SPATIAL STATISTICS

Although geographers and other spatial scientists were pioneers in the early development of GIS, the growth in GIS was largely driven by the software industry whose products were geared towards a big business / big government market. The analytical functions available in most proprietary GIS programs, consequently, are not very sophisticated - i.e. they tend to include things like measuring distances within a coverage, or union or intersect operations by overlaying coverages, but do not always include the more advanced spatial statistical techniques developed by spatial scientists (i.e. geographers, spatial statisticians, etc.). Indeed, most proprietary GIS programs are feature-poor with regard even to 'ordinary' non-spatial statistical techniques. However, as GIS evolves from Geographical Information Systems into Geographical Information Science (and as software companies need to find new niche markets), this is a deficiency which will probably be remedied in the future.

Today I want to look at some of the techniques which are often absent from GIS. We will begin by reviewing a few basic concepts. Following that we will take a fairly detailed look at the principles underlying statistical modelling. This will be followed by a brief review of some of the major techniques, based on a classification derived from the concepts outlined in the first section. Finally we will take a brief look at what software is currently available for these techniques.

BASIC CONCEPTS

A distinction may be made between spatial statistics and statistics in general. The most obvious difference is that spatial statistics are used to analyse data which have a spatial location. Spatial statistics give explicit consideration to spatial properties such as location, spatial patterns, spatial arrangement, distance, etc. This spatial dimension tends to make spatial statistics more complex than ‘ordinary’ non-spatial statistics. However, the book by Bailey and Gatrell (1995) (see references) provides a reasonably accessible introduction to the subject. 1

Bailey and Gatrell divide spatial statistical techniques into four categories depending upon the type of data for which they are designed, viz.

- Point pattern data;
- Spatially continuous data;
- Areal data; and
- Interaction data.

There is a fairly close correspondence between these categories and the basic conceptualisations of space introduced in a previous class: point data use an entity (or object) based conceptualisation, spatially continuous data are field based, area data are entity based (although the associated attributes - e.g. population density - may in some instances also be conceptualised as fields); and interaction data are entity based (or what I identified as network based). These conceptualisations in turn tend to influence the data model used in GIS (i.e. raster or vector).

Entities generally have associated attribute data. These data may be measured on various measurement scales: nominal, ordinal, or interval. The numeric values in a nominal scale simply identify different categories, without any implications of size or importance (e.g. 1=females, 2=males). These numeric values could be replaced by simple labels without any loss of information. The numeric values in an ordinal scale also identify categories, but the numbers imply an ordering in the categories (e.g. 1=very good, 2=quite good, 3=medium, 4=poor). However, the numbers cannot be added, subtracted, multiplied or divided in any meaningful way. The numeric values in an interval scale measure quantity (e.g. temperatures in degrees Centigrade). This means that the difference (i.e. interval) between two values is a measure of quantity. 2

---

1 Most geographical statistics texts provide a simplified version of ‘ordinary’ statistics, with only passing mention to some spatial statistical techniques. Very few books provide an explicit consideration of spatial statistics, and most of those that do are written by statisticians with little allowance being made for ordinary mortals such as geographers.

2 The traditional classification included a fourth category called ‘ratio’, but this may be regarded as a special case of ‘interval’. The distinguishing property of a ratio scale is that a value of zero indicates a quantity of none. This means that the ratio of other values (e.g. 100 and 50) indicates the ratio of their relative sizes (i.e. twice). Other
Measurement scales can be viewed as 'low' (i.e. nominal) or 'high' (i.e. interval). The measurement scale determines whether a particular statistical technique may or may not be used for a particular purpose. For example, Pearson correlation requires data on an interval scale, but a Chi Square test can be used with nominal data or higher (i.e. ordinal or interval).

Bailey and Gatrell also provide a useful classification of statistical techniques depending upon the purpose for which they are used. This classification has three categories:

- Visualisation. This simply refers to plots, graphs, maps and various other ways of graphical depicting the data.
- Exploration. The objective here is to describe the data. This may involve the calculation of descriptive statistics (e.g. mean, standard deviation) or it may involve a form of visualisation after further processing of the data.
- Modelling. The objective here is to examine relationships or quantify descriptive parameters in a more formal manner. Modelling is involved in all forms of statistical inference or hypothesis testing, although this fact may be disguised in the simpler ‘cookbook’ statistics texts. Statistical modelling is therefore explored in more detail in the following section.

These three categories are by no means discrete and tend to merge into one another. Nevertheless, the classification is didactically useful.

