TEACHING SPATIAL STATISTICAL TECHNIQUES AND CONCEPTS

Richard Castle
The University of Brighton, United Kingdom
rcc2@brighton.ac.uk

The last decade has seen a rapid increase in the use of Geographical Information Systems (GIS) and the analysis of spatial data is an important component of this development. Spatial statistics is a relatively young subject and, although there are useful textbooks on spatial statistics theory, there is virtually no literature on how to teach spatial statistical concepts and techniques. This paper suggests ways of teaching some of spatial statistical analysis without recourse to matrix algebra and vectors. By using the graphical features in Excel *it is possible to illustrate and explain the concepts behind the statistical techniques in GIS. The interactive and dynamic features of* Excel *enable students to investigate the effects of changing the spatial location of the data and to develop an understanding of spatial dependence and its impact on Kriging and regression techniques.*

INTRODUCTION

Spatial statistics is a relatively new area of statistics which started in geography departments in the 1970s. Spatial statistics has developed rapidly in the last 10 years; the increased processing power of modern computers has enabled the analysis of large data sets using complex numerical techniques.

Spatial statisticians have developed their own techniques for solving problems; firstly because of the nature of spatial data and, secondly, for the types of problems that geographers want to solve.

Spatial data analysis is used when data has been collected at a number of different sites; in addition to the variables of interest, we also have information about the location of each site. It is not possible to assume samples are independent and, therefore, we must use techniques which build the covariance structure into our models.

Spatial data analysis has its own techniques for tackling problems unique to spatial data. For example, Kriging models are used to predict values at different location and thus produce a surface map of the variable under study.

Often the students who study GIS and spatial statistics do not specialise in statistics and consequently are weak at mathematics. This paper shows how 'visual' learning through the use of interactive spreadsheets can be used, alongside the usual GIS packages such as Arcmap, to explain difficult statistical concepts and techniques; it is not necessary to use complicated matrix algebra and vectors. Two techniques are discussed to illustrate the approach – the teaching of covariance structures (semivariograms) and Kriging techniques used to produce predict surfaces.

'VISUAL' LEARNING

Hunt (1994) discussed the use of spreadsheets in teaching statistics and Snee (1993) pointed out the importance of using a variety of teaching methods to develop learning and understanding. Castle (2002) explained how spreadsheets can be used in a variety of ways to aid learning.

Carr (2002) demonstrated that *Excel* "provides an extremely flexible and versatile teaching tool for constructing computer demonstrations" and Cumming (2002) advocated the use of 'live' figures to aid statistical understanding. He pointed out that "animation, multiple representations, engagement through interactivity, and vivid 'take-home' images can all be valuable" learning tools.

Bailey (1995) stresses the importance of explaining the rationale behind spatial statistical methods and the spreadsheets which have been developed enable the lecturer to explain and illustrate spatial statistical concepts and techniques dynamically. Alternatively the student can use these features on the spreadsheet to experiment and explore and thereby develop their understanding.

Two examples of this approach are discussed in the next two sections – spatial dependence and spatial prediction.

