions (or lack of assumptions) about the source data and degree of statistical complexity. Choosing an appropriate interpolation method can be difficult, as archaeologists deal with a variety of different data (from biotic, physical, and cultural), data are unlikely to have been optimally sampled, and samples may contain sources of noise or error. To add to this, interpolation outcomes can also be considerably different depending on the method chosen, so care is therefore required to ensure that the most appropriate method is selected that is sensitive to the underlying data as well as the goals of the analysis.

The purpose of this chapter is to describe the basic concepts of spatial interpolation, to review and to provide guidance on the use of three common interpolation methods, and to offer some examples and discussion of how spatial interpolation provides opportunities for data visualization and prediction that can build insight into the behaviours which generate patterns in the archaeological record.

First, to illustrate the utility of spatial interpolation, consider the following three scenarios, which capture a representative variety of the types of applications that can benefit from spatial interpolation.

Scenario One You are a conservation or commercial archaeologist, and you need to define the spatial characteristics (i.e. the varying intensity) of a large artifact scatter identified by a sample of test units. In this scenario, the goal is primarily data visualization, such that the spatial properties of the cultural materials can be effectively portrayed and communicated to planners for the purposes of avoidance. Spatial interpolation provides a solution for illustrating where the highest artifact intensities are located, and for estimating the edges of the scatter. If removal of the site is required, spatial interpolation can also offer a guide to the placement and number of excavation units, and even estimates of artifact recovery rates, to aid in the calculation of costs. A basic but robust method such as an *inverse distance weighted* (IDW) algorithm is a good first approach.

Scenario Two You are a graduate student and your research project involves modelling an inundated (underwater) cultural landscape. The goal is to obtain a more precise understanding of the surface of the lake bed to assist in the identification of submerged shorelines. The data consists of a combined set of terrestrial topographic and underwater bathymetric measurements that together provide an opportunity to model a palaeolandscape. The interpolation methods require consideration of the possibility of localized features such as submerged shorelines but also to the possibility of errors in data capture. An appropriate resolution sensitive to the required analytic scale and the intensity of data observations must also be considered. A *spline interpolation* with a tension set to respond to local features is a reasonable starting point.

Scenario Three You are researching the spatial characteristics of a sample of radiocarbon dates obtained from a cultural phenomenon, for example, the spread of a cultivar such as maize throughout northeastern North America. The goals of the analysis are to define and visualize the geospatial trend and to identify regions where radiocarbon dates deviate from a global trend, perhaps exhibiting an early or late adoption pattern. In this scenario the data comprise a set of georeferenced radiocarbon dates across a large geographic region (1000s of square kilometres). The spatial interpolation algorithm needs to be sensitive to potential local effects of varying scale, as well as a likely directional north-south trend. For these types of highly complex problems, an appropriate choice of spatial interpolation is *kriging*, in which the parameters of the interpolation are derived empirically from the data. Geospatial interpolation improves the outcome as it provides analysts with the ability to identify where the modelled surface is more or less accurate due to data sampling issues.

Although these three examples are typical uses of spatial interpolation, they do not cover the full range of scenarios in which spatial interpolation provides a solution to an archaeological problem. Other applications of archaeological interest, such as the use of spatial interpolation to model a continuous distribution of soil chemistry values taken from point samples across a historic living surface, have similar requirements. At its most basic, the process begins with a set of discrete point observations recording the changing values of a phenomenon distributed across geographic space. This is followed by the selection and implementation of an appropriate interpolation method, which requires the selection of a range of parameters, as well as the grid resolution of the output model. The first issue is the selection of an algorithm; options are presented in the following section, including some guidance on which of the myriad options is likely to provide the best balance between the data requirements, computational and statistical simplicity, and the archaeological problem to be solved.

Method

In this section I explain the methods behind different forms of spatial interpolation, beginning with the relatively straightforward, followed by the more complex. My goal here is to provide sufficient knowledge such that the underlying concepts of different forms of spatial interpolation are sufficiently understood so that practitioners can make an appropriate decision as to which method is the most suitable for a specific application. Although some forms of spatial interpolation are highly complex and require some understanding of more advanced statistical concepts, other approaches are relatively easy yet are also extremely versatile and powerful tools for surface modelling and data visualization.

Fundamentally, interpolation is a predictive modelling tool used to estimate the value of a quantitative phenomenon (such as an artifact count) between measurements. Interpolation differs from methods such as kernel density estimation (KDE – see Bevan, this volume) because interpolation is concerned with prediction, rather than characterization. Whereas KDE can take a set of point-based frequency data, such as artifact counts, and convert observations to densities, KDE does not predict the densities between point locations – it only characterizes them. Conversely, interpolation predicts the values between observations. Interpolation is thus by its nature a more complex method than density estimations, but I have written this chapter for archaeologists without a specialized background in spatial statistics, so I've avoided the use of mathematical formula and have focused instead on written descriptions of the methods to present the core concepts. It is nevertheless worth defining the fundamental concept of spatial interpolation, to show how it is a weighted average of sampled data. This is illustrated by the equation (Li & Heap, 2008, p. 4):

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \tag{7.1}$$

which simply says that to obtain an interpolated estimate $\hat{z}$ at location x_0, then one must take the sum n of the weighted values from known locations $z(x_i)$, the weighting of each location being given by a procedure defined by λ_i. Where interpolation methods vary is in how many samples are needed to estimate accurately the value at the point of interpolation, and how these samples are to be weighted. At the most basic level, different methods frame how many of the surrounding known values should be considered, and to what extent should nearby points be given greater weight than those far away.

With reference to the three scenarios presented in the previous section, I next consider three classes of interpolation: (1) distance-weighting; (2) thin plate splines; and (3) kriging (also see Lloyd & Atkinson, this volume). There are many more approaches to spatial interpolation beyond these three, a few of which I mention at the end of this section; but most archaeological problems in which spatial interpolation offers a solution can be solved by the judicious application of one of these approaches.

Distance-weighting interpolation

Linear interpolation is illustrated in Figure 7.1(a). Using data presented in the figure as an example, it can be observed that at any location on the imaginary line joining the two points in geographic space, an interpolated value can be reasonably estimated as a function of the linear distance (d) between the known values. Thus, the value at the mid-point of the line will be the mean of the two known observations; but as the point of interpolation moves closer to one of the two points the interpolated value is linearly weighted towards the closest point. On this basis, point D in Figure 7.1(a) is therefore reasonably estimated as being equal to 15.

Figure 7.1(b) shows a scenario in which the values at three points are known. This enables an interpolation to be made at any point within the polygon formed by the three points (A, B and C).