**MODELLING SPATIAL DATA**

Statistical models of one description or another are implicit in all formal statistical inference and hypothesis testing, even though the term 'model' may not be used explicitly in many elementary statistical texts. If you have experience with some basic statistical techniques such as ‘t’ tests or regression, then you have ‘done’ statistical modelling. We need to introduce the fundamental ideas and principles more explicitly.

**(a) Basic Concepts**

Since statistical models are concerned with phenomena which are **stochastic**, that is to say phenomena which are subject to uncertainty, or governed by the laws of probability, we need a 'language' which allows us to represent such uncertainty mathematically. This is provided by the concept of a **random variable** and its associated **probability distribution**.

Random variables are used to represent stochastic phenomena mathematically; informally, we may think of them as simply being variables whose values are subject to uncertainty. Typically, we might represent the result of a single throw of a die, or the measurement of precipitation at a particular site, by a random variable say, \( Y \). The name is somewhat unfortunate, since ‘random’ tends to imply that such variables are equally likely to take any of their possible values, whereas all that is really meant is that the values are subject to some form of uncertainty. One value, or maybe a set of values, may be more likely than others, but we still refer to the variable concerned as 'random'.

The relative chance of a random variable taking different possible values is characterised by its associated probability distribution, so that we might refer to the random variable \( Y \) as having a probability distribution \( f_Y(y) \). \( f_Y(y) \) is a mathematical function which specifies the probability that \( Y \) has the specific value (or ranges of values) \( y \). The random variables are represented by upper case letters, and specific values of a random variable by lower case letters. For example, \( Y \) may be a random variable that conceptually represents the result of throwing a die, but when it is actually thrown we get a particular observed value \( y \) of this random variable.

In general, random variables may be **discrete**, that is, only able to take a finite number of values; in this case \( f_Y(y) \) is the **probability** that \( Y \) takes the specific value \( y \). They may also be **continuous**, able to take any value within a continuous range, in which case \( f_Y(y) \) is the **probability density** at the value \( y \). The probability that a random variable \( Y \) takes values in some range \((a, b)\) is therefore:

---

classifications include further categories (e.g. ‘binary’ as a special instance of nominal, or ‘percentage’ as a special instance of ratio), but these may be ignored for present purposes.

---

3 This section is a slightly edited version of the discussion by Bailey and Gatrell (1995).
We may also occasionally be interested in the cumulative probability distribution $F_Y(y)$ of a random variable $Y$, sometimes referred to as its distribution function. This is simply a mathematical function which specifies the probability that $Y$ takes any value less than or equal to the specific value $y$. Thus:

$$F_Y(y) = \begin{cases} \sum_{u=-\infty}^{y} f_Y(u) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{y} f_Y(u) \, du & \text{if } Y \text{ is continuous} \end{cases}$$

We may also be interested in the expected value or mean, of a random variable $Y$, or perhaps some function, $g(Y)$, of that random variable. As its name suggests, this is simply the 'average' value that we would expect $Y$ or $g(Y)$ to take. It is therefore a weighted sum of the possible values, where the weights used for each value are the probability associated with that value. Therefore:

$$E(Y) = \begin{cases} \sum_{y=-\infty}^{\infty} y \cdot f_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} y \cdot f_Y(y) \, dy & \text{if } Y \text{ is continuous} \end{cases}$$

or

$$E(g(Y)) = \begin{cases} \sum_{y=-\infty}^{\infty} g(y) \cdot f_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} g(y) \cdot f_Y(y) \, dy & \text{if } Y \text{ is continuous} \end{cases}$$

The expected value of one particular function of a random variable is often of particular interest, namely the expected squared deviation of the random variable from its mean. This is the variance of the random variable and it is a broad measure of how much its values tend to vary around their 'average'. Formally, $VAR(Y) = \Sigma [(Y - E(Y))^2]$. The positive square root of this quantity is referred to as the standard deviation ($\sigma_Y$) of the random variable.

All of these ideas generalise to the case where we are interested not just in one random variable but in more than one. If we have two random variables $(X, Y)$, then we can speak of their joint probability distribution, $f_{XY}(x,y)$, which specifies the probability, or probability density, associated with $X$ taking the specific value $x$, at the same time that $Y$ takes the specific value $y$. As well as the mean and variance of $Y$ or $X$, we may then also be interested in the expected tendency for values of $X$ to be 'similar' to values of $Y$. A broad measure of this is the covariance of the two random values, defined as $COV(X,Y) = \Sigma((X - E(X))(Y - E(Y)))$. The covariance of two random variables divided by the product of their standard deviations is referred to as the correlation between them, i.e.

$$\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y},$$

Two random variables are said to be independent if the probabilistic behaviour of either one remains the same, no matter what values the other might take. In that case their joint probability distribution is simply the product of their individual probability distributions, so that $f_{XY}(x,y)=f_X(x)f_Y(y)$. If this is the case, then their covariance and correlation will each be zero.
(b) Statistical Models.

