PREAMBLE

I had rather be hissed for a good verse than applauded for a bad one.

Victor Hugo

Some scholars practicing what is called pure science are convinced that their ways of doing science are theoretical, and hence superior to that done in what is called applied science. On the other hand, many scholars in the applied sciences stress that superiority of theory over practice is a myth, and that theory and practice cannot be separated. They continually point out the numerous possibilities of doing science that mix pure theoretical research goals and applied research goals, each worthy of equal respect and dignity. The relationship between statisticians and geographers in the realm of spatial statistics is a point in question here. The purpose of this paper is for Martin to give his personal view of the application of spatial statistical analysis in geographic research, mostly noting shortcoming of its use by selected geographers. Martin argues that if geographers believe they have theoretical research findings that contribute to statistics, then statisticians should be allowed to scrutinize this research. Throughout his discussion he hints that geographers do not have the expertise necessary for making such contributions, and that geographers should restrict themselves to applications while enticing statisticians into undertaking the requisite theoretical developments. A number of publications concerning spatial statistics have made clear that it is both an oversimplification and even an error to view geography as solely application–oriented, and statistics as theory–oriented, for scholars in both areas hold a variety of talents and viewpoints. Richardson softens Martin's message, noting that all scholars have an interest in avoiding abusive uses of statistics, and echoes Martin's belief that theoretical developments in spatial statistics need to be linked to relevant examples and realistic geographical problems. In many ways, this paper effectively highlights the contrasts between statistical and geographical approaches to spatial statistics.

The Editor

The Role of Spatial Statistical Processes in Geographic Modelling

R. J. Martin

Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, England

Overview: In this paper I give a personal view of the role of spatial statistical processes in geographic modelling. I consider models used by geographers, and comment on the statistical shortcomings of their use. I discuss the role of the geographer in statistical research, and the role of the statistician in geographic research. I also expand on the discussion of two particular topics of interest to geographers—boundary effects and missing values.

1. Introduction

There has been a considerable amount of published research in geography in which spatial statistical models have been used or investigated—see for example the review papers of Cliff and Ord (1975) and Bennett and Haining (1985), and the references therein. I will discuss in this paper one particular part of this research—that part in which spatial stochastic processes are used to model the dependence between observations on the same variable at different spatial sites (or on different regions). This is the topic in Section 6 of Cliff and Ord (1975), and Sections 3.1 and 4.1 of Bennett and Haining (1985). Even in this restricted area I am only going to discuss aspects of which I have some statistical knowledge. My viewpoint is that of a theoretical statistician, and I claim no geographical knowledge or understanding.

In the discussion to Bennett and Haining (1985), I expressed my reservations (Martin, 1985) about the published research in geography that I had seen. Some stronger views were given by Besag (1985). My reservations concerned two main aspects. Firstly, that the models used by geographers did not appear to receive the validation from data that has become standard in current statistical analyses, and there was no indication that geographers felt that such validation was necessary. Secondly, that research by geographers that purported to advance statistical theory and methodology was being published in non-statistical journals, and was clearly receiving inadequate refereeing and not receiving the attention of statisticians.

In Section 2 I shall discuss in detail the role of statistical models in geography, whilst in Section 3 I shall discuss the role of the geographer in statistical research and of the statistician in geographical research.

Two particular topics that have received much attention in geographical publications on spatial statistics are boundary effects and missing values. I discussed boundary effects in Martin (1987), and will reconsider some of that discussion in Section 4. In response to a question from a geographer, I wrote up some research of mine on missing values in Martin (1984). Some further comments are in Martin (1987). A subsequent publication (Haining, Griffith and Bennett, 1989) has considered numerically one aspect of this—the information loss. As a result I derived some theoretical results covering this aspect, which are in Martin (1989a). I discuss and extend some of these results in Section 5.

2. Statistical Models in Geography

2.1. Justification of models

I am not a geographer, and I have no basis for discussing geographic models unless these models are statistical. Unfortunately, I have been unable to understand those statistical models that I have seen used in geography. I mentioned in the introduction my concerns about these models expressed in Martin (1985). In their reply, the authors (Bennett and Haining, 1985) confirmed that the 'model is paramount', and justified this by stating that 'It must be remembered that in human geography and planning we are dealing with individuals and social groups and this results in a problem of legitimizing models, and often in planning applications, it requires the participation of individuals who often will be non–numerate.', and that data analytic methods are 'not appropriate for planning a city'.