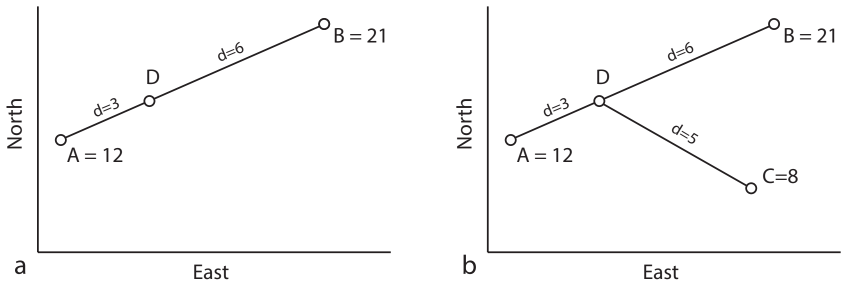

FIGURE 7.1 Simple interpolation examples: (a) with two known point values, using linear interpolation estimates D = 15; (b) adding a third sample location and using inverse distance weighted squared interpolation estimates D = 12.5.

This is in fact the basis of a form of interpolation that is called a *triangulated irregular network* (TIN), in which three neighbouring points form the vertices of a triangular surface that is empty of other points. TIN-based approaches were common in the formative days of GIS-based spatial analysis, as they were computationally easy and visually effective; however, they are one of the least accurate interpolation methods and have been superseded by alternative ways of surface modelling. To interpolate the value, rather than two points, we form an estimate from three or more points. As with the first example, we consider the distances between the interpolation point and other known points and provide more weight to the closer points. This assumes that points closer to the unsampled location are more similar than those lying further away – in other words, that there is some positive spatial autocorrelation in the dataset. A basic way of weighting points is to use the inverse (reciprocal) of the distance, or more often the inverse of the distance raised to a power of two (i.e. 1/d2), to reduce the weight of more distant points. Thus, the weight of each of the known point values to the interpolation value *decreases* with distance to the location of interpolation, giving rise to the method known as an *inverse distance weighted* (IDW) interpolation. The formula defines the method for estimating the value Z at point p, based on the values of the other i points and their distances.

$$Z_p = \frac{\sum_{i=1}^{n} (Z_i / d_i^2)}{\sum_{i=1}^{n} (1 / d_i^2)} \tag{7.2}$$

With reference to Figure 7.1(b), the equation 7.2 is applied to give the interpolated value at point D:

$$D = \frac{\left(\frac{12}{3^2} + \frac{21}{6^2} + \frac{8}{5^2}\right)}{\left(\frac{1}{3^2} + \frac{1}{6^2} + \frac{1}{5^2}\right)} = 12.5$$

As the number of available sample points with known values increases beyond three, then the accuracy of the predication may be improved by using more points to calculate the prediction. This is defined as the neighbourhood search area, which can be limited either by a distance or a defined number of closest points. Because most GIS programs with interpolation tools will by default use the nearest 12 neighbours and apply a power of 2 to the distance weighting, this gives rise to its common shorthand name of IDW-12.

As the neighbourhood search size or number of neighbouring points increases, this will influence the prediction, but it is not always clear what search size to choose and how to weight more distant points. In fact, the choice of neighbourhood size (n) and the power weighing (p) is arbitrary and depends both on the goals of the interpolation and the characteristics of the sample points. For example, increasing p to values above 2, to 3 or even 4, will increase 'bumpiness' as it pays more attention to local values. Conversely, the smoothness of the surface can be controlled by using a greater number of neighbours. Selecting $n = 12$ or more may work well with uneven or noisy samples, in which the goal is to find a more general trend. IDW methods thus usually require some degree of experimentation to dial-in the balance between a surface that is locally sensitive and one that illustrates the regional pattern.

It is worth considering an alternative interpolation method in cases where the underlying data leads to problems with IDW results. This is typically manifest by the surface showing multiple peaks and pits around the original data – and adjustment of the weighting parameters does not improve the result. The two most common alternatives are thin plate splines (TPS) and interpolation with geostatistics, called kriging.

Thin plate spline interpolation

A spline is a continuous curve that connects a set of points. The term 'thin plate spline' (TPS) describes a method conceptually analogous to a stiff plate of metal being warped to lie across a sample of points in three-dimensional space. Each point's attribute is treated as a z-value (elevation), and the resulting surface should ideally pass through the data points with the least amount of curvature – i.e. it should be as smooth as possible. Because the plate intersects each point and curves smoothly between points of different values, it is a good approach for smoothly undulating surfaces such as land surface elevations or climate measurements, such as rainfall, in which localized anomalies are minor (Hutchinson, 2007; Hancock & Hutchinson, 2006).

As with IDW, splines have a set of parameters that can be manipulated, including the size of the neighbourhood in which the spline obtains predictions. Like IDW, choosing parameters is typically by experimentation, with 12 neighbours often set as the initial default. There are two basic approaches with their own parameters that control the relationship of the model in regard to the original points:

Regularized splines allow for curves that extend beyond the range of z-values, and this allows for smoother gradients in the output surface. This can be controlled by a parameter in the function called the *regularization weight*: higher weights reduce the stiffness of the spline and result in smoother surfaces that can extend beyond the z-values of the known samples (Figure 7.2(a)). The correct regularization parameters can be obtained by experimentation, although some implementations of spline interpolation include a built-in procedure for estimating the correct weighting through cross-validation. This is a method of error estimation created by generating a series of surface models using different weighting parameters in which a sample of known observations are withheld from each model. Each of the surface models can then be compared to the known data values to obtain an estimate of the error in each model, and the model with the lowest error then defines the optimum weighting value.

Tension splines are constrained to the original data values and prevent the surface curving beyond the z-range of the point sample (Figure 7.2(b)). If the data collection strategy has defined the range of values (e.g. within an elevation survey, whether the highest and lowest elevations were captured), then a tensioned spline is likely to give more accurate results at the expense of smoothness. A weighting parameter can also be included for tensioned splines, providing some control over the amount of smoothing in the final surface – although with tension splines, increasing the weight will decrease the smoothness.