SPATIAL DEPENDENCE

Spatial dependence is a natural idea for geography students to grasp: locations that are nearby tend to have similar observations. For example, our experience of weather tells us that rainfall is similar in places only a few miles apart. The map in Figure 1 shows the sample locations in California where rainfall (z) data have been collected. The difficulty for students studying spatial statistics is to understand how this intuitive idea is reflected in a semivariogram.
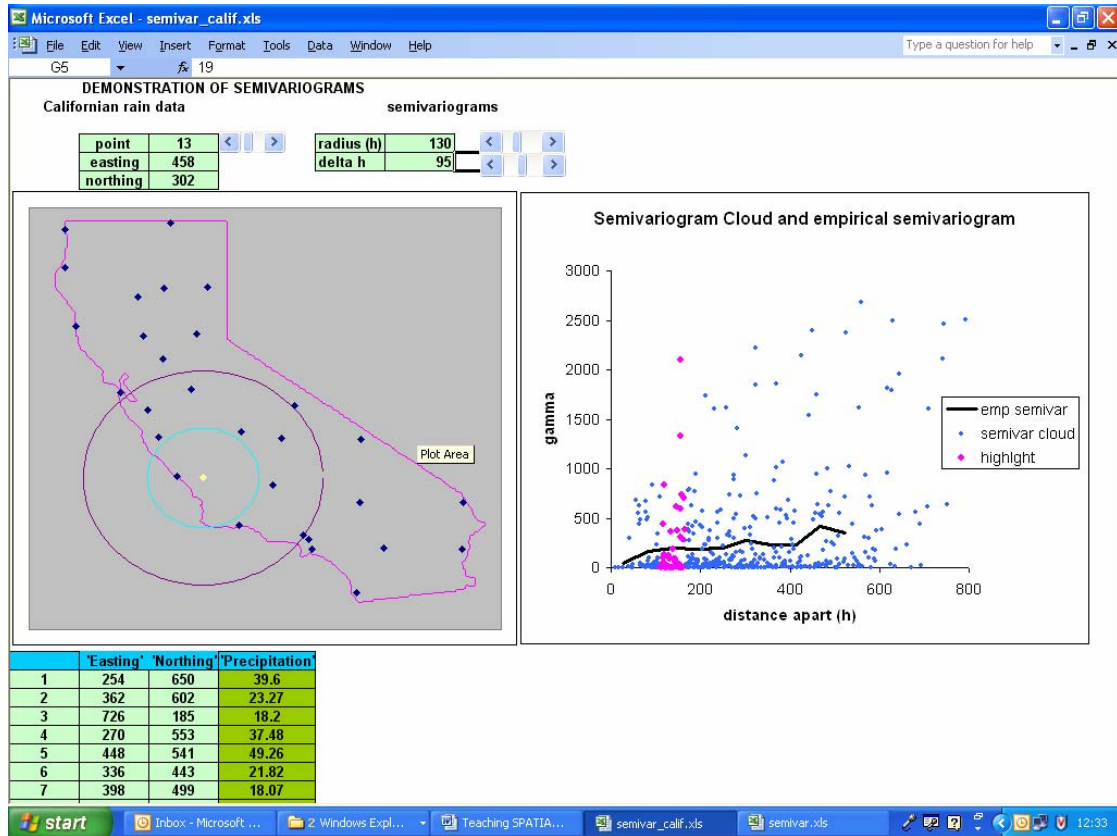

Figure 1: Demonstration of spatial dependence

If z_i and z_j represent the rainfall at two locations, i and j, then the semivariance $(0.5(z_i - z_j)^2)$ is a simple measure of spatial continuity between these two places. If two locations, i and j, are close together then we would expect their semivariance to be small; however, as the distance between locations (h) increases we would expect this measure to increase. The semivariogram cloud is a plot of $0.5(z_i - z_j)^2$ against distance apart (h) for all pairs of locations, i and j.

Figure 1 shows a typical semivariogram cloud and its associated map. It is not obvious to students how this plot is constructed and how it relates to the data on the map. The spreadsheet shows the link between the two graphs. The concentric circles on the map show all the sample points that are (approximately) a fixed distance (h) from the central data point. We can now see how the semivariances for all these sample pairs are plotted at distance h on the semivariogram. Additionally, the empirical semivariogram has been drawn; this shows how the semivariance changes with distance, h, on average. At small distances, h, the semivariance is small and, in a stationary process, this will increase to an asymptotic maximum.

The dynamic features of *Excel* allow instantaneous adjustment of the radius of the concentric circles, or, alternatively different data points can be chosen for the centre of the circles. In this way the student can see how the semivariogram is constructed.

Further dynamic spreadsheets have been developed to show dynamically how the semivariogram is related to covariance and how a semivariogram model can be fitted to the data.

SPATIAL PREDICTION

Surface prediction is a key problem in geography and Kriging is a common technique for achieving this. Text books generally confront the student with the Kriging estimate:

$$\widehat{z}(\underline{s}) = \underline{c}'(\underline{s}) C^{-1} \underline{z}$$

This matrix expression is generally beyond the abilities of most geography students. It is more easily expressed and understood as a weighted average of the observed data values.