I will reply to this in two ways. Firstly, if the data are of no relevance to the model, I do not see the point of presenting the models to statisticians and hoping that 'a research agenda ... may have stimulated the Fellows of the Society to help in their solution'. However, whatever the context and whether or not individuals are involved, I would still be concerned that models are not validated through an examination of relevant data. I also cannot see the relevance of the possible non–numeracy of the participating individuals. My concern is with the non–numeracy of the geographical researchers. Secondly, my comments were not actually aimed at planning models, but at the spatial dependence models discussed below. To concentrate the reply on one area, which did not appear to be represented in the paper, is misleading.

I will now elaborate on my misgivings about models for spatial dependence. When it is possible to envisage a development over time, in which present events depend in some way on previous events, it may be reasonable to attempt to model this development using a 'generative' model. For purely spatial data it is not possible to imagine such a development. Besag (1974) says of spatial models that '... our models will not be mechanistic and must be seen as merely attempts at describing the "here and now" of a wider process'. Some of the early discussions in statistics over simultaneous and conditional models appeared to depend on the belief that such generative models had some meaning outside describing the data. This attitude still appears to pervade geographic research.

Thus, without consulting the data, it is forthrightly assumed that the covariance structure is modelled by, for instance, a one–parameter first–order conditional process or a one–parameter first–order simultaneous process. For example, Haining, Griffith and Bennett (1989) state that 'a first–order conditional autoregressive model ... has ... a monotonically decaying correlation function which seems appropriate for social and environmental spatial data'. They then use the non–stationary edge–corrected version of this model on some remotely sensed data, with only a cursory check for suitability, although they do allow the possibility that the model represents the deviations from a second–order trend surface. I know of no physical or geographic reason why the dependence should of necessity be adequately fitted by this model. The data collection may require the participation of non–numerate individuals (and instruments), but I would not find the argument convincing. There are of course many other correlation functions that monotonically decay with lag.

2.2. First–order models

Assume henceforth that for a given set of n sites or regions there is an n–vector of observations $\mathbf{y}$ with mean $\mu = E(\mathbf{y})$ and covariance (or dispersion) matrix $V\sigma^2$. The one–parameter first–order conditional process is usually taken in geographic modelling as specifying the inverse covariance matrix V^{-1} in the form $I - \beta W$, where W is a symmetric matrix of non–negative weights that are assumed known, and are usually taken as zero down the main diagonal. In this form it is a very simple and convenient model. Parameter estimation is particularly simple. Note that there is no need for the row sums of W to be constant, nor any great advantage when they are; and that elements could be negative. Also, the diagonal elements do not need to be zero, although the conditional means are not then linear in β, as noted below.

Gaussian maximum likelihood requires (see, for instance, Martin, 1984) the minimization with respect to β of $|V^{-1}|^{-1/n}\mathbf{e}'V^{-1}\mathbf{e}$, where $\mathbf{e} = \mathbf{y} - \hat{\mu}$, and $\hat{\mu}$ is an estimate of μ. This involves the calculation of the quadratic form $\mathbf{e}'V^{-1}\mathbf{e}$ and the determinant $|V^{-1}|$. Both of these are very easy for a given β, since $\mathbf{e}'V^{-1}\mathbf{e} = \mathbf{e}'\mathbf{e} - \beta\mathbf{e}'W\mathbf{e}$ and so is linear in β, whilst $|V^{-1}| = \prod(1 - \beta\lambda_i)$, where the λ_i are the eigenvalues of W.

Thus exact Gaussian maximum likelihood is easily performed using a one–dimensional search over the admissable range of β (to ensure that V^{-1} is positive definite), which is in general, provided $\lambda_{\min} < 0$ and $\lambda_{\max} > 0$, $(\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$, where $\lambda_{\min} = \min_i\{\lambda_i\}$ and $\lambda_{\max} = \max_i\{\lambda_i\}$. This is the appropriate range when the diagonal elements of W are zero. Note that this range is more general than that given in many geographical publications—see, for example, Haining (1987, 1988), and Haining, Griffith and Bennett (1989). For example, if $n = 3$ and the off–diagonal elements of W are all $1/2$, as used by Brandsma and Ketellapper (1979), then the admissable range of β is $(-2, 1)$. Note also that that if W consists of non–negative elements and does have constant row sums c, then $\lambda_{\max} = c$, so that we require $\beta < c^{-1}$. Also, by the Perron–Frobenius theorem, $\lambda_{\min} \leq -c^{-1}$. Thus β has a simple upper bound, rather than $|\beta|$ having a simple bound, as was stated in Martin (1987).