A statistical model for a stochastic phenomenon consists of specifying a probability distribution for the random variable (or variables) that represent the phenomenon. Once this probability distribution is fully specified there is effectively nothing further that can be said concerning the behaviour of the phenomenon - by definition it cannot be specified to any greater degree of precision than the probability that particular values may occur.

In the case of modelling something simple, like throwing a die, which involves a single random variable, say Y, this means specifying a corresponding probability distribution, \( f_Y(y) \). However, for more complex phenomena rather more can be involved. Consider, for example modelling levels of a photochemical oxidant such as ozone in a large rural region. The ozone level at each location, \( s \), in the region \( R \), will vary during the course of a day and from day to day according to some probability distribution.

Note that \( s \) is a \( (2 \times 1) \) vector \( (s_1, s_2)^T \) which refers to a point location, where \( T \) indicates the matrix should be transposed – i.e. it is a column vector. This is simply a shorthand way of referring to the x-coordinate \( s_1 \) and the y-coordinate \( s_2 \) of a point. The bold typeface indicates a vector. If we have two point locations we can refer to them by the two vectors \( s_1 \) and \( s_2 \) where \( s_1 = (s_{11}, s_{12})^T \) and \( s_2 = (s_{21}, s_{22})^T \).

The particular form of the ozone level probability distribution may well differ from location to location in \( R \). Furthermore the ozone levels at neighbouring locations may well be related in some way. For example, ozone levels at sites separated by 5 kilometres are probably quite similar, while those separated by 50 kilometres may be very different. Thus, to represent ozone levels in \( R \) we require a set of possibly non-independent random variables, \( \{Y(s), s \in R\} \). Such a set is often referred to as a spatial stochastic process. A complete statistical model for the ozone level in region \( R \) involves specifying the joint probability distribution of every possible subset of these random variables. Without simplifying assumptions, this would be a formidable task.

So, how do we go about formulating a statistical model? In rare cases, like that of a throw of a die, the appropriate statistical model may be apparent purely from our theoretical knowledge about the phenomenon. For example, an obvious model for a fair die is: \( f_Y(y) = 1/6 \), \( y = 1 \ldots 6 \). However, when we attempt to model phenomena more interesting than dice, like ozone levels, we are unlikely to know enough about the particular phenomenon to be able to fully specify a model in such a way. We must therefore rely on observational data to help us to arrive at an appropriate model. However, it should be appreciated that whilst data will help, alone they can rarely provide a definitive model. A typical data set in the ozone level example would consist of a set of observations \( y(1), y(2), \ldots \) each at specific sites, \( (s_1, s_2, \ldots) \) in \( R \).4 This data set, \( \{y_1, y_2, \ldots\} \), is often referred to as a realisation of the spatial process. It is just one observation from the joint probability distribution of the random variables \( \{Y(s_1), Y(s_2), \ldots\} \) - or in simpler notation \( \{Y_1, Y_2, \ldots\} \). Unfortunately, one observation does not give much information about a joint probability distribution, even if we were prepared to accept that this particular set of sites was 'typical' (i.e. an unbiased sample) of sites in general in \( R \), which may itself be open to question.

The specification of a model generally involves using a combination of both data and 'reasonable' assumptions about the nature of phenomena. Such assumptions may arise, for example, from background theoretical knowledge about how one expects a phenomenon to behave, the results of previous analyses on the same, or a similar, phenomenon, or alternatively, from the judgment and intuition of the modeller. How ‘reasonable’ these assumptions are, in certain cases, can be assessed by exploratory analyses of aspects of the observed data appropriate to those particular assumptions. Once specified, they provide a basic 'framework' for a model. Usually this framework will amount to the specification of a general mathematical form for the probability distribution appropriate to the phenomenon, but one which involves certain parameters whose values are left unspecified. This general form is then further refined, or fitted (i.e. the values of the unknown parameters are estimated), by reference back to the observed data. The fitted model can then be evaluated, which may lead to modified assumptions and a different model being fitted, and so on.

