INT. J. GEOGRAPHICAL INFORMATION SCIENCE
VOL. 18, NO. 5, JULY–AUGUST 2004, 447–457

Taylor & Francis
Taylor & Francis Group

Research Article

Some thoughts on inference in the analysis of spatial data

A. STEWART FOTHERINGHAM

National Center for Geocomputation, National University of Ireland, Maynooth, Co. Kildare, Ireland

CHRIS BRUNSDON

School of Computing, University of Glamorgan, Pontypridd CF37 1DL, UK; e-mail: cbrunsdo@glam.ac.uk

(*Received 18 February 2003; accepted 5 October 2003*)

Abstract. Statistical inference is important for all those who engage in the analysis of spatial data. The issue is becoming increasingly important given the explosion in the availability of spatial data and the proliferation of Geographic Information Systems (GIS) across different academic disciplines and application areas. The aim of this paper is to provide a brief overview of some of the concepts and controversies inherent in statistical inference in the hope of raising the level of awareness within the geographic information science community that different points of view exist when it comes to inference. We argue that the concept of statistical inference in spatial data analysis and spatial modelling is perhaps broader than many GIS users imagine. In particular, we argue that different types of inference exist and that process inference is just as valid as sample inference, even though the latter appears to dominate the GIS literature.

1. Introduction

The notion of inference is important for all those who are involved in the analysis of spatial data. It is extremely useful to make informed statements about something that, ultimately, is unknown: it is equally useful to consider how likely these statements are to be wrong. A great deal of research is based on the collection of data, both qualitative and quantitative, in order to understand processes that cannot be observed or to draw conclusions about larger numbers of unobserved subjects. In both cases, one must generally engage in an inferential process. Here, we primarily consider the role of statistical inference in spatial data analysis.

This is a useful time to highlight this topic given the explosion in the availability of spatial data and the proliferation of Geographical Information Systems (GIS) to assist in the analysis of such data. It is particularly pertinent given that discussions of inference are relatively rare in GIS texts. One recent exception to this, which is to be applauded, is that by Longley *et al.* (2001). However, even there, in line with what we suspect is a fairly widely held view in the GIS community, the authors espouse the view that there is *one* method of statistical inference, based on classical

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810410001658065

Neyman–Pearson concepts of significance testing. The notion either that this method might not be applicable in some instances or that it might just be one of several competing and equally valid methods of statistical inference is not generally considered. However, within the statistical community, there is much debate about the relative validity and utility of a variety of approaches to statistical inference (*inter alia*, Nester 1996). These debates need to reach a wider audience within the Geographical Information Science (GISc) community so that alternatives to the Neyman–Pearson framework of inference might be considered in the analysis of spatial data.

Thus, the aim of this paper is to provide a brief overview of some of the concepts and controversies of statistical inference in the hope that this might raise the level of awareness of these other approaches within the spatial analysis and GIS communities. We do this by asking two major questions: 'Why do we need to infer?' and 'How do we infer?'.

2. Why do we need to infer?

There appears to be a fairly widely held belief among geographers that statistical inference is a valid method only when one is making inferences about a population from a sample. This is perhaps most explicitly stated in the following passage:

Before using inferential tests on spatial data, therefore, it is advisable to ask two fundamental questions:

- Can I conceive of a larger population that I want to make inferences about?
- Are my data acceptable as a random and independent sample of that population?

If the answer to either of these questions is no, then inferential tests are not appropriate (Longley *et al.* 2001: 322).

These certainly are important questions, but they are not the only questions to ask in terms of inference, and the conditional statement that follows the questions is therefore wrong. A third question should be added to the above list: 'Can I conceive of a data-generating process about which I want to make inferences?'.

Neither of the original two questions, for example, addresses the need to make inferences about *processes* as well as populations, and neither recognizes that inferences about processes can, and should, be made on the basis of population data. To make the difference between these two types of inference explicit, we will refer to them as *process inference* and *sampling inference*. In process inference, the null hypothesis is a statement about the data-generating process rather than about a population. In sampling inference, the null hypothesis is about the characteristics of an unknown population. It would seem that most GIS researchers are more familiar with sampling inference than with process inference, but both are equally valid forms of inference.