The differential and second differential of the likelihood can also be easily found, so that maximisation routines that use the differential can be used. For example, the Newton–Raphson procedure given by Ord (1975) for the one–parameter first–order simultaneous scheme can easily be adapted. In this case, using

$$f(\beta) = -n^{-1}\sum \ln(1 - \beta\lambda_i) - \ln(\mathbf{e}'V^{-1}\mathbf{e}),$$

where $\mathbf{e}'V^{-1}\mathbf{e} = \mathbf{e}'\mathbf{e} - \beta\mathbf{e}'W\mathbf{e}$, the iteration for $\hat{\beta}$ becomes

$$\hat{\beta}_{r+1} = \hat{\beta}_r - f_\beta(\hat{\beta}_r)/f_{\beta\beta}(\hat{\beta}_r),$$

where

$$f_\beta(\beta) = n^{-1}\sum\{\lambda_i/(1 - \beta\lambda_i)\} - (\mathbf{e}'W\mathbf{e})/(\mathbf{e}'V^{-1}\mathbf{e})$$

and

$$f_{\beta\beta}(\beta) = n^{-1}\sum\{\lambda_i/(1 - \beta\lambda_i)\}^2 - \{(\mathbf{e}'W\mathbf{e})/(\mathbf{e}'V^{-1}\mathbf{e})\}^2.$$

However, care should be taken to ensure numerical accuracy and to monitor convergence, as $\hat{\beta}$ is often close to the upper limit of its admissable range. Ripley (1988, Section 2.1) notes that the Newton–Raphson procedure may fail for some data sets.

For some configurations of the sites, the eigenvalues are known theoretically. For example, for an n_1 by n_2 rectangular lattice with W containing ones for immediate horizontal or vertical neighbours, the $n_1 n_2$ eigenvalues of W are given by

$$\lambda_{ij} = 2\cos\{\pi i/(n_1+1)\} + 2\cos\{\pi j/(n_2+1)\} \text{ for } i = 1,\ldots,n_1 \text{ and } j = 1,\ldots,n_2.$$

Gasim (1988) has obtained eigenvalues of W when further neighbours are included, although it is difficult to see the practical use of such W (he actually obtains his results for a one–parameter simultaneous process, but they hold equally for the conditional). In other situations the eigenvalues of W need only be calculated numerically once.

Results on the (Fisher) information under Normality can also be easily obtained for this model. Formulæ for the information, defined as the expected value of the second differential of the log likelihood, are given by Mardia and Marshall (1984)—see also Martin (1984). Now, when $V^{-1} = I - \beta W$ we get $\frac{\partial^2 V^{-1}}{\partial \beta^2} = 0$, so that the most convenient form to take for twice the information on β is $2I_{\beta\beta} = -\frac{\partial^2 \ln|V^{-1}|}{\partial \beta^2}$, which is $-\frac{\partial^2\{\sum \ln(1-\beta\lambda_i)\}}{\partial\beta^2}$. Therefore, $2I_{\beta\beta}$ is $\sum\{\lambda_i/(1-\beta\lambda_i)\}^2$. Since $\frac{\partial V^{-1}}{\partial\beta} = -W$, another convenient form for $2I_{\beta\beta}$ is

$$\text{trace}(V\frac{\partial V^{-1}}{\partial\beta}\ V\frac{\partial V^{-1}}{\partial\beta}) = \text{trace}\{(V\frac{\partial V^{-1}}{\partial\beta})^2\}.$$

For small n, this is easiest found as the sum of squares of the elements of $V\frac{\partial V^{-1}}{\partial\beta} = -VW$, using $\text{trace}(A^2) = \Sigma\Sigma a_{ij}^2$ when A is symmetric. Otherwise, we can use the fact that the trace of a matrix is the sum of its eigenvalues, and that the eigenvalues of $-VW = (I-\beta W)^{-1}W$ are $\lambda_i/(1-\beta\lambda_i)$, $i = 1,\ldots,n$, so that those of $(-VW)^2$ are $\{\lambda_i/(1-\beta\lambda_i)\}^2$.

