Spatial Interpolation of Minimum Temperatures in Southland, New Zealand

Irfon Jones & Katrina Richards
Geography Department
University of Otago, Dunedin, New Zealand
Email: irfonj@yahoo.com

Presented at SIRC 2001 – The 13th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 2nd –5th 2001

Abstract

This study addresses the spatial interpolation of one kind of extreme climate event in Southland, New Zealand: air frosts. The presentation of temperature data as a continuous, interpolated surface, rather than as a set of individual points, facilitates an easier understanding of spatial patterns in the study area. The methodology developed in this study has potential application for mapping frost hazards for crop insurance companies, identifying microclimates for new crops, and is transferable to other types of data besides temperature.

1.0 INTRODUCTION

Spatial data have two characteristics: location and attribute. Location is represented using a coordinate system, usually in two dimensions (three dimensions is also possible). Latitude/Longitude and Grid Referencing are two systems that are commonly used in Geography. Attribute is the value of the variable of interest, e.g. temperature, vegetation or soil characteristic. In many natural systems attributes at nearby locations are comparatively similar to each other, whereas attributes are less similar for locations that are far apart. This expectation of spatial correlation has been formalised as the "First Law of Geography" (Tobler 1970, p236). It is impractical to sample attributes at all possible locations in any study area. Therefore data collection is conducted using a sampling scheme whereby attributes are found for point locations. Points may be randomly or regularly spaced, and can be stratified according to areal divisions, e.g. different soil classes or landscape units. From these point data, inferences can be drawn about the nature of the attribute surface and attribute values at unsampled locations, i.e. between points. Assuming there are no sharp boundaries in the data, it is possible to accurately interpolate values between the sampled point data, and to generate a continuous surface of information. Methods used to interpolate at unsampled locations can be grouped under the term "spatial interpolation".

2.0 SPATIAL INTERPOLATION

There are two broad approaches to spatial interpolation: global and local. Global interpolation methods are dependant on all data points in the study area (Franke 1982). That is, each sampled data point, irrespective of its location and thus its relative distance from other points, contributes information to every interpolating site. Changing the location or attribute of any one data point modifies the value obtained at each interpolating site. Thus a different map results. In contrast, local interpolation methods are dependant only on ‘nearby’ data points (Franke 1982). Local interpolation methods use only the data points within a specified search radius to interpolate. The definition of ‘nearby’ is difficult and hard to quantify.

A review of the literature shows that several interpolation techniques have been developed. All make use of a weighting scheme. These are based on the distance between the interpolating site and surrounding sampled sites, with nearby sites being weighted preferentially. For example, a nearby site may receive a weighting of 0.9, whereas a distant site may have a weighting of 0.1. Thus the attribute of the interpolating site will be composed of 90% of the attribute of the nearby site and only 10% of the attribute of the distant site. The method by which these weights are calculated holds important assumptions regarding the outcome and reliability of the interpolation model. This consequently impacts on the validity of a particular interpolation method to a particular dataset. Each interpolation method results in a different map because the underlying model differs.

2.1 Kriging

Of all the interpolation methods, kriging is the only technique that considers two sources of information regarding the attribute in question. These are:

1. Variation between points
2. Distance between points

2.2 The Semivariogram

Variation between points is measured using semivariance. By pooling together pairs of data at a geographic distance and direction h apart, the semi-variance ($2\gamma(h)$) of the sample can be calculated by:

$$2\gamma^*(h) = \frac{1}{n}\sum[g(x) - g(x+h)]^2 \text{ (Clark 1979)}$$

The above formula generates the sample semivariance, $2\gamma^*(h)$, based upon a grade (g), geographic location (x) of the grade and the distance and direction (h) of the pair from (x).

A plot of semi-variance values against h gives the semivariogram (Figure 2.1). The plot displays the variation of a variable with increasing distance. Fitting of the semivariogram can be made using Weighted Least Squares (WLS), as proposed by Cressie (1985). Several models can be fitted, depending on how the semivariogram plots (Figure 2.2).