To see how process inference can be applied, and indeed is the only form of inference that is plausible in some instances, suppose a newly manufactured six-sided die were thrown 100 times and the score recorded at each roll. At the conclusion of the experiment, the die is destroyed: the act of destruction ensures that the 100 recorded numbers constitute a population data set. Despite the fact that we now have a 'population' data set, it is still reasonable to use inferential tests

here to examine the issue of whether the die was a true one or not. This is an example of process inference. The task of the analysis is to determine the probability that the die was fair (that is, that the probability of obtaining any given score was one-sixth).[1] Notice that although a population is being analysed, it is still reasonable to draw inferences about the process that produced the population.

The example above may appear contrived, but it does demonstrate the nature of process inference and how it is used in practice. Inference in this context is used as a type of benchmark. The actual set of die scores is compared with that which would be obtained in an ideal situation—the perfectly fair die. The significance level provides a measure of how close the real scores are to those expected from this ideal. In the real world, this interpretation is extremely useful when considering many types of spatial data, such as those on the spatial distribution of diseases, for example. Suppose we examine the spatial incidence of some type of disease, and we plot the home addresses of those with the disease on a map. Again, we may have a 'population' of disease occurrences, but it is still useful to ask: 'How likely is it that the observed spatial pattern of the disease has arisen from a random process given the underlying spatial distribution of the population-at-risk?'. Obviously, if the data are very unlikely to have been generated by a random process, this will lead us to explore different causal mechanisms from the situation in which the data appeared to result from a random process. We could determine the probability that the data have been generated by a random process by comparing the actual data with those expected from a Poisson process where the expected values in each zone are proportional to the population at risk (for an example, see Fotheringham and Zhan 1996). The Poisson process provides a benchmark of 'randomness'. It is therefore useful in such a situation to consider statistical inference, even though the data set might constitute a population.

A further example of inference not based on populations is found in randomization testing. As Manly (1991) and Jacquez (1999) observe, randomization methods allow inference to be made about patterns in a *sample* regardless of the population from which it is drawn. For example, consider a variable, x, distributed across a set of spatial units $\{x_1, x_2, \ldots x_n\}$. Suppose we are interested in the degree to which similar values of x are found in adjacent locations. We could compute some measure of spatial autocorrelation for the distribution of x and examine the null hypothesis that the observed distribution of x has arisen from a random process. If the null hypothesis is correct, any permutation of x-variables among a set of geographical zones would be equally likely. We can randomly rearrange the values of x across the n zones and recalculate the measure of spatial autocorrelation. Repetition of this process allows us to build an experimental distribution for the spatial autocorrelation statistic from which we can make some inference about the observed distribution. If the statistic based on the observed distribution of x is sufficiently distant from the mean of the experimental distribution, we may claim that the spatial distribution of the observed data is highly unlikely to have arisen from a random process. Conversely, if the statistic for the observed distribution is not sufficiently far from the mean of the experimental

[1]Given that 100 is not an integer multiple of 6, it is impossible to obtain equal or 'fair' proportions of each possible score in this population data set, so a simple description of the population would automatically lead to the conclusion that the die is not fair, regardless or whether it is or not!

distribution, we would fail to reject the null hypothesis and conclude that a random process is in operation. Once again, the null hypothesis is not about a population but about a process.

3. How do we infer?

The above discussion highlights the situations in which inference is useful. It claims that process inference is just as useful as sampling inference. Here, we consider the different ways of making inferences from statistical data regardless of the situation in which the inference is being made.