If we want to get the asymptotic variance of $\hat{\beta}$, the maximum likelihood estimator of β, then we also need $I_{\beta\sigma^2}$. The simplest form for $2\sigma^2$ times this is $\text{trace}(-V\frac{\partial V^{-1}}{\partial\beta})$, which is therefore $\sum\{\lambda_i/(1-\beta\lambda_i)\}$. Although the previous result on $I_{\beta\beta}$ has been used by geographers, this result on $I_{\beta\sigma^2}$ has not—see Haining, Griffith and Bennett (1989), and Martin (1989a). These results give a simple formula for the asymptotic variance of $\hat{\beta}$ as 2 over the corrected sum of squares of the $\lambda_i/(1-\beta\lambda_i)$, $i = 1,\ldots,n$, (Martin, 1989a), although asymptotically equivalent forms are easier to use. The latter are considered in Besag and Moran (1975) and Besag (1977b).

It is important to realise that the form $V^{-1} = I - \beta W$ is not the inverse variance matrix of a second–order stationary one–parameter first–order conditional process when the sites or regions form a finite regular lattice. That is, V is not proportional to a correlation matrix. There are several ways of seeing this. A simple way is to invert numerically a given V^{-1}, and note that, for instance, the diagonal elements are not constant. Although geographers are becoming more aware of this fact—see, for instance, Haining (1987)—there does still appear to be some confusion. For example, Haining, Griffith and Bennett (1989) use $V^{-1} = I - \beta W$, but also appear to assume V proportional to a correlation matrix—see my comments in Martin (1989a).

For a given W with zeroes on the main diagonal, the one–parameter first–order conditional process can be written in the form

$$E(y_i\,|\cdot) = \mu_i + \beta\sum w_{ij}(y_j - \mu_j), \quad \text{with } \text{var}(y_i\,|\cdot) = \sigma_\eta{}^2,$$

where the conditioning is on all other values, y_j, $j \neq i$. Note that the assumption that all sites have the same conditional variance is often not reasonable for a finite set of sites—in particular, it is usually preferable that the conditional variance is smaller for the interior points. It is possible to postulate unequal conditional variances for the one-parameter first-order conditional process, but W must then be asymmetric with $w_{ij}\sigma_j^2 = w_{ji}\sigma_i^2$, where $\mathrm{var}(y_i\,|\,\cdot) = \sigma_i^2$ (Besag, 1975). Another possibility would be to use a symmetric W, but for W to have non-zero diagonal elements. Then, provided $1 - \beta w_{ii} > 0\ \forall i$, $\mathrm{var}(y_i\,|\,\cdot) = \sigma_\eta^2/(1 - \beta w_{ii})$. However, the conditional mean would now have the form

$$E(y_i\,|\,\cdot) = \mu_i + \frac{\beta}{1 - \beta w_{ii}} \sum_{j \neq i} w_{ij}(\mathbf{y}_j - \mu_j)$$

which is non-linear in β.

On an infinite lattice, the second-order stationary process is such that w_{ij} only depends on the lag $i - j$. For a finite lattice, define an interior site as a site i for which all the sites appearing in the expression for $E(y_i\,|\,\cdot)$ are observed. Then, part of the confusion about stationary is due to the fact that provided either site i or site j is an interior site, the (i, j) element of V_S^{-1}, where S denotes the stationary form, is precisely that element of $I - \beta W$. This is easily seen by direct multiplication of V_S and V_S^{-1}, and using known relationships between the correlations—see equation (5.12) of Besag (1974).

Therefore, many results obtained for $V^{-1} = I - \beta W$ do hold for interior sites of the stationary process without modification. Nevertheless, there are many results that do not simply carry over from one form to the other. Of particular importance are the results of Guyon (1982), who showed that using Gaussian maximum likelihood for one form may lead to estimators with undesirable properties for a different form. Thus, great care should be taken to precisely define which form is being postulated. This care is not yet sufficiently in evidence.