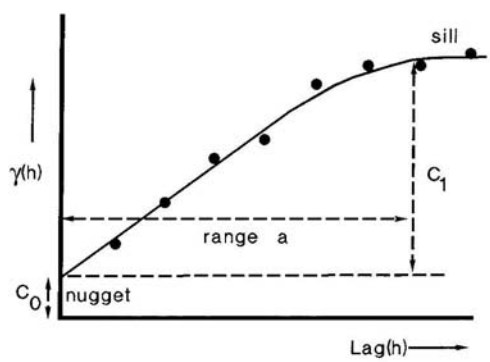

Figure 2.1 Semivariogram fitted with a spherical model (from Burrough & McDonnell 1998)

The semivariogram displays several characteristics of the plotted variable. The nugget effect (C_O) is a measure of spatially uncorrelated noise. One would expect the semivariogram to always pass through the origin, as two samples taken at the same location would be identical. However this is often not the case, as Figure 2.1 shows. In the presence of a nugget effect or spatially uncorrelated noise, the semi-variogram crosses the y axis at a positive value.

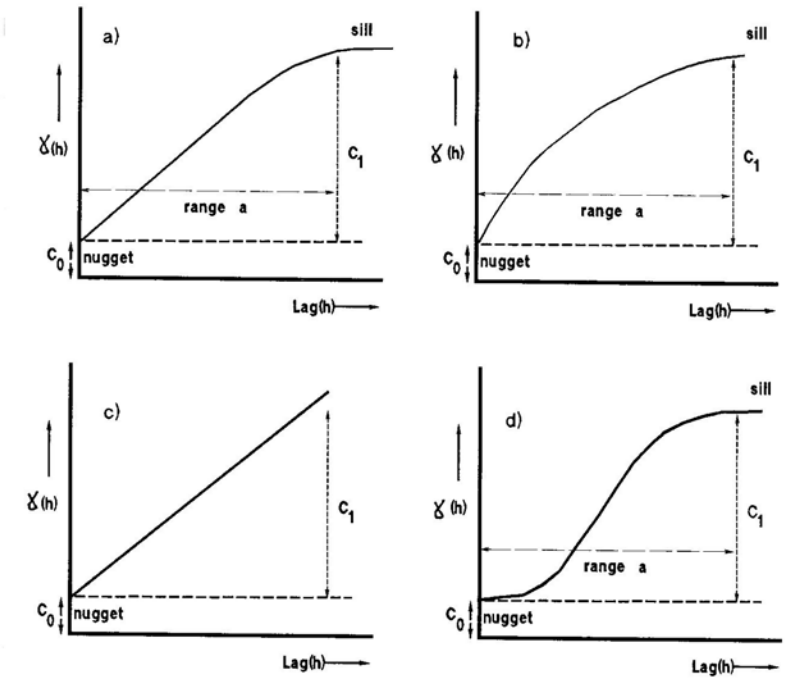

Figure 2.2 Various models fitted to semi-variograms (a) Spherical; (b) Exponential; (c) Linear; (d) Gaussian (Burrough & McDonnell 1998))

As the distance between samples increases, the semivariance between them also increases until the samples become independent. In the semivariogram plot, this feature is shown by the fitted line becoming horizontal. The distance at which this occurs is termed the range (a) and the corresponding semi-variance value is the sill (C_1). In physical terms, when samples become independent of each other, they are no longer 'related' in a spatial sense.

A large number of sampled data points at varying distances are required to produce a semivariogram. Burrough and McDonnell (1998) recommend between 50 and 100 sampled data points are necessary. If insufficient data exist, then kriging should not be attempted due to the inappropriate modeling of the data.

2.3 Interpolation by Kriging

The process of interpolation by kriging is intended to fulfill several characteristics:

1. Exact Interpolator: When the attribute at a location is known as part of the data set, the interpolated value is equal to the known value. Therefore there is no under or over estimation, characteristic of other techniques.