Inferential methods can be categorized in several ways; here, we have chosen to divide them into two types: *formal* and *informal*, with the former being subdivided into *theoretically formal* and *computationally formal*. Formal approaches are generally used to test a hypothesis that has been suggested by theory before the data have been collected. When using a formal approach, one attempts to quantify the plausibility that the observed data lead to the hypothesis. A part of formal inferential tests is a set of 'rules' which, although perhaps arbitrary, if applied consistently will lead different people to reach the same conclusion about whatever hypothesis is being examined. With theoretically formal methods, given the same data and the same set of rules, everybody will reach the same conclusion. For instance, if 100 researchers are asked to calculate a p value for a t-test associated with the difference in two sample means using the same data and if they are given the same set of rules (e.g. that the null hypothesis is rejected is the p value is less than 0.05), they should all reach the same conclusion because the test makes reference to a standard theoretical distribution. In computationally formal tests, the reference distribution is generated computationally from the data and not from a theoretical model. Hence, different researchers are likely to generate different reference distributions and therefore are not guaranteed to reach the same conclusion about the null hypothesis. The p value used in the computationally formal test determines the proportion of researchers (in an infinite sample) who would not reach the same conclusion as the majority. Informal approaches are less rigorous and much more subjective: the aim is simply to provide some information on the system under investigation rather than to test a specific hypothesis. For instance, graphical techniques may be used to display data so that any patterns become apparent. However, there is no 'black and white' test available, and what one person 'sees' in the data might be quite different from what another sees. For example, looking at a scatterplot of points in which variable Y is plotted against variable X, one researcher might conclude that there is a relationship between Y and X, whereas another might conclude that no such relationship exists. The objective in devising new methods for informal inference is to develop ways of presenting data that maximize the proportion of reasonable conclusions that are reached regarding relationships within the data. For instance, mapping data based on administrative boundaries, where the size of units is often inversely related to their populations, can often be misleading, and cartogrammetric representations might offer the chance of more accurate conclusions being drawn regarding trends in the data (Tobler 1973, Dorling 1991). Prescribed hypotheses could be considered in informal tests, but informal approaches are perhaps more useful as *generators* of hypotheses. Identifying patterns in the observed data, for example, can often suggest new research questions. We now consider these different methods of inference in more detail.

3.1. *Theoretically formal tests*

Theoretically formal tests of significance fall into two main types: those which lead us into a discrete decision which is either to 'reject' or to 'fail to reject' a null hypothesis and those which indicate which is a better model of the data being analysed. Of the former type, there is a distinction between *Classical* (sometimes also referred to as '*Frequentist*') and *Bayesian* modes of inference.

Classicial inference

The vast majority of examples of statistical inference in spatial analysis and GIS still make use of classical inference, which involves ideas of confidence intervals and significance testing. In the context of hypothesis testing, the difference between the Bayesian and the Classical approaches is very clear. Regardless of whether probabilities are subjective or objective, a Bayesian draws conclusions of the form 'the probability that hypothesis X is true is r'. In classical inference, hypothesis X (the *null hypothesis*) is not regarded in probabilistic terms. It is assumed to be a state that is either true or false. To determine the truth or falsehood of the null hypothesis, a numerical test is applied. This returns a verdict of 'fail to reject' or 'reject', but of course, this verdict could be wrong, and one can consider the probability of being wrong. Thus, in classical inference, there are four possible outcomes, each with an associated probability (table 1).

The probability a is referred to as the *significance level* of the test, and the probability b as the *power* of the test. Thus, the probabilities in classical inference do not refer to the truth of the null hypothesis but refer to the success of the testing process. They are *operating characteristics* of the test. The significance level of the test is fairly easy to specify, provided one can specify the probability of the test outcome in terms of the null hypothesis. There are a vast number of examples of tests of this kind in many standard statistical texts—typical are tests of whether a particular model parameter is equal to zero. The power is generally harder to specify. Typically, the null hypothesis is a simple mathematical statement such as a particular regression coefficient being zero. However, there are an infinite number of alternatives to this, and the power of the test will depend on how different the regression coefficient is from zero in addition to the sample size. Classical inference can also be applied to parameter estimation where confidence intervals are constructed denoting a range of values having the probability 1–a of containing the true parameter.