Similarly, the one-parameter first-order simultaneous process has V^{-1} of the form

$$(I - \beta W)'(I - \beta W),$$

where in this case W does not need to be symmetric. Note that the diagonal elements of $W'W$ usually differ, and are usually greater for interior sites, so that the conditional variance at these sites is reduced. In fact, assuming $w_{ii} = 0$, $\mathrm{var}(y_i\,|\,\cdot) = \sigma_\eta^2/(1 + \beta^2 \sum_j w_{ji}^2)$. Whilst it may be desirable that the conditional variance is smaller for interior sites, the precise variances arise from the modelled neighbours, rather than being directly specified.

This model has also been used without question—for instance, see Haining (1987). The model is also simple and convenient, although not as simple as the conditional process. Since $|A'A| = |A|^2$, exact Gaussian maximum likelihood can easily be performed using the eigenvalues of W (Ord, 1975), although these eigenvalues may be complex if W is not symmetric. In many cases, the postulated W is the same as a possible W for the conditional process, in which case W is symmetric and has the same eigenvalues as before.

In this common case that W is symmetric, it is possible to get simple results for the information. Then $V^{-1} = (I - \beta W)^2$, so that its eigenvalues are $\{(1 - \beta\lambda_i)^2\}$, and V and W commute. Therefore $-V\frac{\partial V^{-1}}{\partial \beta} = 2W(I - \beta W)^{-1}$, with trace $2\sum\{\lambda_i/(1 - \beta\lambda_i)\}$. Since W

and W' have the same eigenvalues, this formula also holds even when W is not symmetric. Also

$$V\frac{\partial V^{-1}}{\partial \beta}\, V\frac{\partial V^{-1}}{\partial \beta} = 4W^2(I-\beta W)^{-2},$$

with trace $4\Sigma\{\lambda_i/(1-\beta\lambda_i)\}^2$. These are just multiples (2 and 4) of the values for the conditional process. The asymptotic variance of $\hat{\beta}$ is therefore exactly one quarter of the value for the conditional process, which was discussed above. Since Haining (1987) uses a symmetric W, the general formulæ misquoted from Ord (1975), and the approximation given, are quite unnecessary. Note that Ord's (1975) α should be $-\Sigma\{\lambda_i/(1-\beta\lambda_i)\}^2$.

2.3. More general models

Extensions to the one–parameter conditional or simultaneous forms have been suggested. For instance it is easy to extend the conditional form for V^{-1} to the two–parameter form $I-\beta_1 W_1-\beta_2 W_2$, which either extends the range of dependence or can be used for the same range of dependence as before, but with W split into two parts to allow different degrees of dependence in different directions. This extension, at least in the simultaneous form, is usually attributed in the geographic literature to Brandsma and Ketellapper (1979), although the idea was hardly new to statistics. Even in spatial statistics the use of more than one parameter is well established—see Whittle (1954). The simultaneous process can itself, at least on an infinite lattice, be represented as a special case of a conditional form with separate parameters for each of the immediate horizontal or vertical neighbours, the immediate diagonal neighbours, and the lag–two horizontal or vertical neighbours—see Besag (1974).

However, if it is wished to keep some of the simplicity of the one–parameter conditional or simultaneous forms, there are not many extensions available. The ability to obtain eigenvalues of V^{-1} that are linear in the parameters β_i is hampered by the requirement that the W_i matrices need to commute. Using powers of W is possible, but is not always satisfactory. Unless W is triangular, W^2 has some diagonal elements positive, so that the conditional variance for those sites is reduced. For a rectangular lattice the most general conditional form with known eigenvalues that does not use powers has $V^{-1}=I-\beta_1W_1-\beta_2W_2-\beta_3W_3$, where W_1 is for horizontal neighbours, W_2 for vertical neighbours, and W_3 for the four diagonal neighbours. Squaring this gives the most general simultaneous form.