As with IDW-based interpolation, there is a lot of subjectivity and arbitrariness in the selection of parameters that control the output. Some experimentation is thus necessary to obtain a model that meets

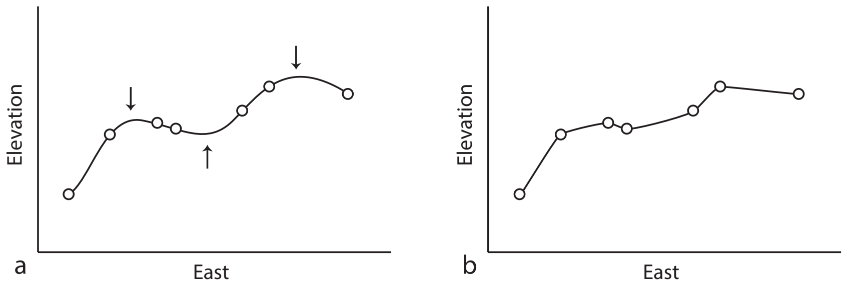

FIGURE 7.2 Spline as a concept: (a) regularized with high weighting, allowing the interpolation estimates to exceed the z-values to maintain smoothness at points marked by arrows; (b) a tension spline, which adheres to the original data values at the expense of smoothness.

the needs of the analysis. Overall, however, thin plate spline approaches are usually better than IDW methods for data in which smoothness is valued in the final product, such as in elevation models. Be aware though that smoothness may be misleading if the underlying data is in fact rough and the goal of the analysis is to illustrate or model this characteristic. In instances where roughness could be attributed to measurement noise (e.g. in bathymetric or topographic survey in which vegetation may be impacting true ground height), or in which more generalized trends are desired (at the expense of local accuracy), then splines do offer a robust solution.

One inherent problem with both IDW and spline interpolation methods is that the neighbourhood size and weighting parameters are typically arbitrarily assigned and visually evaluated. Although it is certainly possible to estimate parameters less subjectively by using cross-validation, this is rarely implemented in the GIS environments in which interpolation is typically performed, and thus it is not routinely applied. This is potentially problematic, as a visually-satisfying surface may be erroneous, and without some form of evaluation of its accuracy it may be uncritically adopted and used as the basis for further interpretation, compounding the error.

Kriging

Kriging (pronounced with a hard 'g', after the South African Daniel Krige) is an interpolation method in which the parameters of the interpolation are estimated empirically using geostatistics. The integration of geostatistics means that kriging is a more complicated form of interpolation, but it provides some advantages over IDW and spline approaches. The primary advantage is that by integrating geostatistics to estimate weighting parameters, the interpolation is sensitive to the characteristics of the samples, and generally produces a more accurate surface model. Some forms of kriging can also generate an error surface so that the interpolation's accuracy can be evaluated across the sampling window. Finally, because kriging requires the analyst to examine the spatial characteristics of the data set before interpolation, it provides opportunities to examine underlying spatial patterns in sample data, which can lead to a better overall decision about the type of kriging to apply that will in turn improve the outcome.

Kriging works by first measuring the spatial autocorrelation of sample points. Spatial autocorrelation is a measure of the relationship between distance and similarity: positive spatial autocorrelation describes a situation in which the value difference between pairs of points is correlated with distance, such that the closer two points are in space, the more similar they are likely to be (also see Hacıgüzeller, this volume).

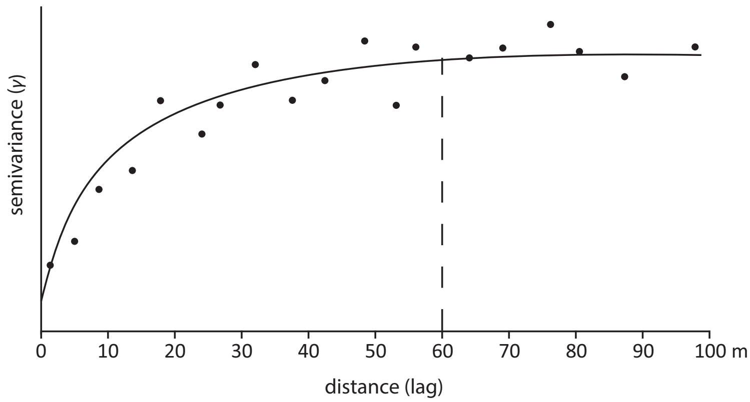

FIGURE 7.3 A variogram showing increasing variance between samples of values drawn from increasing distances apart. After a distance of 60 m there is no increase in variance.

This relationship between distance and similarity is expressed on a graph called a variogram, which plots distances between pairs of points on the x axis (referred to as the 'lag'), and a statistic called the *variance*, denoted by γ on the y axis. The variance is a measure of the variability in differences between all pairs of point values within a defined lag (such that as lag increases, the variance normally increases too, up to the level of the variation in the entire sample). The reason that a graph of lag against variance is useful is that for each unsampled location surrounded by points at varying distance, the variogram provides information about the distance-based weighting needed to estimate the value at unsampled locations. This means that a sample-specific distance weighting can be derived that is sensitive to the original data and removes some of the subjectivity inherent in IDW or TPS interpolation methods. For example, a variogram like the one shown in Figure 7.3 shows that the variance in differences between observed pairs of values increases rapidly up to about 15 m, variance then increases slowly to about 40 m and hits a maximum that does not increase any further from distances of 60 m and high. This means an appropriate weighting factor will consider values within 15 m more significantly than those further away up as these have the lowest variances and thus predictive power; whereas points more than 60 m away have little predictable influence on estimates of local values.

As well as the changing influence of distance, some forms of kriging can include the directionality (or geographic orientation) between pairs of points into the weighting calculation. The term *anisotropic* refers to situations when direction independently influences the rate of change in a sample (e.g. Figure 7.4). This occurs commonly in topographic surfaces but also in other situations when there is a directionality to the source of the response variable, such as might be expected in the amount of material reaching settlements at progressive distances if primarily distributed along geographically oriented distribution routes. To detect the presence of anisotropy requires the creation of a two-dimensional variogram surface, in which the degree of change by direction can be measured, which will typically manifest itself as an ellipse on the variogram surface. If there is anisotropy, then a method called *anisotropic kriging* will produce more accurate surfaces. This requires calculating separate variograms at major and minor axes of the ellipse, and these angles are then provided to the kriging solution to provide a weighting function that considers the directionality of the surface.