One might ask why alternative forms of statistical inference exist and why some people prefer not to follow the classical approach to inference. Although the classical approach is very frequently used, it is not without its problems. Most opponents of the classical approach identify problems with hypothesis testing. The main problem is that most null hypotheses are expressed in the form $1 \sim 0$, or some

Table 1. Four possible outcomes in classical inference.

Outcome	Probability
Null hypothesis is true, and test result is 'fail to reject'	1–a
Null hypothesis is true, but test result is 'reject'	a
Null hypothesis is false, and test result is 'reject'	b
Null hypothesis is false, but test result is 'fail to reject'	1–b

other specific number, where l is a parameter. If the sample size is sufficiently large, a classical hypothesis test would reject the null hypothesis of l~ 0 for only very small differences between the estimated parameter and its value under the null hypothesis. This serves as a reminder that there can be a difference between statistical significance and real-world importance.

Another problem with classical inference is demonstrated in the following example. Suppose a six-sided die were thrown 24 times, and the only time a six was thrown was on the 24th throw. How reasonable is this for a fair die? The problem could be investigated using classical inference. Since we only have information about sixes being thrown, we could estimate the probability of a six being thrown on any roll. Call this h. We would estimate the quantity h by computing the proportion of sixes in the sample. Call this sample statistic t. Here, t~ 1/24. To test for fairness, we need to test the null hypothesis, H_o: h~ 1/6, against the alternative, H_a: hv 1/6.

In this case, it is assumed that the concern is whether the die provides *enough* sixes, so the test is one-tailed. Under the null hypothesis, the distribution of the number of sixes thrown will be binomial with parameters n~ 24 and h~ 1/6. n is the number of trials, and h is the probability of a success in any one trial. If we let S be the number of successes (sixes thrown), then

$$P(S\sim s)\sim C_{ns}h^{s}(1\{\ h)^{n\{\ s}$$

The lower one-tailed significance of our observed test statistic t is $P(t$f 1/24$|H_o)$, which is equivalent to $P(S$f 1$|H_o)$. Using the binomial distribution formula, this is approximately 0.073, which is the p value of our test. Common practice at this stage is to note that p exceeds the magical value of 0.05 and conclude that our result is not unreasonable for a fair die.

However, now suppose that the experiment was not to throw the die 24 times but instead to throw it *until a score of six occurred*. In this case, the number of sixes thrown, S, is fixed at one, and n, the number of trials, is the random variable. Here, n has a negative binomial distribution. In this case, $P(t$f 1/24$|H_o)$ is equivalent to $P(n$o 24$|H_o)$. This can be shown to be approximately 0.010. Thus, the result now seems to be highly significant, and we would conclude that we do not have a fair die!

Thus, the method of data gathering seems to affect the significance as well as the experimental outcome itself. In the above example, the same physical events happened for both analyses—from a scientific viewpoint, the experimental outcomes were identical. The only differences were in the intentions of the experimenters about how the experiment would be terminated—either throw the die 24 times, or throw it until a six was scored. The consequences are quite controversial and determine whether or not we conclude that we have a significantly low proportion of sixes. Hence, we can make the statement that: In classical inferential theory, significance depends not only on the observed data but also on the data collection policy.

This can be important to those who use GIS because it is sometimes possible to vary the form of data collection in the course of an experiment. For example, if fieldwork required to collect data is deemed too dangerous or too costly during the course of an investigation, it may be curtailed. Similarly, if, during an interview process, it becomes apparent that a particular question is too sensitive or too controversial, that question might be dropped. In such instances, it is difficult to derive formal inferential statistical tests for the variables affected.