An illustration of this approach would be to apply a simple form of the familiar standard linear regression model, widely used in non-spatial analysis, to our ozone level example. An example of a model of this type would involve: firstly, the assumption that the random variables \( \{Y(s), s \in R\} \) are independent; secondly, that their probability distributions only differ in their mean value, all other aspects being the same; thirdly, that this mean value is a simple

---

4 Strictly speaking the observations \( y_1, y_2, \ldots \) should be written as \( y(s_1), y(s_2), \ldots \) as they are observations on the random variables \( Y(s_1), Y(s_2), \ldots \). However, the references to \( s \) may be dropped for convenience. \( Y(s_1), Y(s_2), \ldots \) may likewise be referred to as \( Y_1, Y_2, \ldots \).
linear function of location, say $E(Y(s)) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$ where $(s_1, s_2)$ are the spatial coordinates of the location $s$; and fourthly that each $Y(s)$ has a normal distribution about this mean with the same constant variance $\sigma^2$. In short, our model says that $Y(s)$ are independent and the probability distribution of each is $N(\beta_0 + \beta_1 s_1 + \beta_2 s_2; \sigma^2)$ – i.e. normal with mean $\beta_0 + \beta_1 s_1 + \beta_2 s_2$ and variance $\sigma^2$.

Under such assumptions, we are relieved of the task of specifying a joint probability distribution for every subset of $Y(s)$, since independence implies this would just be the product of the distributions at each of the sites involved in any such subset. Furthermore, although each of these distributions differ, they do so only through a simple relationship involving a small number of parameters, $\beta_0$, $\beta_1$ and $\beta_2$. Hence, our observed data $(y_1, y_2, ...)$ at specific sites, $(s_1, s_2, ...)$ in $R$, cease to be just rather unhelpful single observations from the joint probability distribution of the random variables $(Y_1, Y_2, ...)$, and become usable to estimate the unknown parameters of our postulated model - there is now the necessary 'replication' in observations to achieve this. Parameter estimation is discussed in the next subsection; the point here is simply that our assumptions provide a framework under which final model specification reduces to a problem of the estimation of unknown parameters. Furthermore, enough replication is present in the data to make this estimation feasible.

(c) Parameter Estimation

Given such a model framework exactly how are unknown parameters to be estimated from the data? In other words, how do we fit a proposed statistical model? The single most commonly used approach is that of maximum likelihood.

The idea of maximum likelihood is really quite simple. Our exploratory analyses, background knowledge, intuition and judgment have led us to a model framework, which, as discussed previously, will usually amount to the specification of a general mathematical form for the probability distribution appropriate to the phenomenon under study, but one which involves certain parameters whose values are left unspecified. In general let us denote these unknown parameters by the vector $\theta$. In the above example of the linear regression model, $\theta$ would have four elements ($\beta_0$, $\beta_1$, $\beta_2$ and $\sigma^2$), but in general it might have any number of elements. Now, if the model framework provides a general mathematical form for the probability distribution appropriate to the phenomenon under study, then from this we must be able to write down a general mathematical form for the joint probability distribution of the set of random variables $\{Y_1, Y_2, ... Y_n\}$ for which our data $(y_1, y_2, ... y_n)$ constitute a set of observed values. Suppose that this joint probability distribution is $f(y_1, y_2, ... y_n; \theta)$. Here we have explicitly indicated that as well as being a function of the observed values $(y_1, y_2, ... y_n)$, this will also in general depend on the values of the unknown parameters $\theta$.

How is $f(y_1, y_2, ... y_n; \theta)$ to be interpreted? It is a joint probability distribution so, given particular values for $\theta$, it specifies the probability, or probability density, associated with $Y_i$ taking the specific value $y_i$ at the same time that $Y_j$ takes the specific value $y_j$ and so on. Hence, if $(y_1, y_2, ... y_n)$ are our actual data observations, $f(y_1, y_2, ... y_n; \theta)$ is effectively the probability or probability density associated with them occurring, given the general model framework proposed. We refer to this as the likelihood for the data, and we would normally denote it as $L(y_1, y_2, ... y_n; \theta)$. Notice that given a known set of observed data $(y_1, y_2, ... y_n)$, the likelihood depends only on the unknown parameter values $\theta$. Recall that our objective is to estimate values for $\theta$ and an obvious way of doing this is now apparent. We should choose values for $\theta$ that maximise the likelihood for the observed data. Often, because it is easier and equivalent, we maximise the logarithm of the likelihood or the log likelihood, which we denote $l(y_1, y_2, ... y_n; \theta)$. Under the maximisation of either function, we are effectively 'tuning' our proposed model framework by choosing parameter values which give the greatest possible likelihood of observing the data that we actually observed. This would seem a sensible way to proceed.