Note that whenever V (or V^{-1}) has a simple eigenvalue/eigenvector decomposition, $V = P\Lambda P'$ where the columns of P are the standardized (or normalized) eigenvectors of V and Λ is a diagonal matrix of the corresponding eigenvalues, then a simulation is easily obtained using $\mathbf{y} = \mu + P\Lambda^{1/2}P'\varepsilon$, where $\Lambda^{1/2}$ is a diagonal matrix of the square roots of the eigenvalues, and ε is a vector of simulated independent random variables with mean zero and variance σ^2. There is no need to numerically find the Cholesky square root of V, as suggested by Haining, Griffith and Bennett (1983). Similarly, if V has the simultaneous form $B'B$ where B has a simple eigenvalue/eigenvector decomposition, $B = P\Lambda P^{-1}$, we can use $B^{-1} = P\Lambda^{-1}P^{-1}$ where Λ^{-1} is a diagonal matrix of the inverses of the eigenvalues, so that $\mathbf{y} = \mu + B^{-1}\varepsilon$. This should be preferable to numerically inverting B, as suggested by Haining, Griffith and Bennett (1983), and reasonable for even moderately large n.

Another simple extension is to use the above forms for V^{-1} as forms for V—finite dependence or moving–average models. Note that the finite dependence models are perfectly reasonable models of the covariance, although the attempt to derive them through a 'generating' mechanism in Cliff and Ord (1981, p. 150) is incorrect. The eigenvalues of V will again be linear in the parameters. The quadratic form $\mathbf{e}'V^{-1}\mathbf{e}$ can be quickly computed as $\mathbf{f}'\mathbf{\Lambda}^{-1}\mathbf{f}$, where $\mathbf{f} = P\mathbf{e}$ and V has the eigenvalue/eigenvector decomposition $P\mathbf{\Lambda}P'$. Simple results can be obtained for the information using, this time, trace($V\frac{\partial V^{-1}}{\partial\beta}$) and trace($V\frac{\partial V^{-1}}{\partial\beta}\ V\frac{\partial V^{-1}}{\partial\beta}$), so that essentially the same results are obtained as before. The rapid decay to zero of the covariances makes this form less attractive in many practical situations. It is also possible to combine the two forms, and still keep the same eigenvectors provided the W_i matrices commute.

One other possible extension that does preserve some simplicity in the likelihood is the errors–in–variables formulation (Besag, 1977a). Essentially, this approach uses one of the above V matrices, and adds to it αI, so that $\text{var}(\mathbf{y}) = (V + \alpha I)\sigma^2$. This is useful when the sample correlation function does not appear to tend to 1 as the lag tends to 0, as when there is an extra independent error, such as measurement error, at each site. The quadratic form $\mathbf{e}'V^{-1}\mathbf{e}$ is found in the same way as when V is specified.

Another extension for data on a rectangular lattice is to the separable processes, which can often be very easily fitted (Martin, 1990). These processes are somewhat restrictive in the range of possible covariance structures—in particular correlations must be reflection symmetric and decay exponentially—but the ease of specification and fitting makes them attractive whenever the assumption is reasonable. Simulation of a separable process is relatively easy provided the sites can be represented as a subset of a rectangular lattice, since on an n_1 by n_2 rectangular lattice V is a Kronecker product of dispersion matrices of orders n_1 and n_2, and square roots of these matrices are usually easily found (Martin, 1990).

2.4. Comparison of models

From a data analytic point of view, it is important to be able to fit different models, and compare their fits. If the models are hierarchical, each more general than the previous, then standard statistical tests often can be used as a guide, although the theoretical justification is frequently lacking. Note that if n is not too large, say less than 100, then provided care is taken to ensure numerical accuracy, any model can be fitted, whether or not V has simple eigenvalues. If computing problems are ignored, it is easy, for regularly arranged sites or regions, to postulate a series of models, each more general than the previous one, with natural orderings of the neighbours (with, in general, different parameters for the two directions, but the possibility that the two parameters are the same). The next extension is to include the four diagonal neighbours (again, in general there would be different parameters for the two directions). For the subclass of separable processes, this procedure can actually be easily accomplished, because of the ease of fitting most processes (Martin, 1990).

It is much harder to say what should be done with irregular sites or regions. Note that despite the attempts at developing theory on a rectangular lattice, it is the irregular sites or regions that are most common for natural geographical data. Irregular sites can still be modelled with a particular dependence structure, and the form for V deduced from it. Irregular regions cause the most problems. Modelling for data on irregular regions has tended to be extremely arbitrary. A particular form is postulated, such as the one–parameter conditional form of $V^{-1} = I - \beta W$, and then the weights w_{ij} also are arbitrarily chosen from a wide range of possibilities. These include functions of the distance between arbitrary centroids, and functions of the common boundary length (see Besag, 1975). In most applications the weights are more like parameters than known constants.

