Spatial Interpolation Notes

December 14, 2020

1 Introduction

We want to use interpolation because it is reasonable to assume that spatially distributed variables are also spatially correlated. It is not always true, but often worth exploring as part of an analysis. There are multiple different methods to interpolate data that depend on different underlying assumptions. These methods are described below.

Information from ARCGIS and Applied Spatial Data Analysis With R (2013) unless listed otherwise

2 Inverse distance weighting (IDW)

IWD determines cell values using a linearly weighted combination of surrounding values. The weights are function of the inverse distance. The general form for the IDW function is:

$$\hat{Z}(s_0) = \frac{\sum_{i=1}^{n} w(s_i) Z(s_i)}{\sum_{i=1}^{n} w(s_i)}$$

where: $w(s_i) = ||s_i - s_0||^{-p}$
$||\cdot||$ = Euclidean distance
p = an inverse distance weighting power, defaulting to 2

The value of p determines how much closer values are preferred. As p increases, IDW approaches a one-nearest-neighbour interpolation model. p can be selected using cross-validation.
Another way to control IWD interpolation is through selecting the number of neighbouring observations to include. This can improve speed of interpolation, and may be used when there is reason to believe that distant points have little correlation. There are two approaches for varying the number of points used for interpolation:

1. Varying search radius
    - The number of points to include is fixed, and the radius changes to include that set number
    - Depends on the density of observations fluctuating
    - The maximum radius can also be set, in which case all points will be included if that max radius is reached before n
2. Fixed search radius
    - Set a radius and minimum number of points
    - If $n <$ minimum number of points at set radius, the radius increases until the minimum is reached.

In addition to these two approaches, barriers can be created to limit the searches for neighbouring points, i.e. only search for this side of a river.

3 Kriging

Kriging is a way to interpolate data and estimate a surface for geospatial data. There are two main ways of doing this:

- Inverse distance weighting (IDW) and spline methods
- Kriging

IDW are deterministic interpolation methods as they are directly based on the surrounding values or smoothed formulas. Kriging is different as it uses autocorrelation and takes position into account in the statistical models. Kriging uses a certain number of neighbouring points, or all points within a specified radius (cf kNN).
The general formula for IDW and kriging is:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

where: $Z(s_i)$ = the measured value at the ith location
λ_i = an unknown weight for the measured value at the ith location
s_0 = the prediction location
N = the number of measured values

The difference between IDW and kriging is that in IDW, λ_i only depends on distance to prediction location. In kriging λ_i also depends on autocorrelation i.e. spatial relationship between prediction locations.

3.1 Assumptions in kriging

Information for assumptions from Columbia

For kriging to be used, there are a number of assumptions/conditions to be met. These conditions can be checked in exploratory data analysis.

1. The data should be stationary
    - Means that the joint probability distribution does not vary across the study space, so the same parameters (e.g. mean, range and sill etc) are valid across the space
    - Means one variogram is valid across the space
2. Assumption of isotropy
    - Uniformity in all directions (semivariance identical in all directions)

3.2 Creating a prediction map with kriging

There are two steps:

1. Create the variograms and covariance functions to estimate the spatial autocorrelation values that depend on the model of autocorrelation (fitting a model).
2. Predict the unknown values

3.2.1 Variography (spatial modelling/structural analysis)

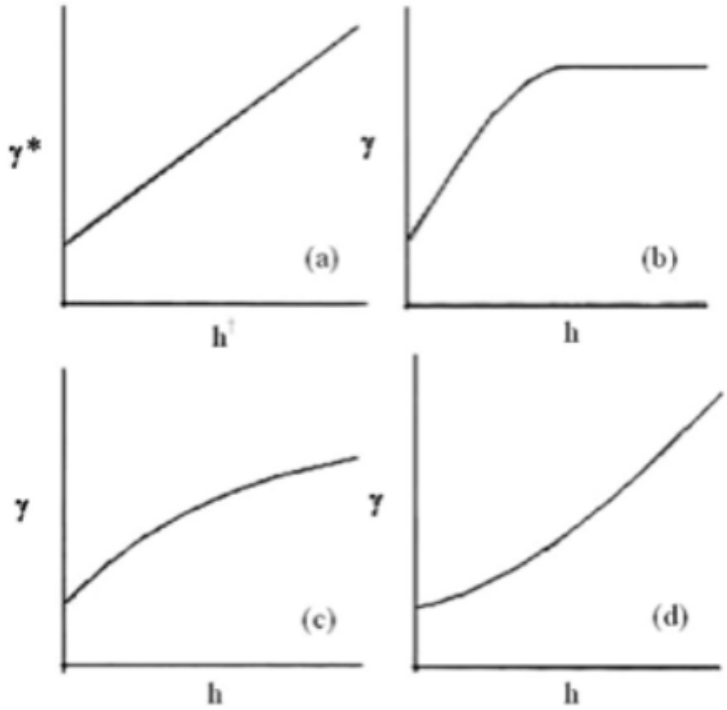
There are often too many pairs of spatial points to calculate and plot the distance for each pair. Instead, spatial distances are put into lag bins i.e. all points in the range $40m < h \leq 50m$ of point A, and calculate the semivariance. The semivariance is equal to half the variance of the differences between all possible points spaced a constant distance apart.