These are the foundational concepts for all kriging methods, but they can be implemented differently depending on starting assumptions and the use of additional parameters. Although there are many forms

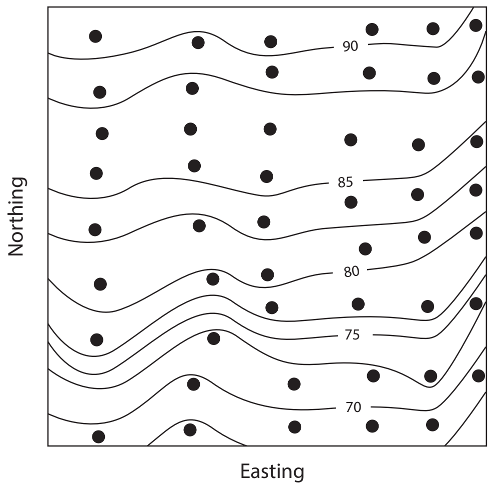

FIGURE 7.4 Anisotropy in a hypothetical sample of semi-regularly spaced test units. The isolines depict sherd counts in 5-sherd intervals, illustrating how the rate of change is greater on the north-south axis than on the east-west axis.

of kriging, the three basic forms are simple, ordinary, and universal. Note that all three of these methods can be implemented to estimate values at unsampled point locations (which is typically the default), or in large polygon units, which is referred to as *block* kriging. Simple and ordinary kriging both assume that the variance in point samples within distance ranges is stationary across the sample (i.e. the observed variation between point samples is not higher in one part of the distribution than another; also see Crema this volume). Unlike simple kriging, ordinary kriging does not assume a constant mean value within a sample of points – it allows for the likelihood that values may have a trend and be higher in one part of the sampling window. Of the two, ordinary kriging is usually the better choice as it has the fewest assumptions. Finally, universal kriging establishes, in a separate process, an equation that describes the first order trend of the observed data within the neighbourhood search window. The kriging function then is a model of the residuals from the trend function. The advantage of universal kriging is that first-order patterns are managed by a separate model, allowing the kriging function to focus on the variability around the global trend. In addition to these three are further forms of kriging, such as co-kriging, which integrates a secondary (correlated) variable into the weighting function to provide estimates that are described well in overviews of geostatistics (see, for example, Haining, Kerry, & Oliver, 2010; Webster & Oliver, 2007).

In general, kriging methods work well when there is spatial structure to the underlying data – i.e. some trend can be observed or is expected in the mean or variation in observations in one or more directions across the sampling window. In addition, as error surfaces can be created in some forms of kriging, a measure of certainty can also be provided. This may be more useful in some situations than others, but if interpolated surfaces are being used for forming decisions about where the highest concentrations in a distribution, or where the edge of artifact scatter is likely to be, then having some understanding of the probable error in the modelled surface is likely to be valuable.

Finally, as these descriptions of kriging methods show, interpolation with geostatistics involves some additional calculations and interpretations to generate the weighting functions, and most GIS packages have limited capacity for full geostatistical analysis. Dedicated spatial statistics software or geostatistical plugins provide more customizable solutions including construction of permutation analysis for evaluating the stability of the models. Lloyd and Atkinson provide a detailed explanation of kriging in this volume; for further extended discussions see Sen (2016).

Sample intensity, measurement scale and edge effects

While the choice of interpolation method and selection of parameters will impact the final surface model, equally significant are the characteristics of the starting point sample. In fact, comparison of model output using identical samples has shown that differences in interpolation method had smaller impact on surface models than differences arising from changing the sample size and the spatial complexity of the phenomenon being modelled (Aguilar, Aguilar, & Carvajal, 2005). It should also be appreciated that some data sets may be too sparsely or irregularly sampled to provide a meaningful output (see Banning, this volume).

Increasing the resolution of the interpolated surface (i.e. by reducing the grain, or raster pixel size in GIS terms) increases the amount of information produced by the analysis (Lam, 2004), but also increases the storage demands and processing time. While more information is generally better (within the limits of available computational time and space), the chosen resolution must also reflect the sample spacing of the observed data points. This is because the ability of the interpolation algorithm to predict an output value correctly decreases when the measurement scale greatly exceeds the spacing of the observation points. Most GIS platforms default to output resolutions that are not based on empirical evaluation of the sample spacing, or instead leave it to the user to subjectively decide. Hengl (2006) provides some useful guidance and suggests the following equation for deriving an appropriate grid resolution:

$$p = w \cdot \sqrt{\frac{A}{N}} \tag{7.3}$$

where p is the grid resolution (i.e. pixel edge length), which is based on the sample area (A) and number of observation points (N) multiplied by a weighting factor of w. Hengl suggests that a weighting of $w = 0.0791$ to $w = 0.25$ is appropriate for random samples, but if the observation sample is regularly spaced (e.g. as might be the case following a defined sampling interval), then a more appropriate weight is $w = 0.5$. It is better to err on the side of less precision (i.e. larger pixels) to maintain accuracy, although (as shown later) an evaluation of the impact of larger or smaller resolutions is often worthwhile.

Edge effects are a major concern in interpolation. Many processes we wish to model are continuous beyond our sampling window, and thus suffer from a lack of sampled information beyond the window. As the accuracy of a prediction is partially dependent on being surrounded by locations where the true values are known, one solution is to reduce the extent of the area to be interpolated so that it is surrounded by known points. This is formally known as a border-area edge correction (Yamada, 2009). How much to step in though is dependent on the intensity of the point sample and thus is related to average n-th nearest neighbour distance, where n is the number of neighbours used in the interpolation. For example, in Figure 7.5, to avoid edge effects in a routine interpolation based on eight nearest neighbours, an estimate of the step-size can be calculated by deriving the average distance from each point to its eighth nearest neighbour (nn). In this case, the average nn distance is 11 m, and this provides a width of the internal buffer around the sample distribution. The obvious disadvantage to this approach is potentially considerable data loss at the edges, but the error rates in this zone are high because of edge effects, so keeping it risks false confidence in the model. There are other methods, but all solutions require compromises between predictive accuracy and data loss. Yamada (2009) provides several examples of ways to manage these concerns.

Model comparisons

There can be considerable variation in the output between different interpolation methods and as well within methods when parameters are adjusted. For example, consider the following typical scenario which consists of a scatter of 202 small (30×30 cm) test units across an area approximately 70 m wide by 130 m long (Figure 7.5(a)). The data was collected in advance of an excavation project on a grassy field interspersed with buildings (the most substantial of which is marked as such), and each test unit records the count of pottery artifacts observed at that location (for general context, please see Conolly et al., 2014). There is a considerable amount of local variation (noise) in the data, but there is a global trend of higher values in the southwest declining to the northeast. The goal of interpolation is to visualize this trend as a continuous process in order to estimate the scale of the sub-area(s) with substantially higher artifact counts.

Three interpolation methods were used to construct the surfaces: inverse distance weighting (IDW), splines, and kriging. In each, different parameters were selected to evaluate the impact these had on the predictive accuracy of the modelled surface. Model resolution was also considered. There are 202 observations in the study area of 9100 m^2, and from equation 7.3 with a weighting factor of 0.4 to reflect the semi-regular spacing, an appropriate pixel dimension is 2.7 m. To evaluate the impact of resolution on accuracy, two models were run using each set of parameters for method: one with a pixel dimension of 1m, and one at 3m. Table 7.1 summarizes the methods and parameters used.

