GIS and Spatial Statistics: Recent UK experiences and developments.

Derek Bond  
Northern Ireland RRL  
University of Ulster at Coleraine  
Coleraine, United Kingdom  
D.Bond@ulst.ac.uk  

Peter Brown  
Urban Research & Policy RRL  
University of Liverpool  
Liverpool, United Kingdom  
pjbbrown@liverpool.ac.uk  

Robin Flowerdew  
North West RRL  
Lancaster University  
Lancaster, United Kingdom  
R.Flowerdew@lancaster.ac.uk  

1. Introduction

The uptake and use of GIS in the public sector in the United Kingdom was critically reviewed in a recent paper by Worrall & Bond (1996). Their main conclusion was that the high hopes about the use of GIS that had been prevalent during the late 1980s had failed to materialise. This they argued was due mainly to organisational and implementation issues but also highlighted the lack of effective spatial analysis techniques available to practitioners who have made the heavy investment needed in GIS.

In this paper some of the practical developments in spatial analysis that have emerged from the work of the Regional Research Laboratories in the United Kingdom will be outlined. In particular, the issue of whether these approaches provide cost effective solutions justifying the large development costs associated with GIS is discussed. The Regional Research Laboratories were established in 1988 by the United Kingdom’s Economic and Social Research Council (ESRC) to provide a decentralised network of centres of excellence in the handling of large scale datasets. Given the timing of their establishment it seemed reasonable that they should, at least initially, concentrate on the handling of spatially referenced datasets. The funding model for the RRLs was also interesting and experimental, as after the initial funding from the ESRC they were expected to be self-financing. Given the extremely competitive environment for ‘blue skys’ research money this meant that from the start the RRLs were also involved in ‘practical’ research. They have thus been leaders in areas such as geodemographics, spatial indicators, GIS interfaces, policy evaluation and implementation issues. Further detail of the work programmes of the RRLs can be found in Bond (1999), Bond & Robinson (1995), Bond & Robinson (1994), and Walker (1993).

In Worrall and Bond (1996) an extended definition of GIS was given:
‘a computer-based system for capturing, storing, manipulating, analysing and visualising spatially referenced data and integrating it with other computer based information; a toolkit for the modelling and analysis of complex research, management and planning problems; and a system to support decision makers by enabling them to structure problems and identify potential solutions for evaluation’

(Worrall and Bond (1996) p. 366)

This paper will concentrate on the manipulation, analysis and visualisation aspects of GIS to show how the work of the members of the RRL network has helped develop the available toolkit for policy makers.

2. Visualisation Issues
One of the earliest selling points of GIS was its visualisation capabilities. This was normally to produce detailed computer based maps of utilities location or choropleth maps of socio-economic variables. Some of the early work in the RRLs was on the cognitive aspects of GIS, c.f. Hearnshaw and Unwin(1994). With the growth of windowing environments and in particular web based interfaces the use of maps as an interface with complex datasets has developed. Various RRLs have been involved in this research and one of the best developed products is the Cas-web interface to the 1991 Census of Population. Casweb is a web-based interface to the 1991 Census "area statistics" SAS/LBS). These large and complex datasets comprise aggregate counts of persons and house-holds for various geographical units. Extracting data from Casweb is a four step process. First the area is select, then the data required, complex expressions can be built and finally the data retrieved.

Example of Cas-web interface

Other work conducted by the RRLs on visualisation range from the use of exploratory techniques c.f. Haslett(1994) to using DTM models to assess the visual impact of wind farms c.f. Kidner (1997).

3.Manipulation and Analysis

The growth of GIS and related spatial data handling facilities have led users to demand greater flexibility from providers of statistics in terms of the areas for which datasets are made available. Considerable work has been done by the RRLs in this area. For example The North East RRL has been in the forefront of developments of defining ‘Travel to Work Areas’-c.f. Coombes(1992) and ONS & Coombes(1998) - unfortunately as Coombes (1998) concludes: ‘...the methods and software which are needed for defining new sets of areas -...- are still too specialised to be found in many GIS packages’.