This, then, is the general approach to parameter estimation or model fitting, known as maximum likelihood. Of course, in any specific case the maximisation of $L(y_1, y_2, ... y_n; \theta)$ or its logarithm, with respect to $\theta$, may not be particularly easy and may involve intensive computation and the development of algorithms to numerically approximate the solution. However, for some 'standard' and commonly used model frameworks it turns out to be relatively straightforward and mathematically equivalent to ways of fitting models with which you may be more familiar. For example, for the multiple linear regression model discussed earlier, where the model framework involves the assumption of independently distributed random variables, having normal distributions with the same variance, maximum likelihood parameter estimation reduces to using the method of ordinary least squares. That is, parameters are estimated by minimising the sum of squared 'residuals' or differences between the data values and
those predicted under the model. If we relax the assumption of independence and equal variance, but maintain the
other aspects of the model, then maximum likelihood leads to parameter estimation by **generalised least squares**, which is the minimization of a weighted sum of squared 'residuals'.

Maximum likelihood not only provides parameter estimates, but also general measures of how reliable these estimates are (i.e. their **standard errors**) and of how well alternative models fit a particular set of data. Standard errors are, in essence, derived by consideration of how 'peaked' the likelihood function is at its maximum. Informally, the idea is again simple. If the likelihood is highly 'peaked' at its maximum and falls sharply in value as you move away from this, then you can be pretty sure that the estimated parameter values are reasonably reliable (small standard errors). If, alternatively, the maximum occurs at some point on a slowly changing 'plateau', then the value of the likelihood is similar all over this 'plateau' and therefore for several different sets of parameter values; hence you should be less confident in the estimates derived (large standard errors). When it comes to comparing the overall fit of two different models we can consider the ratio of the likelihoods associated with each (duly maximised for the parameters involved). Informally, we ask whether one value is significantly better than the other.

**(d) Hypothesis Testing**

So where does statistical hypothesis testing fit into our picture of statistical modelling? Testing a hypothesis is a question of comparing the fit to the data of two models, one of which incorporates assumptions which reflect the hypothesis, the other incorporating a less specific set of assumptions. Usually a hypothesis will amount to the specification of values for certain of the parameters involved in the model. Testing hypotheses is therefore one facet of statistical modelling – i.e. we simply ask whether a model, in which certain parameters have pre-specified hypothesised values, fits the data significantly less well than one where these parameters are allowed to be optimally estimated from the data. No new theory is really involved, although of course we can wrap all this up in the language of 'p-values' and so on, if we so wish. Notice, however, that all modelling inevitably involves some assumptions about the phenomenon under study; hence hypothesis testing will always involve comparison of the fit of a hypothesised model with that of an alternative which also incorporates assumptions, albeit of a more general nature. The validity of a particular form of hypothesis test often relies critically on these alternative assumptions being, in turn, valid for the phenomenon in question. For example, in the ozone model discussed above, the standard multiple regression test of whether the parameters $\beta_1$ and $\beta_2$ are significantly different from zero depends on both the assumption of the independence of $Y(s)$ and of a normal distribution for these random variables.

**(e) Spatial Data Modelling**

Thus far, our very brief review of the general process of statistical modelling and hypothesis testing has not been oriented specifically towards models of spatial phenomena. However, having established sufficient general background we can now introduce some concepts of particular importance with regard to spatial data.

As already emphasized earlier, spatial phenomena often exhibit a degree of **spatial correlation**. Spatial analysts need to incorporate the possibility of such spatial dependence into their models if the models are to provide realistic representations of such phenomena. For example, the independence assumption of the standard multiple regression model would be unlikely to be realistic in relation to our ozone level example. In addition to the mean value of the ozone level varying in $R$, the distribution of values about this mean at any site is likely to be related to that in neighbouring sites. In fact, a simple linear relation between mean value and location is also unlikely! But the point to be emphasised here is that in general the behaviour of spatial phenomena is often the result of a mixture of both first order and second order effects. **First order effects** relate to variation in the mean value of the process in space - a global or large scale trend. **Second order effects** result from the spatial correlation structure, or the spatial dependence in the process; in other words, the tendency for deviations in values of the process from its mean to 'follow' each other in neighbouring sites - local or small scale effects.