Bayesian inference

One inferential framework that avoids these problems because it contains no concept of a sampling set is that of Bayesian inference. In Bayesian inference, rather than considering a *test statistic* as a random variable, the *model parameters* are considered as the random variables. The output of a Bayesian analysis is a probability distribution for the model parameters (termed the *Posterior* distribution). In the die example, the model parameter is h. We would use the distribution of h to assess hypotheses about the procedure producing the observed data. For example, we could find the probability of hf 1/6. Since sampling sets are not involved, the data-collection policy does not influence the determination of this probability. However, this is not without its price, since Bayesian analysis does require input from the analyst in the form of a *Prior* distribution, which represents the analyst's *a priori* beliefs about the values of the parameters before data collection. One way of combating this, assuming the analyst wishes to aim for an objective analysis, is to use a so-called *non-informative* prior, representing a state of no prior belief that any one set of parameter values is more likely than any other. In the die example, this would be a uniform distribution for h in the range 0–1. Fotheringham *et al.* (2000: chapter 8) provide a detailed account of Bayesian inference in a geographical context.

Selecting the best model

The methods we have considered until now, whether they be classical or Bayesian, all work with the notion that there is a 'true' model that can be either tested (significance testing in classical inference) or calibrated (confidence intervals in classical inference, posterior distributions in Bayesian inference). However, this is not always a helpful notion. In some instances, we do not really believe that x is exactly normally distributed, or that y is exactly linearly related to x. In these situations, we adopt such models more because they are close to reality, and therefore reasonable predictors or providers of explanations, than because they represent reality exactly. In this case, most conventional approaches to inference are flawed. What intellectual ground do we have for testing some hypothesis when we have *a priori* reasons for thinking that it is not strictly true? A more practical approach in these situations is to consider a number of 'candidate' models and attempt to decide which of these is closest to reality. The term *closest* deserves emphasis here—it is not assumed that any of the models are perfect, just that some reflect reality better than others. The inferential task here is identifying the best model, not the 'truth'. This approach is that adopted by Akaike (1973) and others, and is outlined very clearly in Burnham and Anderson (1998). The general idea is that statistics can be constructed, consisting of a goodness-of-fit term and a penalty term. The penalty term reflects model complexity, typically as a function of the number of parameters in the model. These statistics are used as measures of the information lost when approximating reality with a given model. The key idea is that the data are generated by the 'true' model (regardless of what that is), and the statistic is a measure of how well a given model measures up to this in information-theoretic terms.

This approach is one that is generally overlooked when applying ideas of statistical inference to spatial problems but could often be an appropriate one. In many practical cases, we have no compelling theoretical arguments for the functional forms we use in our models or for the distributions used when considering

random errors in models; in such situations, we are forced to specify *ad hoc* models. In other cases, we are faced with competing theories—and all theories arise from simplifications of reality. In either of these situations, an approach to inference that attempts to find the closest model to reality from a set of competitors has promise as a practical quantitative tool.

3.2. *Computationally formal tests*

In theoretically formal classical inference, a test statistic is calculated from the sampled data set and then compared with a theoretical distribution with known probability properties (for example, a normal distribution). On the basis of this comparison, we either reject or 'fail to reject' the null hypothesis according to some *a priori* and arbitrary cutoff point, and we can also calculate the probability that we rejected the null hypothesis when it is true. Alternatively, and often preferably, we can derive a confidence interval for the parameter which consists of a range of values within which the unknown population value of the parameter lies with a stated degree of confidence. Whichever approach we take to making inferences using this classical approach, it is necessary to be able to assume some form of theoretical distribution for the test statistic. For some statistics, such as the sample mean and ordinary least-squares parameter estimates, the theoretical distributions are well known and, in most circumstances, can be used with confidence that the assumptions concerning the distributions are met. However, for some statistics, either there is no known theoretical distribution against which to compare the observed value or, where the distribution is known, the assumptions underlying the use of that particular distribution are unlikely to be met. In the former case, a formal theoretical test may still be possible (for example, a Mann–Whitney U-test for comparing two sets of rankings when the underlying distribution is unknown); however, there are other instances when no such formal theoretical test is available. In these instances, which are common in the analysis of spatial data, the construction of experimental distributions is especially useful (*inter alia*, Hope 1968, Costanzo 1983, Diaconis and Efron 1983, Efron and Gong 1983, Efron and Tibshirani 1986, Mooney and Duvall 1993). We term this 'computationally formal' inference here because although it relies on the procedures adopted in classical inference, it replaces a theoretical distribution of the test statistic with one computed from the data being analysed.