2. Unbiased Estimator: The sum of the weighting adds to 1. This allows weighting to be apportioned in a manner which is not biased towards certain spatial arrangements (e.g. clustering, as in the case of IDW interpolation).

3. Estimation Variance is Minimised: This achieves the minimum variance at each interpolation site. It can also be mapped. Thus a map of the variance of the estimate can be produced in tandem with the predicted value.

Several forms of kriging exist, each with their particular characteristics and each suited to a specific task. A brief overview follows:

1. Simple Kriging: The system is identical to ordinary kriging as defined above, however the area mean is known beforehand.

2. Universal Kriging: This method allows the application of kriging with background trend.

3. Co-Kriging: This is the kriging of a surface of values using two data sets. The system can be used when too few samples of one variable have been taken but data of another spatially correlated variable exists which has been sufficiently sampled. A crossvariogram is then used to generate the weighting scheme.

4. Stratified Kriging: By creating individual semivariograms for discrete areas, e.g. of different soil classes, and kriging according to those semivariograms, stratified kriging relaxes the requirements of stationarity. The technique involves cutting an unstationary area of interest into regions complying with stationarity. Therefore kriging can be applied to those areas without compromising its assumptions. However, this method requires a large amount of data to be feasible.

5. Non-Linear Kriging: This is the kriging of a skewed data set by logarithmically transforming the data before the semivariogram is created. Bias is often introduced by back-transforming the data.

3.0 METHODS

3.1 The Topoclimate South Project

The Southland region of New Zealand includes substantial areas of high class soils and micro- to local-scale (0.1-50 km) climates. The region is capable of growing a wide range of crops and contributes significantly to New Zealand's Gross Domestic Product (about 18%). However, profits from traditional farming methods (e.g. sheep grazing) have declined. Efforts to increase regional wealth and employment levels have lead to the Topoclimate South project. By increasing climatic knowledge, informed decisions about rural productivity, sustainability and potential new crops can be made. Air temperature was measured from 1998 to 2001 with a network of automatic dataloggers positioned with regard to key features of the landscape, e.g. valley floors, terraces and hillsides of varying aspect and slope. These sensors recorded air temperature electronically at 6-minute intervals at a height of 1.2m for a full year at each of a total of 2550 sites. From this data daily extremes, such as minimum daily temperature, could be extracted.

3.2 Study Area

The study area for the project described in this paper is the Waimea Plain and surrounding hills. To the south, the North Range forms a ridge of around 500m in altitude which is part of the northern edge of the Southland syncline. The Lintley Range to the north reaches a height of around 630m. Farming is the main land use in this area with no significant settlements. The villages of Lumsden and Balfour, contained within the study area have a population of 190 and 54 respectively (Statistics New Zealand 1996). Therefore anthropogenic modification of the temperature should be minimal. 212 temperature sensors were present in this area from 1998 to 1999.

On the 21st of June 1999 an extreme frost occurred in the study area. Temperatures as low as -8.3^{o}C were recorded. A map of this frost displaying areas of worst frost would be useful and would test the kriging procedure for the interpolation of temperatures. The minimum temperature recorded by each sensor within the study area was extracted from Topoclimate South's records and used as the data for kriging.

The Idrisi GIS was utilised in preparing the data and ordinary kriging was used to interpolate by using the Gstat program included with Idrisi. Isotropy was found to be present in the data at an angle of 135°. A suitable semivariogram was generated from this data and is shown in Figure 3.1. The model is fitted with a spherical model. Nugget value is 0.673039, range is 14938.1 and the sill is 3.67064. Ordinary kriging was then undertaken by Gstat resulting in Figure 4.1. Cross validation was used to measure the accuracy of the map. This involved the interpolation of a map with one data point removed. Residuals were then generated by subtracting the known value at that location from the interpolated value. This acts as a measure of interpolation accuracy. Cross validation results are contained in Table 1.