A somewhat artificial illustration may illustrate these ideas better. Suppose we imagine scattering, entirely at random, iron filings on to a sheet of paper marked with a fine regular grid. The numbers of iron filings landing in different grid squares can be thought of as the realisation of a spatial stochastic process. As long as the mechanism by which we scatter the filings is purely random, there should be an absence of both first and second order effects in the process - different numbers of filings will occur in each square, but these differences arise purely by chance.

Now suppose that a small number of weak magnets are placed under the paper at different points and we scatter the filings again. The result will be a process with spatial pattern arising from a first order effect - clustering in the numbers in grid squares will occur globally at and around the sites of the magnets. Now remove the magnets, weakly magnetise the iron filings instead, and scatter them again. The result is a process with a spatial pattern arising from a
second order effect - some degree of local clustering will occur because of the tendency for filings to attract each other. If the magnets are now replaced under the paper and the magnetised filings scattered again, we end up with a spatial pattern arising from both first and second order effects.

Because 'real life' spatial patterns frequently arise from this sort of mixture of both first and second order effects, independence in the random variables representing a spatial stochastic process is often too strong an assumption for the spatial modeller. By definition, such an assumption rules out second order effects and therefore needs to be replaced by some weaker alternative which allows for the possibility of a covariance structure. A common approach is to think of the variable of interest (such as the ozone level at a location) as comprising two components. The first order component represents large scale spatial variation in mean value. This is similar to the dependence proposed in the simple regression model used earlier, although the relationship between the mean and location need not be linear, whilst 'covariates' might be included in the relationship, instead of, or together with, location. The second order component is concerned with the behaviour of stochastic deviations from this mean. Instead of assuming these to be spatially independent, they are allowed to have a covariance structure which may give rise to local effects.

The second order component is often modelled as a stationary spatial process. Informally, a spatial process \( \{Y(s)\} \) is stationary or homogeneous if its statistical properties are independent of absolute location in \( R \). In particular, this would imply that the mean, \( E(Y(s)) \), and variance, \( VAR(Y(s)) \), are constant in \( R \) and therefore do not depend upon location \( s \). Stationarity also implies that the covariance, \( COV(Y(s_i), Y(s_j)) \), between values at any two sites, \( s_i \) and \( s_j \), depends only on the relative locations of these sites, the distance and direction between them, and not on their absolute location in \( R \). We say further that the spatial process is isotropic if, in addition to stationarity, the covariance depends only on the distance between \( s_i \) and \( s_j \) and not on the direction in which they are separated.

If the mean, or variance, or the covariance structure 'drifts' over \( R \) then we say that the process exhibits non-stationarity or heterogeneity.

In terms of our earlier example, the case where weakly magnetised iron filings are scattered onto paper with no magnets underneath would roughly equate to an isotropic (stationary) process. The case with unmagnetised filings and magnets underneath the paper, approximates a process with heterogeneity in the mean value and independence in deviations from mean value - a simple form of non-stationary model, similar in spirit to our simple regression model, although obviously not in respect of the relationship between mean and location, which would be non-linear in this case. The experiment with both magnetised filings and magnets under the paper is a more complex non-stationary process that mixes the previous two cases and might be considered akin to a 'two component' model often useful in practice for spatial processes. It involves first order variation or heterogeneity in mean value, combined with a stationary second order effect.

Heterogeneity in the mean, combined with stationarity in second order effects, is a useful spatial modelling assumption, where it may be regarded as 'reasonable' and acceptable, since it implies that the covariance of the process has the same structure from area to area within the region studied. Without the assumption of some form of stationarity in spatial models, one begins to have great difficulty in fitting them to observed data, since the number of parameters involved becomes unmanageable. If all locations in space have potentially different covariance structures as well as means, and, as is usual, we have only a set of single observations at a particular subset of locations, then we stand little chance of estimating all parameters involved in the model.

So, in broad terms, the modelling of a spatial process often tends to proceed by first identifying any heterogeneous 'trend' in mean value and then modelling the 'residuals', or deviations from this 'trend', as a stationary process.

However, it must be acknowledged that concepts such as stationarity and first or second order effects are artefacts of the modeller and not reality. In practice, effects are confounded in observed data and the distinction between them is difficult and ultimately to some extent arbitrary. Both types of effect can give rise to similar spatial patterns. If high values of a process are found in one region and low values among a set of adjacent sites in another region, then how do we know whether the underlying process is non-stationary (i.e. heterogeneous) or if these are local effects resulting from a homogeneous spatial dependence in the data? In other words, how can we distinguish spatial dependence in a homogeneous environment from spatial independence in what is a heterogeneous environment?