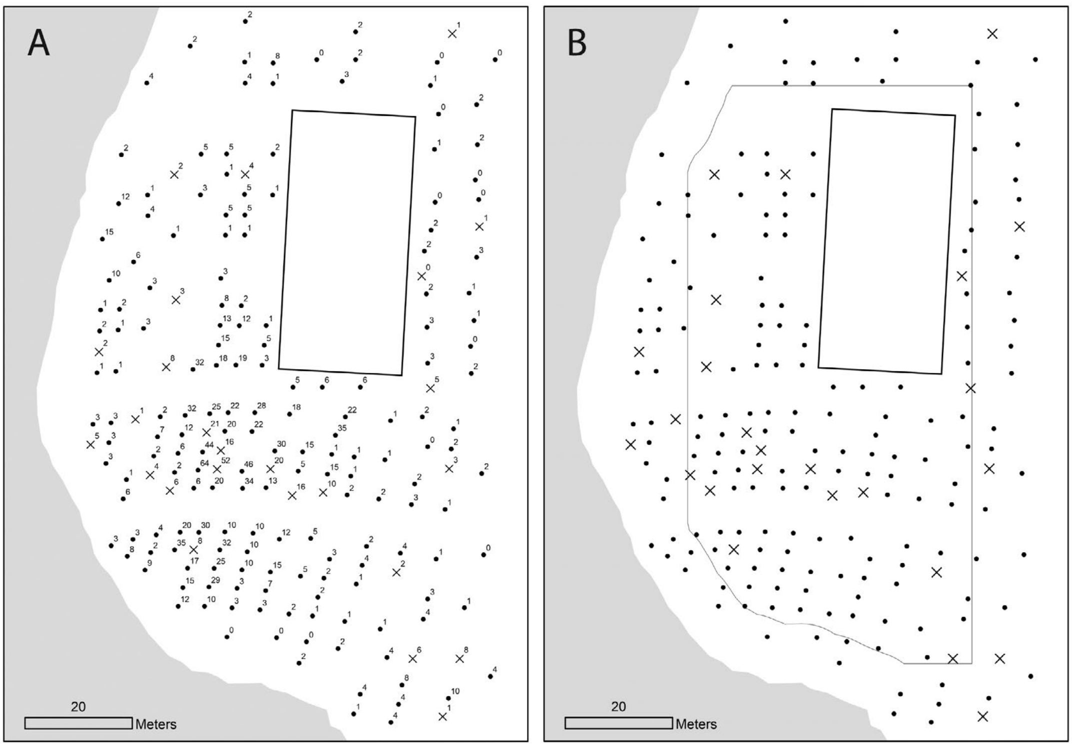

FIGURE 7.5 Archaeological point sample. (a) location of samples and artifact counts; (b) sample with border-area edge correction. The random test samples are designated by an 'x'; the building location by the rectangle.

TABLE 7.1 Interpolation methods and parameters used in the analysis.

Method	*Parameters*	*Anisotropic*	*Grain*
IDW	Neighbours = 4, 8, 12	No	1m, 3m
Regularized spline	Neighbours = 4, 8, 12	No	1m, 3m
Tensioned spline	Neighbours = 4, 8, 12	No	1m, 3m
Simple kriging	Spherical model	No	1m, 3m
Simple kriging	Spherical model	Yes	1m, 3m
Ordinary kriging	Spherical model	No	1m, 3m
Ordinary kriging	Spherical model	Yes	1m, 3m
Universal kriging	Polynomial trend, Spherical model	No	1m, 3m
Universal kriging	Polynomial trend, Spherical model	Yes	1m, 3m

TABLE 7.2 Ranked RMS results by method and resolution.

Rank	*Method*	*1-m*	*3-m*
15	Regularized spline (4)	11.2	10.7
14	Regularized spline (8)	10.7	10.6
13	Regularized spline (12)	9.5	9.4
12	Tension spline (4)	9.3	8.9
11	Tension spline (8)	8.7	8.8
10	Tension spline (12)	8.6	8.3
9	IDW (4)	7.3	7.5
8	Universal kriging (anisotropy)	7.1	7.2
7	IDW (8)	7.0	6.8
6	Ordinary kriging (anisotropy)	6.8	6.7
5	Simple kriging	6.8	6.8
4	IDW (12)	6.7	6.4
3	Universal kriging	6.6	7.2
2	Simple kriging (anisotropy)	6.5	6.6
1	Ordinary kriging	6.5	6.5

To evaluate the relative predictive error in each model, a random evaluation sample of thirty test units was selected and removed from the analysis. The models were constructed on the remaining sample of 172 observations. A root-mean-square (RMS) error was calculated for the difference between the evaluation sample and the predictive surface for each model. (Visual comparison of the models and the RMS results are presented in Figure 7.6 and Table 7.2).

First, as expected, there is a slight increase in accuracy afforded by a 3-m resolution over 1-m pixel resolution, but it is not significant and the latter has been selected for model outputs. Second, the RMS results establish that kriging is consistently able to produce more accurate predictions than other interpolation