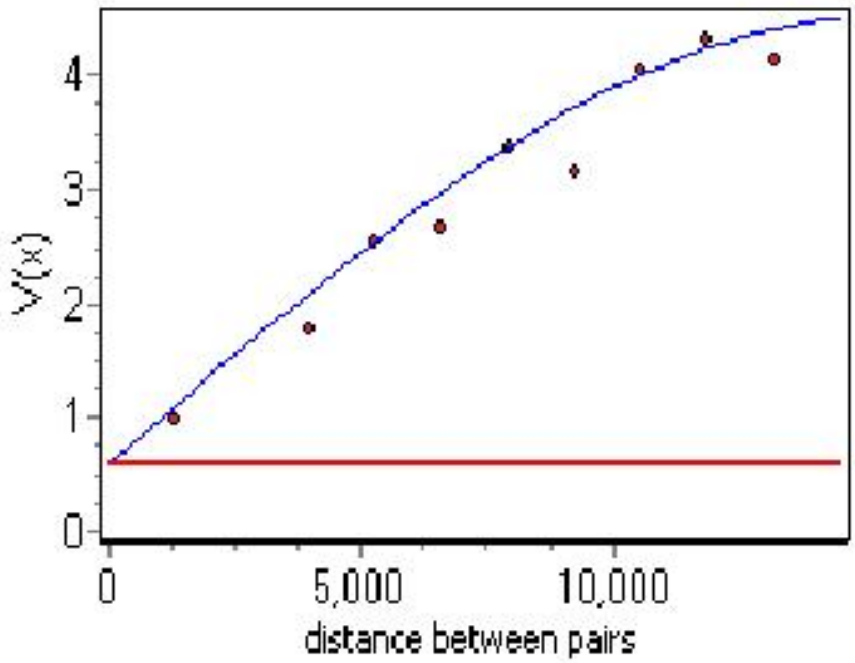

Figure 3.1 Semivariogram and model used for interpolation

4.0 RESULTS

Figure 4.1 displays the resulting interpolated map. The resulting kriged map displays a range of temperatures from 0°C down to -8.3°C. The map shows an area of very cold temperature (less than –8°C) in the centre of the map surrounded by warmer areas (up to 0°C). This clustering of temperatures suggests that they were influenced by topography. Comparison with a topographic map shows that the lowest temperatures were located along the valley floor, whereas the highest temperatures were on surrounding ridges.

The variance map produced in tandem with the kriged surface is shown in Figure 4.2. The map shows that locations close to sampling sites have less variance than locations far from sampling sites. The variance at sampling sites is zero, which is consistent with the kriging process. Variance around the boundary of the study area is relatively high (up to 2.07) due to the lack of sampling sites in these areas. Therefore these peripheral locations are associated with the least confidence in their associated kriging values.

Statistic	Observed	Predicted	Pred-Obs	Prediction std
Minimum	-8.3	-7.1	-1.2	0.9
1st Quartile	-6.3	-6.0	-0.3	1.0
Median	-5.0	-4.6	-0.4	1.0
3rd Quartile	-2.6	-2.7	0.1	1.0
Maximum	0.3	-0.1	-0.2	1.6
N	212	212		
Mean	-4.4	-4.4	0.0	1.0
Standard Deviation	2.2	1.8	0.4	0.1

Table 1 Results of the cross-validation

Table 1 contains the results of the cross validation. The mean residual is 0°C. This shows that, on average there was no difference between the interpolated temperature and the value recorded by sensors at that location. The minimum kriged value was –7.1°C which would have been interpolated when the minimum sampled value –8.3°C was removed during cross validation. Maximum kriged value was –0.1°C which corresponds to the maximum interpolated value produced when the maximum sampled value of 0.3°C was removed from the dataset during cross validation. The standard deviation of the interpolated dataset (1.8°C) is less than that of the sampled data (2.2°C). Correlation between observed and predicted values was reported as 0.8192.