The central idea in the use of experimental distributions for statistical inference is that the sampled data can yield a better estimate of the underlying distribution of the calculated statistic than making perhaps unrealistic assumptions about the population. The sampled data are resampled in some way to create a set of samples, each of which yields an estimate of a particular statistic. If this is done many times, the frequency distribution of the statistic forms the experimental distribution against which the value from the original sample can be compared. Consequently, experimental distributions can be constructed for any statistic, even if the theoretical distribution is unknown as is the case, for example, with statistics such as the mean nearest-neighbour distance for an irregularly shaped study area. Fotheringham *et al.* (2000) provide an example of the use of experimental distributions to assess the significance of a spatial autocorrelation coefficient and compare the results with classical tests.

3.3. *Informal tests*

Informal tests are those that rely on a subjective interpretation of data. Frameworks that encapsulate this concept include non-visual exploratory data analysis techniques (Tukey 1977, Ehrenberg 1982), visualization (Cleveland 1993, Hearnshaw and Unwin 1994) and data mining methods such as cluster analysis and neural networks (Hertz *et al.* 1991, Gurney 1995). These give the user a 'feel' for trends within the data and exceptions to these trends. Although they are not usually thought of as inferential tools, and indeed some might argue that they are not, they can provide a useful guide to inference. For instance, simply mapping a data set on disease incidence against the population-at-risk can be sufficient for us to infer that there is a clustering process taking place and that the incidence of the disease is highly unlikely to be random. In very large and complex data sets, more sophisticated data mining techniques such as various forms of cluster analysis or neural networks can be used to infer something about the data. A 'good' informal inference procedure is one that would lead to a large proportion of users drawing the correct inference from the data. Some methods are clearly better than others in this regard but ultimately we often do not feel comfortable relying entirely on what we know are subjective decisions. Even when the inference drawn from a data set is virtually without doubt, we still like to undertake a formal test in order to convince ourselves and to convince others. Additionally, if the informal inference procedure relies on a mapped or graphical representation of the data, issues regarding visual perception need to be taken into account (Tufte 1983, MacEachran *et al.* 1998) and if the data are being displayed dynamically, further issues on how individuals perceive animated displays need to be considered (Dorling and Openshaw 1992, Cook *et al.* 1998). We have all been fooled by visual tricks of display, and the reluctance to rely entirely on our visual perception of trends and relationships is strong.

4. Summary

We argue here that the concept of inference in spatial data analysis and spatial modelling is perhaps broader than many GIS users imagine. Equally, we argue that different types of inference exist and that process influence is just as valid as sample inference, the latter being more commonly examined in geographical studies. Ultimately, inference is similar to the presentation of evidence in a court of law: we use it to try to convince people of a point of view. For this reason, total reliance on informal inferential procedures is often discouraged, even though in many cases the trends or relationships exhibited by the data are obvious to any reasonable person.

In a similar vein, it is interesting to note the longevity of the 'magical' p values such as 0.05 and 0.01. Despite the fact that almost all researchers use high-speed computers and statistical packages that routinely display exact p values associated with a test statistic, there is still a great use made of formal inferential tests in which the outcome is a discrete 'accept'[2] or 'reject' of a null hypothesis with the exact p value unreported. All that is often reported is that the exact p value is either less than or greater than some magical mark. This desire for a crisp answer may be simply an anachronism that will eventually be superseded by the widespread use of exact p values, or it may be part of our innate desire for 'clean' answers—we prefer unambiguity (however apparent) to fuzziness. For instance, even when exact p

[2]Strictly speaking, this should be 'failure to reject'.

values are reported on a test statistic, we often mentally compare these values to the mental yardsticks of 0.05 and 0.01 and make judgements based on whether or not the reported p values exceed these yardsticks.

To summarize our position in this paper, just as there are different types of evidence, so there are different types of inference and even though we might express a preference for one type of inference, we should recognize the validity of other approaches. Not to do so would lead to a very restrictive type of quantitative spatial analysis that would undoubtedly miss many important findings.