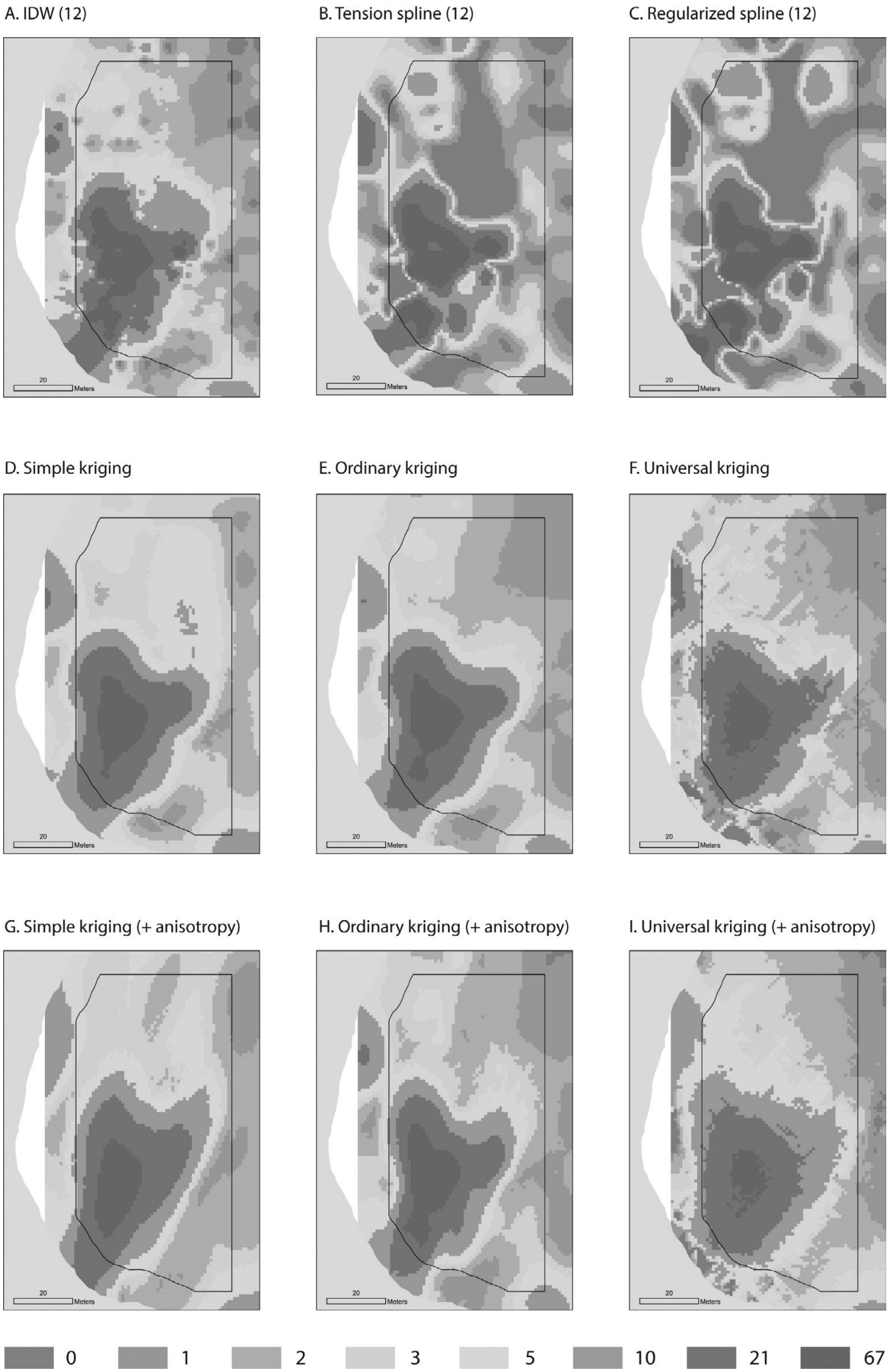

FIGURE 7.6 Visual differences in the surfaces of nine interpolation methods at 1 m resolution. RMS errors for each model are provided in Table 7.2. A colour version of this figure can be found in the plates section.

methods, even if in this instance the results are only marginally more accurate over IDW-12. In this set of data, the spline models performed relatively poorly, as they are better suited to phenomena with smoother transitions and higher spatial autocorrelation, such as elevation surfaces. Clearly, a more locally sensitive interpolation method like IDW provides some advantages over splines. In fact, in this implementation, IDW-12 is roughly equivalent to ordinary kriging, suggesting that it is a reasonable model to use if simplicity in implementation is worth more than the additional insight and potential increase in accuracy geostatistical modelling provides.

Case studies

Two case studies are considered, at two different scales of analysis, that illustrate how interpolation methods can convert samples of discrete observations into critical insights into the spatial patterns of human behaviour. The first case study concerns the reconstruction of use-of-space in Late Neolithic (LN) and Copper Age (CA) sites in Hungary by Salisbury (2013). The second examines the use of interpolation to provide insight into the geographic patterns at continental scale generated by the spread of Neolithic agricultural practices from Southwest Asia into and across Europe by Fort (2015).

In the first example, the goal of the analysis is to visualize the spatial variation in soil chemistry from habitation sites in order to reveal patterns in use of space. This is a popular form of spatial analysis that depends on interpolation and has been approached in different ways in a variety of contexts (see, for example, Rondelli et al., 2014; Mikołajczyk & Milek, 2016; Negre, Muñoz, & Lancelotti, 2016). To illustrate the potentials and a few pitfalls in the application of interpolation methods work by Salisbury (2013) is used. The specifics involve a sample of six LN and CA habitation locales in eastern Hungary and the data consists of element abundance (ppm) in soil samples taken systematically (at a 10 m or 5 m grid interval) across each site. The multivariate data was reduced using principal components analysis (PCA) to identify correlated variation in groups of elements that are assumed to reflect different anthropogenic processes (e.g. cooking, food discard, metal working). The first five of the PCA component axes (cumulatively representing over 80% of the variation) were then used as the variables to be interpolated. Each component was examined separately by assigning each of the original samples a value based on the sample's position on the component's axis.

Salisbury (2013) used ordinary kriging to produce the interpolated maps for visualizing the spatial patterns. As described in the previous section, this is an excellent approach as it measures local variation in the mean observed values across different portions of the data to generate a weighting function for the interpolation. The maps generated from this analysis illustrate clear differences between the different PCA components, which the author uses to interpret patterns in the use of space and their different chemical signatures. However, note that the output scale appears to have been defined at too fine a grain given the distances between sample points and this appears to have led to some instability in at least one of the outputs (Figure 7.7).

Unfortunately, the author did not provide any information about the specifics of the scale used, nor about the variogram model, leaving it to readers to trust that kriging methods were correctly and appropriately derived. More fundamentally, as the raw data is not provided in this paper, there is no opportunity for interested readers to replicate the analysis or build on this work using different methods. I highlight this not to criticize the analysis or interpretation, only to illustrate that without sufficient supplementary data the interpolation methods and output can only be taken on trust. Nevertheless, the interpolations do provide information about patterns in different chemical signatures in the soil, and these patterns can be evaluated and substantiated using other archaeological data. Without the methods provided by spatial interpolation, these insights would be difficult to obtain.