$$\gamma(h) = \frac{1}{2}[z(x_i) - z(x_j)]^2$$

Plotting the distance vs semivariance produces an empirical semivariogram. Closer items should be more similar, therefore lower semivariance. The opposite is true for further points.
A model is fit to the empirical semivariogram (cf regression). Different types of models can be fit the the semivariogram:

- Spherical (most common)
- Circular
- Exponential
- Gaussian
- Linear

Figure 1: Different types of models used in spatial modelling (Poilou 2008). a) Linear semi-variogram; (b) spherical semi-variogram; (c) exponential semi-variogram; and (d) power semi-variogram



There are a number of key points on the figures:

- Range
  - The Range is the point at which the semivariance first levels off
  - Items within the range are autocorrelated (distance matters)
  - Items outside the range are not autocorrelated (distance no longer changes the semivariance)
- Sill
  - The Sill is the height at which the semivariance levels off to
- Nugget
  - The minimum value of semivariance ($\gamma(h = 0)$)
  - Theoretically there is no semivariance when $h = 0$, but in reality it is present due to measurement error or spatial sources of variation at distances smaller than the sample interval (or both)
- Partial Sill
  - Amount of semivariance between Sill and the Nugget

3.2.2 Predictions

Now a model has been fit to the semivariance and autocorrelation can be observed, predictions can be made within the domain. Kriging differs from IWD as it uses the semivariogram to calculate the weights. There are two methods used in kriging:

1. Ordinary kriging
   - Assumes the constant mean is unknown
2. Universal kriging
   - Assumes there's a prevailing trend
   - Trend is modelled with polynomial function, and subtracted from observed
   - Semivariogram is modelled on the residuals to produce autocorrelations

4 Natural Neighbour

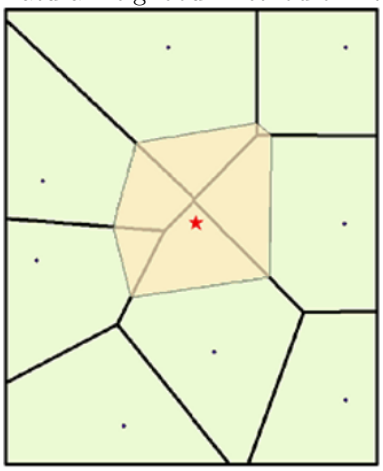
Natural neighbour is a local method that examines samples near the point of interest and evaluates the relative overlap with their areas. The relative overlaps are then used to create the weights for interpolation. Because of this, it is also known as "area-stealing" (Sibson) interpolation. Natural neighbour interpolation therefore does not infer trends that are not already present in the data, and the surface passes through the points, and is smooth in between.
The areas are called Voronoi (Thiessen) polygons. Voronoi polygons are created by examining the space around points and drawing the boundary so that every place inside the boundary is closest to the polygon's point than any other. Formally this is written as:

$$R_k = \{x \in X \mid d(x, P_k) < d(x, P_j) \text{ for all } j \neq k)\}$$

where: R_k = Voronoi polygon of point k
P_k = Point k
P_j = Neighbouring point j

An example of this can be seen in the figure below.

Figure 2: Natural neighbour method of interpolation



5 Splines

Splines are a smoothing function that pass through all the input points and attempt to create a smooth surface between them. The surface is fit to a specified number of neighbouring input points. The basic spline is also known as a thin plate interpolation. There are two conditions that minimum curvature splines must follow:

1. The surface must pass through all data points
2. The surface must have minimum curvature i.e. minimize the cumulative sum of squares of the second derivative terms of the surface at each point

One possible issue with thin plate interpolation is that there may be rapid change in first derivatives around each data point.
Generally, the spline formula is:

$$S(x, y) = T(x, y) + \sum_{j=1}^{N} \lambda_j R(r_j)$$

There are two spline types:

1. Regularized splines
   - Creates a smooth and gradually changing surface, allowing values outside those observed in the data
   -

2. Tension splines
    - Creates a less smooth surface with values more tightly constrained by the sample data range