Acknowledgement

The first author acknowledges the generous support of Science Foundation Ireland in the form of a Research Professorship.

References

AKAIKE, H., 1973, Information theory and an extension of the maximum likelihood principle. In *2nd Symposium on Information Theory*, edited by B. N. Petrov and F. Csaki (Budapest: Akademiai Kiado), pp. 267–281.

BURNHAM, K. P., and ANDERSON, D. R., 1998, *Model Selection and Inference: A Practical Information-Theoretic Approach* (New York: Springer).

CLEVELAND, W., 1993, *Visualizing Data* (Summit, NJ: Hobart Press).

COOK, D., CRUZ-NEIRA, C., KOHLMEYER, B. D., LECHNER, U., LEWIN, N., NELSON, L., OLSEN, A., PIERSON, L., and SYMANZYK, J., 1998, Exploring environmental data in a highly immersive virtual reality environment. *Environmental Monitoring and Assessment*, **51**, 441–450.

COSTANZO, C. M., 1983, Statistical inference in Geography: modern approaches spell better times ahead. *The Professional Geographer*, **35**, 158–165.

DIACONIS, P., and EFRON, B., 1983, Computer intensive methods in statistics. *Scientific American*, **248**, 116–130.

DORLING, D., 1991, The Visualisation of Spatial Social Structure. Ph.D. Thesis, Department of Geography, University of Newcastle-upon-Tyne.

DORLING, D., and OPENSHAW, S., 1992, Using computer animation to visualise space–time patterns. *Environment and Planning B*, **19**, 639–650.

EFRON, B., and GONG, G., 1983, A leisurely look at the bootstrap, the jackknife and cross-validation. *American Statistician*, **37**, 36–48.

EFRON, B., and TIBSHIRANI, R., 1986, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**, 54–77.

EHRENBERG, A., 1982, *A Primer in Data Reduction* (Chichester, UK: Wiley).

FOTHERINGHAM, A., BRUNSDON, C., and CHARLTON, M., 2000, *Quantitative Geography: Perspectives on Spatial Analysis* (London: Sage).

FOTHERINGHAM, A. S., and ZHAN, F., 1996, A comparison of three exploratory methods for cluster detection in spatial point patterns. *Geographical Analysis*, **28**, 200–218.

GURNEY, K., 1995, *An Introduction to Neural Networks* (London: UCL Press).

HEARNSHAW, H. M., and UNWIN, D. J. (eds.), 1994, *Visualisation in Geographical Information Systems* (New Brunswick, NJ: Wiley).

HERTZ, J., KROGH, A., and PALMER, J. A., 1991, *Introduction to the Theory of Neural Computation* (Reading, MA: Addison-Wesley).

HOPE, A. C. A., 1968, A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society B*, **30**, 582–598.

JACQUEZ, G. M., 1999, Spatial statistics when data are uncertain. *Geographic Information Sciences*, **5**, 77–87.

LONGLEY, P., GOODCHILD, M., MAGUIRE, D., and RHIND, D., 2001, *Geographical Information Systems and Science* (Chichester, UK: Wiley).

MACEACHRAN, A. M., BREWER, C. A., and PICKLE, L. W., 1998, Visualizing georeferenced data: representing reliability of health statistics. *Environment and Planning A*, **30**, 1547–1561.

MANLY, B., 1991, *Randomization and Monte Carlo Methods in Biology* (London: Chapman & Hall).

MOONEY, C. Z., and DUVALL, R. D., 1993, *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage Series in Quantitative Applications in the Social Sciences, 95 (Newbury Park, CA: Sage).

NESTER, M., 1996, An applied statistician's creed. *Applied Statistics*, **45**, 401–410.

TOBLER, W., 1973, A continuous transformation useful for districting. *Annals of the New York Academy of Science*, **219**, 215–220.

TUFTE, E., 1983, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press).

TUKEY, J., 1977, *Exploratory Data Analysis* (Reading, MA: Addison-Wesley).