$$\widehat{z}(\underline{s}) = \sum_{i=1}^{n} w_i \, z_i$$

This is a much more intuitive idea for the students to grasp. In discussion, the students quickly suggest sensible weighted average estimates. For example, a local sample mean in which only points within a certain distance of the point to be estimated are included in the estimate, or, an inverse distance weighting average in which all data points are included in the estimate but weights decrease as the distance between the data points and the estimation point increases.
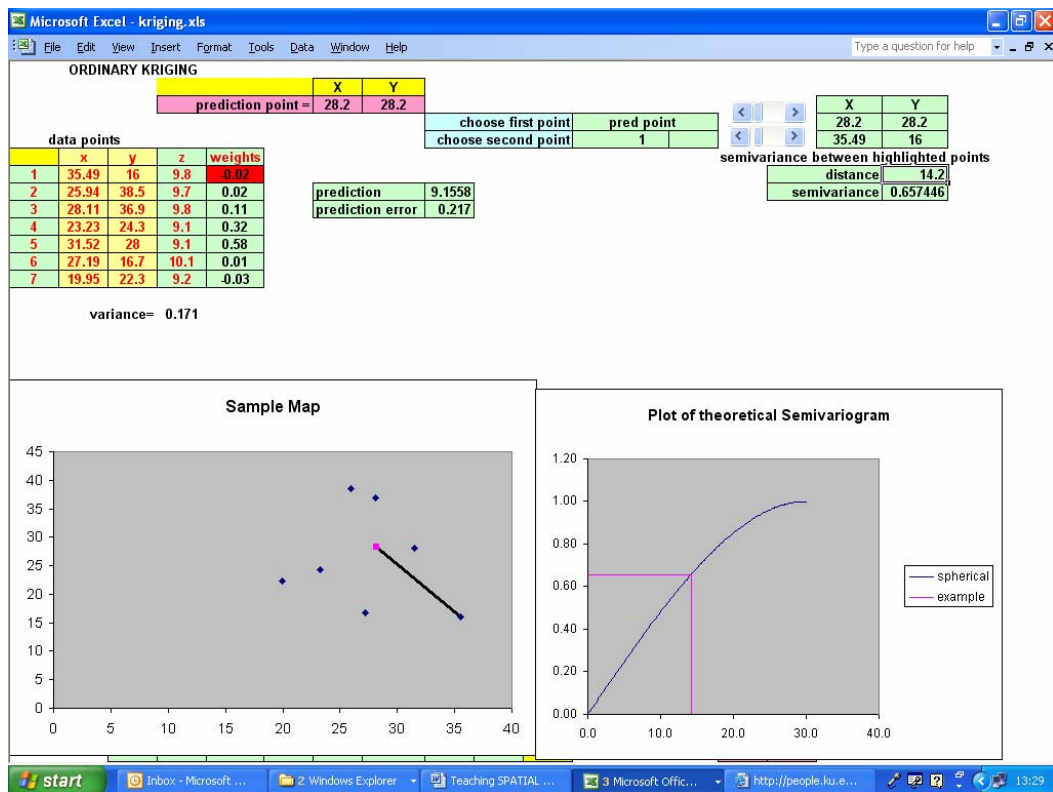

Figure 2: Demonstration of spatial prediction

Kriging is another weighted average estimate and the spreadsheet is a way of illustrating how the weights are calculated. Figure 2 shows a simple example; the graph on the left represents a map comprising of 7 data points and the point to be estimated. The spreadsheet calculates the weights for each pair of data points using a theoretical semivariogram. The scroll bar can be used to show particular pairs of points and the corresponding weight is illustrated in the table (Figure 2). The predicted value and its associated prediction error are also presented.

This, by itself, does not show how the weights are allocated for each data value. But the dynamic features of *Excel* allow data values to be moved. By clicking on a particular data value on the left hand graph, the data value can be moved to a new location and, in this way, it is possible to see how this affects the weights, and, in turn, how the estimated values and prediction error are affected. The weight of a data point will increase as the data point moves closer to the point to be estimated. As the data point is dragged closer to the prediction point the semivariance between them decreases; this is illustrated in the right hand panel in Figure 2.

Further investigations can be carried out. For example, the prediction error of the estimate will increase as the prediction point is moved away from the rest of the data. Alternatively, it is possible to investigate the effect of clustering of data values. The weights of points in a cluster are reduced because the data values are acting as surrogates for each other and not providing unique information.

DISCUSSION

Although there are now an increasing number of explanatory texts on spatial data analysis there is virtually no literature on how to teach the concepts.

Visualisation is key to GIS and by using the graphical and dynamic features in *Excel* it is possible to illustrate and explain the concepts behind the statistical techniques in GIS. The interactive features of *Excel* enable the student to investigate the effects of changing the spatial location of the data and thereby develop an understanding of spatial dependence and its impact on Kriging and regression techniques.

By using spreadsheets it is possible to illustrate key concepts visually. By dynamically investigating spatial data, the key concepts of spatial dependence and surface prediction can be explain and explored.

Further spreadsheets have been developed to explain spatial statistical concepts; these can be downloaded from the author's website (http://www.it.bton.ac.uk/staff/rcc2/).

REFERENCES

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis.* Upper Saddle River, NJ: Prentice Hall.

Bradstreet, T. E. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *The American Statistician*, 50, 69-78.

Carr, R. (2002). Teaching statistics using demonstrations implemented with *Excel.* In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics,* Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Castle, R. (2002). Using interactive visualisation to develop statistical understanding In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics,* Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Cumming, G. (2002). Live figures: Interactive diagrams for statistical understanding. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Hawkins, A., Joliffe, F., and Glickman, L. (1992). *Teaching Statistical Concepts.* London and New York: Longman

Hunt, N. (1994). Teaching statistics using a spreadsheet. In *Proceedings of the Fourth International Conference on Teaching Statistics*, Morocco, Vol. 2, p. 432.

Snee, R. D. (1993). What's Missing in Statistical Education? *The American Statistician*, 47, 149-154.