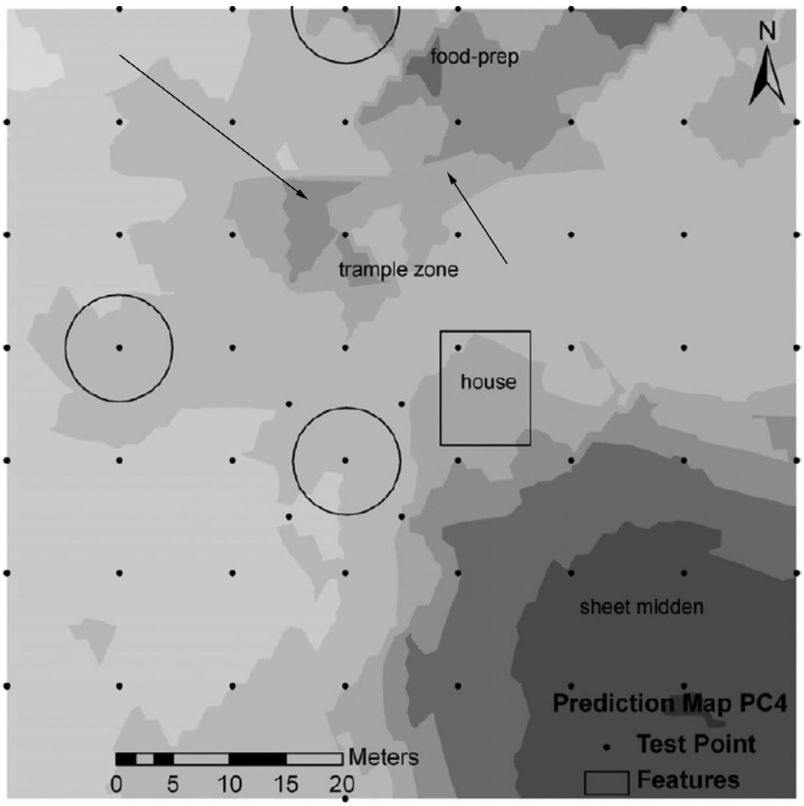

FIGURE 7.7 Interpolation example modified from Salisbury (2013, Figure 5). Note that the high resolution (small pixel size) has exceeded the limits of the original data. The interpolation is thus unstable and noisy where there is higher local variance, for example in the area north of the 'trample zone' (arrows).

The next illustrative case study concerns a much smaller scale and correspondingly larger region of analysis, which is the spread of agriculture across Europe. There are many papers on this topic which use some form of interpolation to calculate patterns of movement (e.g. Ammerman & Cavalli-Sforza, 1984; Gkiasta, Shennan, & Steele, 2003; Bocquet-Appel, Naji, Vander Linden, & Kozlowski, 2009). A recent illustrative example by Fort (2015) uses interpolation to visualize and derive estimates for the absolute speed of demic (i.e. movement of people) versus cultural diffusion related to this economic transformation. Fort's work is based on a point sample of nearly 1000 radiocarbon dates scattered across Europe that record the dates and locations of early farming communities. As there has been a long-standing debate over the relative importance of demic versus cultural processes in this transition, Fort's stated goal was to use the temporal patterns to distinguish between the two processes.

Fort first derives mathematical models for demic, cultural, and demic-cultural diffusion rates and shows that demic-cultural diffusion will spread over space faster than just demic or cultural diffusion alone. Second, Fort interpolates the radiocarbon point data to generate a surface model to visualize the temporal patterns related to the appearance of agriculture. Because there is a clear southeast-northwest spatial trend he uses universal kriging to allow for the first order trend surfaces to be incorporated into the weighting algorithm, although he does note that experimentation with different methods produced

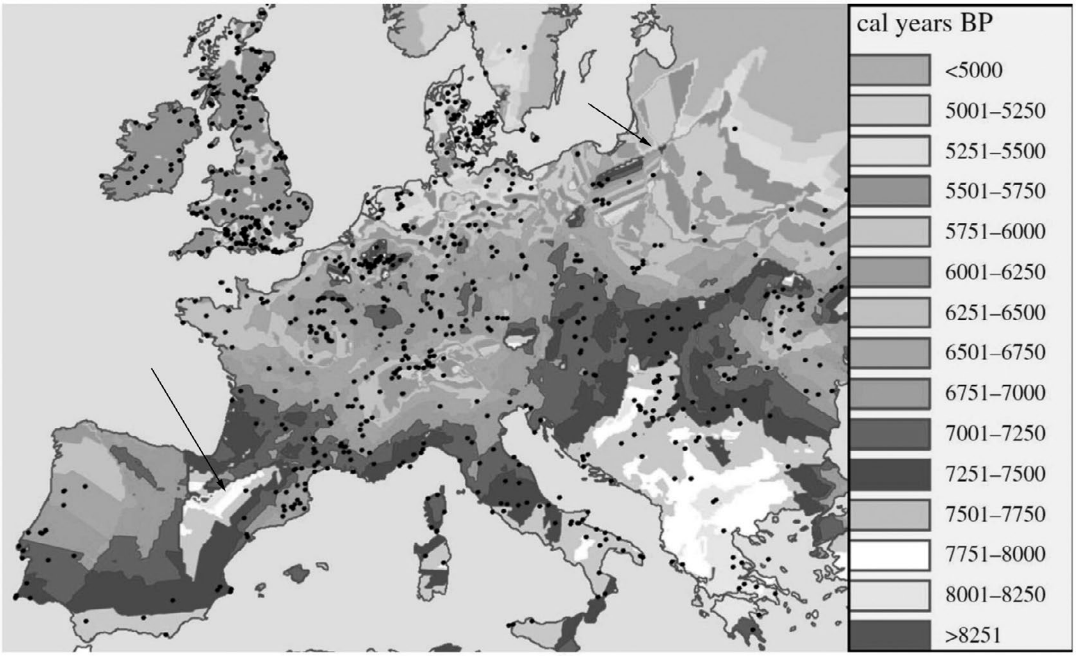

FIGURE 7.8 Interpolation example modified from Fort (2015, Figure 1). An interpolated surface model of radiocarbon dates from Neolithic sites (black dots) depicting the space-time process of the spread of agriculture across Europe. Note the areas indicated by arrows in the southwest and northeast of the model showing where data sparseness causes instability in estimates. A colour version of this figure can be found in the plates section.

similar results. Like the Salisbury paper discussed earlier, there is no information on the model and variogram, but the raw data is made available in Isern, Fort, and Vander Linden (2012) for further analysis and verification by interested readers. Following the interpolation, Fort uses the predictive surface to derive a local 'directional' surface that visualizes the speed and geographic direction of the transmission.

Note in the model (Figure 7.8) how the sparseness of the data samples in some locations (e.g. in Spain) causes some instability in the predictions that create artificial looking temporal boundaries. This is an unavoidable problem with sample data that are unevenly distributed, but contributing to the difficulties in this case could be a grain that is too fine for the density of points. Creating sub-interpolations at different resolutions to consider the heterogeneity of the sample and then combining into a single map is a potential solution. As a further possibility, because calibrated radiocarbon dates are non-normal probability distributions, it would be useful to consider applying a Monte Carlo approach to generate multiple predictive surfaces based on radiocarbon values randomly selected from under each site's pooled probability curve. This would allow for estimation of uncertainty in the surface prediction reflecting the challenge of using radiocarbon data for time-space models. The details of this are beyond the scope of this chapter but it serves to illustrate the potential additional uses of interpolation for the visualization of space-time dynamics.