Other RRLs have also researched this general area. For example the North West RRL has looked at the issue of the modifiable areal unit problem and spatial regression. This has involved using data from the 1991 Census of Population to investigate the relationship between ethnic group and unemployment rates in Northwest England. The results have shown that while correlation coefficients will always suffer from the modifiable areal unit problem effect in the presence of spatial autocorrelation the regression coefficients only suffer if there is also spatial cross-correlation between the response variable and the explanatory variables. However this effect can be modelled by having two effects a local effect and a regional effect. Their work has shown clearly that the results are biased unless the regional effect is also taken into account -c.f. Green & Flowerdew(1996). The work, sponsored by the ESRC, took account of the zonal effect of the MAUP.

The North East RRL are at present researching a related topic that of Geographically Weight Regression a technique for exploratory spatial data analysis. GWR permits the parameter estimates to vary locally; so the standard regression model can be rewritten as:

\[ y(g) = \beta_0(g) + \beta_1(g)x_1 + \beta_2(g)x_2 + e \]
where (g) indicates that the parameters are to be estimated at a location whose co-ordinates are given by the vector g. The parameter estimates for GWR may be solved using a weighting scheme:

$$\mathbf{b}(g) = (\mathbf{X}^\top \mathbf{W}(g) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}(g) \mathbf{Y}$$

The weights are chosen such that those observations near the point in space where the parameter estimates are desired have more influence on the result than observations further away. Two functions we have used for the weight calculation have been (a) bi-square and (b) Gaussian. In the case of the Gaussian scheme, the weight for the i-th observation is:

$$w_i(g) = e^{-d(h)^2}$$

where d is the Euclidean distance between the location of observation i and location g, and h is a quantity known as the bandwidth. (There are similarities between GWR and kernel regression). One characteristic that is not immediately obvious, is that the locations at which parameters are estimated need not be the ones at which the data have been collected. The resulting parameter estimates may be mapped in order to examine local variations in the parameter estimates. One might also map the standard errors of the parameters estimates as well. Hypothesis tests are possible - for example one might wish to test whether or not the variations in the values of a parameter in the study area are due to chance. The bandwidth may be either supplied by the user, or estimated using a technique such as crossvalidation.

Much of the work of the RRLs has involved applying research to practical problems. For example the Wales and South West RRL have developed GIS-based methodologies which can be employed in rural areas to quantify aspects of disadvantage experienced by rural dwellers. They have compared two indicators based on measures of accessibility with standard indicators of disadvantaged. The results of the study has illustrated possible methods of using GIS to help quantify the problems which rurality presents.

A related and extremely practical area for GIS is the field of geodeomographics. Here the Urban and Regional Policy Evaluation RRL based at Liverpool has been very much in the vanguard. Their Super Profiles are topology product developed in collaboration with CDMS Limited is a general purpose classification of the 140000 census enumeration districts in Great Britain -c.f. Brown and Batey(1994) These have been used in many projects, apart from the obvious private sector applications, including health, community safety, higher education, parliamentary constituency profiling and electricity demand forecasting - cf. Brown et.al (1998).

4. Concluding Comments

It is clear from the above that considerable work has been undertaken to develop spatial analysis techniques so that the potential of GIS might be realised. However, the question has to be asked as to whether these are adequate, assuming that management and organisational issues can be overcome, to ensure a successful role for GIS in official statistics? As yet most of these techniques are not readily available to practitioners. The RRLs are currently responding to the Europe Union’s current call for research in Official Statistics with a proposal to bring together the various strands of their work with that of European colleagues to provide official statisticians with a powerful but simple to use GIS toolbox for handling the spatial dimension of official statistics.

REFERENCES


RÉSUMÉ

Cet article trace les grandes lignes d'une partie de la recherche récente conduitee par le RRL.net dans la zone de l'analyse spatiale et consièrre la question de si ces techniques peuvent aider à accomplir la promesse tôt de GIS.