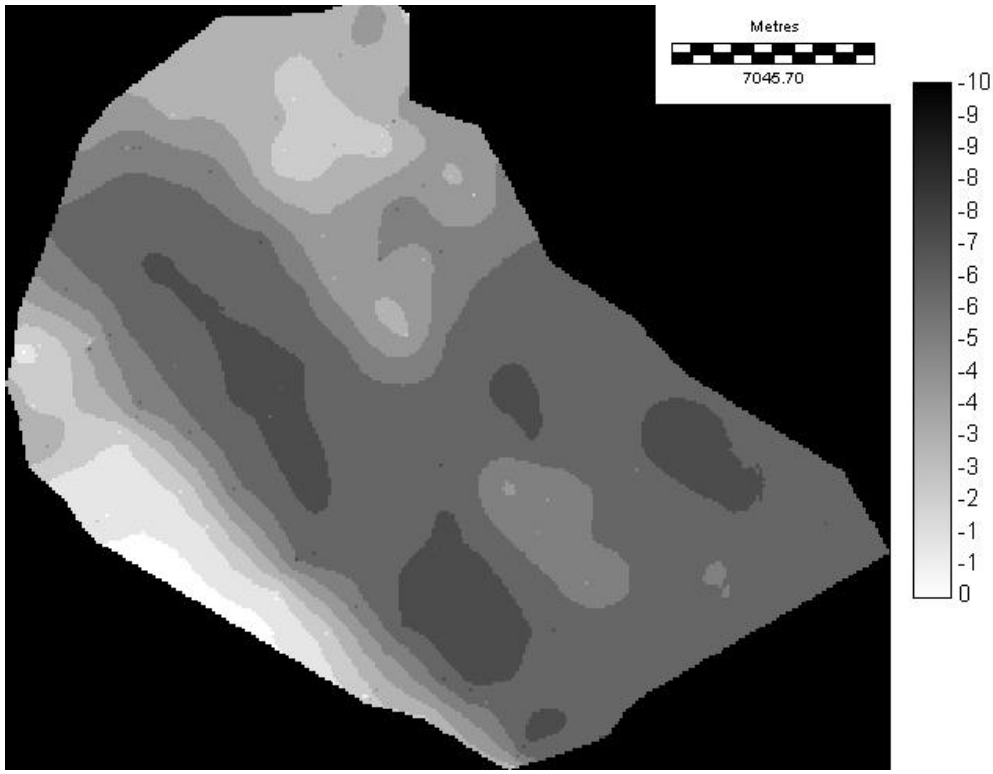

Figure 4.1 Final kriged map (values in ^{o}C)