These are difficult questions with no definitive answers. Analytical methods can help to identify appropriate models and distinguish in some cases between effects that must clearly be global first order trends and those that are more likely to be the result of a second order covariance structure. However, the extent to which they can assist the modeller is limited. Ultimately, models are mathematical abstractions of reality and not reality itself. Statistically we may not be able to distinguish between two equally good but structurally different explanations for variations in a
phenomenon of interest. We have already pointed out that the specification of a model will always involve using a combination of both data and assumptions about the nature of phenomena being modelled. Ultimately, judgement and intuition on the part of the analyst are always involved in statistical modelling. Statistical models are always at best 'not wrong', rather than 'right'.

GEOGRAPHICALLY WEIGHTED REGRESSION

The ozone model discussed above models the relationship between ozone levels and location in space - i.e. it does not consider covariates (i.e. other 'predictor' variables which might impact upon ozone levels, e.g. traffic levels). Covariates are often incorporated in a multiple regression model taking the general form:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \epsilon_i$$

where there are $k$ predictors, $x_{ik}$ is the value of variable $k$ at location $i$, and $\beta_k$ is a parameter indicating the relationship between the dependent variable $Y$ and variable $X_k$. This model presupposes that the regression coefficients are homogeneous or stationary - i.e. that the relationship between the dependent variable and each independent (or predictor) variable is the same everywhere. However, this assumption may not be realistic. For example, house prices might be expected to be influenced by the age of the house. However, in some areas an old house may be valued less than a new house, whereas in other areas an old house may be valued more. In such situations, the relationship between house prices and age is non-stationary.

To accommodate these situations, Fotheringham et al. have proposed a non-stationary model:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \epsilon_i$$

In other words, the values of the regression coefficients $\beta$ are a function of location (where $u_i$ and $v_i$ represent the x and y co-ordinates). One problem with this model is that there are too many unknown parameters, so it is necessary to introduce simplifying assumptions in order to calibrate the model (i.e. estimate the parameters). Geographically weighted regression (GWR) assumes that the parameters are non-stationary but are functions of location.

Each of the regression coefficients estimated by GWR is a function of location. This means that they can be mapped. If the relationship between the dependent variable and variable $k$ is stationary (as assumed by the classical regression model), the map would display no spatial variation, but if it is non-stationary then the map will indicate how the relationship varies over space.

The interpretation of a non-stationary relationship raises interpretation problems. If the relationship between the dependent and an independent variable varies over space, one obvious question is: why should it vary? Post-modernists are quite happy to accept that places are different and therefore so are the relationships between phenomena in different places. Positivists are more likely to seek explanations in terms of other variables which have not been included in the model. If these missing variables could be quantified and included in the model, then the non-stationarity would disappear.

TECHNIQUES

Bailey and Gatrell review a number of spatial statistical techniques which are developed from these principles. This review provides an inventory of some of the techniques which may increasingly find their way into GIS over the next few years. The following simply lists some of the techniques discussed, classified by the type of data they are designed for.

Point Patterns

Point pattern statistics are used to analyse the spatial distribution of features which can be modelled as discrete points.

- Quadrat analysis
• Kernel estimation
• Nearest neighbour analysis
• K-functions

Statistics calculated using these methods are often used to test hypotheses of Complete Spatial Randomness (CSR) – i.e. a homogeneous Poison process throughout the study region. However, other models (e.g. a heterogeneous Poison process) can be explored.

Other methods have been developed to test for interactions between multiple event types, space-time clustering, variations in the underlying population at risk, clustering around specific point sources, etc.

**Spatially Continuous Data**

Spatially continuous data is sometimes termed random field data, or more simply geostatistical data. It should be noted that the term geostatistics refers to the subset of spatial statistics used to analyse spatially continuous data. Spatial statistics is a much broader concept.

Methods developed for continuous data include:

• Spatial moving averages
• Trend surface analysis
• Delauney triangulation / Thiesen polygons / TINs
• Kernel estimation (for the values at sample points)
• Variograms / covariograms / kriging
• Principal components analysis / factor analysis
• Procrustes analysis
• Cluster analysis
• Canonical correlation

**Area Data**

We frequently need to analyse attribute data which refer to polygons (i.e. areas). Methods developed for this purpose include:

• Spatial moving averages
• Kernel estimation
• Spatial autocorrelation (Moran’s I, Geary’s c)
• Spatial correlation and regression

Various models may be developed for spatial correlation and regression (e.g. to relax the assumption of second order stationarity).