Conclusion

I have described the core concepts underlying all forms of spatial interpolation, along with the primary methods that archaeologists have used in the past and continue to use when they need to build and interpret continuous surfaces from point observations. As so much of archaeological data is collected as point observations – or can easily be converted to point observations – this means that interpolation has a very wide range of potential applications.

From the case studies and scenarios described, it should be clear that interpolation methods are also very flexible and offer multiple ways to tailor an analysis to fit the character of the data and the inferred spatial process that is being modelled. However, inexperienced users are strongly advised to not accept the defaults in many of the GIS software platforms, as these rarely provide optimum solutions. Instead, first consider questions such as whether there is a spatial trend in the data, whether there is a potential for anisotropy, whether the surface process is likely to be smooth or rough, whether there are boundaries and how these are to be managed, how the edge effect is going to be managed, and what the appropriate output resolution should be. Careful consideration of these questions will certainly lead to a more successful application of interpolation and is likely to produce a more accurate surface. Experimenting with several approaches and parameters and evaluating them against a test sample of known points withheld from the analysis or in a more formalized way may be the best option. For critical application formal evaluation of predictive accuracy is the preferred solution, such as through test and training samples, cross-validation or by using a method such as ordinary or universal kriging, which is generally seen as a more robust approach than IDW or TPS approaches, especially with data having a global trend.

As a final note, interpolation does not have to end with visualization. There are ways of generating additional insights into spatial processes by using the interpolated surfaces to derive other measurements. The obvious ones are using maps of elevation to derive slope and aspect maps, but as shown earlier, Fort (2015) explains how he obtains a direction of change map based on his interpolation of radiocarbon dates. If maps are produced representing artifact density, they can similarly be converted into insightful visualizations that locate major rates of change using similar methods. Interpolated maps showing different artifact densities (e.g. comparing pottery to lithics) can also be manipulated with map algebra to illustrate how the two are correlated. These types of approaches more generally fall under the umbrella of spatial data manipulation and raster or map algebra but are mentioned here to emphasize that analysis need not end with the construction of a continuous surface of a spatial process – surfaces can be the building blocks for additional forms of data visualization and analysis.

References

Aguilar, F. J., Aguilar, M. A., & Carvajal, F. (2005). Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy. *Photogrammetric Engineering and Remote Sensing*, 71(7), 805–816.

Ammerman, A. J., & Cavalli-Sforza, L. L. (1984). *The Neolithic transition and the genetics of populations in Europe*. Princeton, NJ: Princeton University Press.

Bocquet-Appel, J.-P., Naji, S., Vander Linden, M., & Kozlowski, J. K. (2009). Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14c dates. *Journal of Archaeological Science*, 36(3), 807–820.

Conolly, J., Dillane, J., Dougherty, K., Elaschuk, K., Csenkey, K., Wagner, T., & Williams, J. (2014). Early collective burial practices in a complex wetland setting: An interim report on mortuary patterning, paleodietary analysis, zooarchaeology, material culture and radiocarbon dates from Jacob Island (BcGo17), Kawartha Lakes, Ontario. *Canadian Journal of Archaeology/Journal Canadien d'Archéologie*, 38, 106–133.

Fort, J. (2015). Demic and cultural diffusion propagated the Neolithic transition across different regions of Europe. *Journal of the Royal Society Interface*, 12, 20150166.

Gkiasta, M., Shennan, S., & Steele, J. (2003). Neolithic transition in Europe: The radiocarbon record revisited. *Antiquity*, 77, 45–62.

Haining, R. P., Kerry, R., & Oliver, M. A. (2010). Geography, spatial data analysis, and geostatistics: An overview. *Geographic Analysis*, 42, 7–31.

Hancock, P., & Hutchinson, M. (2006). Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environmental Modelling and Software*, 21(12), 1684–1694.

Hengl, T. (2006). Finding the right pixel size. *Computers and Geosciences*, 32(9), 1283–1298.

Hutchinson, M. (2007). Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems*, 4, 385–403.

Isern, N., Fort, J., & Vander Linden, M. (2012). Space competition and time delays in human range expansions. application to the neolithic transition. *PLoS One*, 7(12), e51106. doi:10.1371/journal.pone.0051.

Lam, N. S.-N. (2004). Fractals and scale in environmental assessment and monitoring. In E. Sheppard & E. B. McMaster (Eds.), *Scale and geographic inquiry: Nature, society, and method* (pp. 23–40). Hoboken, NJ: Wiley.

Li, J., & Heap, A. D. (2008). *A review of spatial interpolation methods for environmental scientists*. Technical Report Record 2008/23, Retrieved from Geoscience Australia, Department of Resources, Energy and Tourism, Commonwealth of Australia. Retrieved December 5, 2018, from https://data.gov.au/dataset/a-review-of-spatial-interpolation-methods-for-environmental-scientists

Mikołajczyk, L., & Milek, K. (2016). Geostatistical approach to spatial, multi-elemental dataset from an archaeological site in Vatnsfjörður, Iceland. *Journal of Archaeological Science: Reports*, 9, 577–585.

Miller, J. R., Turner, M. G., Smithwick, E. A. H., Dent, C. L., & Stanley, E. H. (2004). Spatial extrapolation: The science of predicting ecological patterns and processes. *Bioscience*, 54(4), 310–320.

Negre, J., Muñoz, F., & Lancelotti, C. (2016). Geostatistical modelling of chemical residues on archaeological floors in the presence of barriers. *Journal of Archaeological Science*, 70, 91–101.

Peters, D. P. C., Herrick, J. E., Urban, D. L., Gardner, R. H., & Breshears, D. D. (2004). Strategies for ecological extrapolation. *OIKOS*, 106(3), 627–636.

Rondelli, B., Lancelotti, C., Madella, M., Pecci, A., Balbo, A., Pérez, J. R., . . . Ajithprasad, P. (2014). Anthropic activity markers and spatial variability: An ethnoarchaeological experiment in a domestic unit of Northern Gujarat (India). *Journal of Archaeological Science*, 41, 482–492.

Salisbury, R. B. (2013). Interpolating geochemical patterning of activity zones at Late Neolithic and Early Copper Age settlements in eastern Hungary. *Journal of Archaeological Science*, 40(2), 926–934.

Sen, Z. (2016). *Spatial modeling principles in earth sciences* (2nd ed.). New York: Springer.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. Chichester, UK: Wiley.

Yamada, I. (2009). Edge effects. In R. Kitchin and N. Thrift (Eds.), *International encyclopedia of human geography* (pp. 381–388). Elsevier.