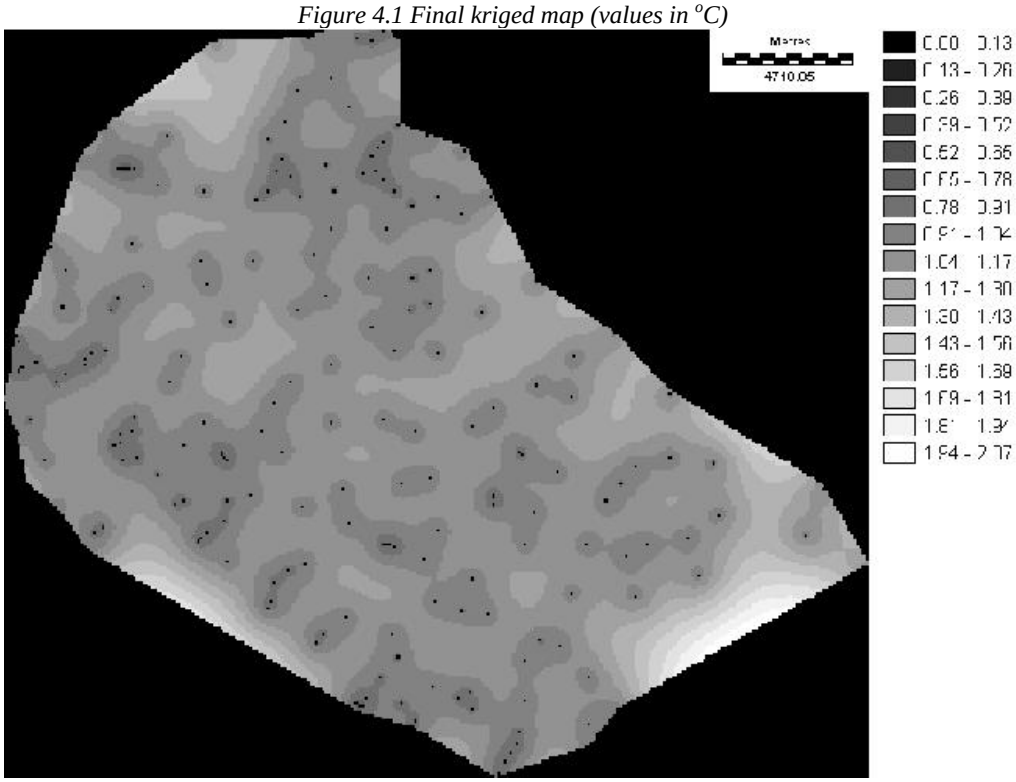

Figure 4.2 Variance map associated with the kriged map

5.0 CONCLUSIONS

Cross validation (Table 1) has shown the interpolated map to be extremely accurate as the mean residual between interpolated and measured temperature was 0.0°C. This shows the usefulness of interpolation and especially kriging to the task of temperature map generation. Table 1 shows the smoothing characteristics of interpolation by kriging. Both minimum and maximum interpolated temperatures were within the sampling range. Additionally the standard deviation of kriging was smaller than that sampled, caused by the smoothing nature of kriging.

The map that has been generated of temperatures in the study area (Figure 4.1) shows the control of topography on temperatures (topoclimate). During the worst frost of 1999, it was hardly freezing on the peak of the North Range, whereas a few kilometres away on the Waimea Plains, temperatures were lower than –8°C. Throughout the study area it can be seen that areas of relatively high altitudes correspond with warm temperatures. The phenomena of increasing temperatures with height is known as an inversion. In this example, the inversion depth is over 400m and encompasses a range of 8°C.

An aim of the Topoclimate South project is to collect data of use to farmers. The analysis of their data by interpolation has resulted in a significant amount of useful information. Areas susceptible to significant freezing have been mapped and are in agreement with climatological theory. That is, cold air collects over the Waimea Plains during harsh frosts. In order to escape freezing conditions, crops or stock may need to be moved to hilltops. A height of around 500m may be suitable for this.

The display of data in graphical forms often assists in the conveyance of important information. It can also assist with the generation of new spatial relationships and confirm the presence of others. This is a strong feature of interpolation as it converts a list of numbers into a graphical and useful form of information. Interpolation also generates new information between sampling sites. Thus the amount of data available for study increases dramatically. It is however important to correctly apply an interpolation technique. Failure to do this may result in poor generation of interpolated data. This impacts upon the quality and validity of the resulting map and any statistical tests or conclusions drawn from such data. Many techniques are available, some are more suited than others to the available data. Those intending to interpolate should consider the methods, assumptions and available data before deciding which method to implement.

6.0 ACKNOWLEDGMENTS

Temperature data used in this project was provided by the Topoclimate South Trust of Mataura, Southland.

7.0 REFERENCES

Burrough P.A. & McDonnell R.A. (1998) *Principles of Geographical Information Systems*, Oxford University Press: New York.

Clark I. (1979) *Practical Geostatistics*, Applied Science Publishers: London.

Cressie N. (1985) Fitting Variogram Models by Weighted Least Squares, *Mathematical Geology* v17, no 5, pp563-586.

Franke R. (1982) Scattered Data Interpolation: Tests of Some Methods, *Mathematics of Computation* v38, no157, pp181-200.

Tobler W.R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography* v46, pp234-240.

Statistics New Zealand (1996*) Census Population and Dwelling Statistics*, Wellington: New Zealand.