Methods have also been developed for special types of area data (e.g. counts, proportions). In the context of regression, many of these involve applications of the generalised linear model (e.g. Poison regression, logistic regression).

When dealing with rates based on small numbers there is a risk of extreme values in areas with a low population (the small numbers problem). One response to this is to map Poison probabilities. This, however, tends to place the emphasis upon the larger areas. Bayesian methods appear to provide a reasonable balance between these two tendencies.

**Spatial Interactions**

Most methods developed for interactions are based upon the gravity model. This postulates that the amount of interaction between two places is a function of their sizes (measured using an appropriate metric) and is inversely related to the distance between them. Distances may be measured as straight line distances or through a network. Apart from measuring interaction, the gravity model can be used for various other purposes e.g. to assign places to alternative centres, identify locations to maximise accessibility, etc.
SOFTWARE FOR SPATIAL STATISTICS

ArcGIS

ArcView 3.2, like most commercial GIS application programs, contained very few options for spatial statistics. ArcGIS is a little better. The Geostatistical Analyst extension, introduced in ArcGIS 8.1, provides some tools but it is still very limited. Although it provides tools for kriging and other simpler forms of spatial interpolation (e.g. inverse distance weighting), many other geostatistical options are missing. Likewise, many of the other non-field spatial statistical techniques mentioned above are not currently available elsewhere in ArcGIS.

Idrisi

Given its origins in an academic Geography Department, Idrisi not surprisingly provides a wider range of spatial statistical options, although it falls short in some areas (mainly because of its preference for raster / field data). Modules in the GIS Analysis | Statistics menu include:

- Pattern – Calculates various descriptive statistics used in landscape ecology for a moving pixel window.
- Regres – Simple regression analysis for images or values files.
- Multireg – Multiple regression for images or values files.
- Logisticreg – Binomial logistical regression for images or values files (multiple independent variables).
- Trend – First, second and third order trend surfaces.
- Autocorr – Moran’s I measure of first-lag autocorrelation in a raster image.
- Quadrat – Quadrat analysis for counts of point features saved as a raster image.
- Center - Weighted or unweighted mean centre and standard radius of point data saved as a raster.
- Cratio – Compactness ratio (compares area of a polygon to a circle with the same perimeter).
- Crosstab – Crosstabulation of images with various statistics (Cramer’s V, Chi-Square, Kappa).
- Validate – Compares similarity of two multicategory maps using Kappa statistic.
- ROC – Relative Operating Characteristic. Compares predicted likelihood of a class occurrence with Boolean image of actual occurrence.
- Sample – Produces vector file of random, systematic or stratified random sample points.
- Standard – Converts values in an image into standardised normal deviates (z-scores).

S-Plus

Some of the deficiencies can be remedied using third party stand alone programs or add-ons. Perhaps the best examples are provided by a company called Insightful (formerly MathSoft). Insightful markets an advanced statistical software package called S-Plus. S-Plus includes a user-friendly GUI to the more commonly used statistical techniques (similar to SPSS – Statistical Package for the Social Sciences), including a very comprehensive suite of graphics routines. It also incorporates its own object-orientated programming language, similar to C++, which enables statisticians to develop their own statistical models. Of more relevance to the present discussion, Insightful produces an add-on called S+SpatialStats which adds a number of important spatial statistical functions to S-Plus. They also market an ArcView extension which allows you to access the full power of S-Plus, including the graphics routines and S+SpatialStats from within ArcView.

S+SpatialStats includes tools for the following:

- Nearest neighbour tests;
- Kernel estimation (plus other density estimation methods);
- Ripley’s k-function tests of second order stationarity;
- Variogram fitting;
- Ordinary and Universal Kriging;
- Moran test for spatial autocorrelation;
- Geary test for spatial autocorrelation;
- Spatial regression using covariance structures.

R

S-Plus is produced by a commercial company (Insightful), but there is also an Open Source equivalent called R. The Comprehensive R Archive Network website (http://cran.at.r-project.org/) provides free downloads of the base system
plus contributed packages for Windows, Macs and Linux platforms. There are several packages and projects dealing specifically with spatial data (e.g. sp, spatstat, DCluster, spgwr).

**BUGS**

Other useful software can be downloaded from the World Wide Web. A good example is a program called BUGS (Bayesian analysis Using Gibbs Simulation). A Windows version (WinBUGS) can be downloaded for free from the web. This is a general purpose program for Bayesian statistics, but it includes a spatial sub-set called GeoBUGS.

**WEB SOURCES**

The module website contains links to R and BUGS and other Open Source software sites for spatial statistics.

**READINGS**