It is easy to criticise, but less easy to make constructive suggestions. My own belief is that for a given set of neighbours the elements of W_i should be parameterized in terms of a small number of parameters, so that different choices of W_i can be compared using standard statistical theory. This approach also has been suggested by Brandsma and Ketellapper (1979). Unfortunately, separability does not appear to have any relevance for data on irregular regions.

So far, I have discussed the modelling of the covariance structure. Of course there are other considerations. The mean structure can also be specified, and may be dogmatically specified, or chosen after examination of the data. The use of a trend surface to represent the large scale variation, as in Haining (1987), is fraught with difficulties when the dependence is also modelled through the covariance. Even a second–order stationary process can exhibit trend–like behaviour, so that the partition into a fixed trend and a random component is not clear. In addition, a parameterized trend surface is usually far too inflexible over a large region, and may require many parameters. Unless there is a clear planar trend over the sites, many statisticians would prefer to model the trend–like behaviour through differencing, or the use of the intrinsic processes introduced in geostatistics (see, for example, Journel and Huijbregts, 1978; and, the extension to intrinsic autoregressions of Künsch, 1987).

Another consideration is the distribution. Normality is almost always assumed, often implicitly. There is usually no check on normality; and, where there is, it usually consists of a univariate histogram of the original data. Apart from the necessity of correcting for the mean function, I have remarked before (Martin, 1983) on how misleading the marginal histogram can be for correlated data. The histogram of normal correlated data can often be multimodal and skewed. The need for correcting the significance values of a goodness–of–fit test for two–dimensional data was shown by Patankar (1954). My view (see Martin, 1990) is that some attempt should be made to obtain approximately uncorrelated residuals, on which standard tests of normality (for example, using as a test statistic the sample correlation coefficient associated with a normal probability plot) can be approximately used.

Note that, whilst it is important to check the distributional properties when simulating data, there is no point in checking the derived data, as suggested in Haining, Griffith and Bennett (1983). It is better and simpler to check that the original simulated data for ε satisfy the assumptions of normality, constant variance, and independence.

3. Statisticians and geographers

There have been calls for statisticians to become involved in geographic research (see, for instance, Cliff and Ord, 1975; Bennett and Haining, 1985). My own involvement has been by a somewhat strange route. The published research in geography, which uses spatial statistics, held, and still holds, no particular interest to me. As I already have mentioned, I cannot see the point of most of it, and much is riddled with statistical errors. If that published research were concerned with applications in geography, I would probably have remained uninvolved. However, somewhat to my surprise I found that the vast majority of the publications that I had seen were not about geography at all. Although published in geographical journals, they were claiming to contain advances in statistical theory (by statistical theory I mean any non–trivial mathematical manipulations performed within a statistical context).

I take care to seek opinions and comments, before submission of a paper, from others who may have more knowledge of the topics in the paper. I do not find it easy to get my papers published in statistical journals. I rigorously check any paper many times between the first draft and the final proof version for errors and misprints. I therefore was disturbed by the apparent ease with which these papers were able to appear in geographical journals, the kudos the authors received, the statistical errors the papers contained, and the lack of generality of reported results. Of course, there are poor papers in statistical journals, and there is good statistical work done by members of disciplines other than statistics. There also are many papers abusing statistics of which I am thankfully unaware. It seems unfortunate to me that non–statisticians who publish statistics, however flawed, are held in high esteem in their professions, whilst statisticians who publish statistics, however good, are seen as doing no more than trade. I can only say that I became aware of these "geographical" papers, and responded to them.

My first response (Martin, 1984) was a purely statistical work, but resulted from seeing some tentative beginnings by geographers. In this case I had already looked at the theory, but had not had the time (or the stimulus) to prepare it for publication (see my comments in Section 5). My second response (Martin, 1987) was a direct result of seeing published work by geographers. In that paper I attempted to straighten out what I saw as some muddled thought (see my comments in Section 4), and to correct some of the errors I had noted. To the credit of the geographic community, this response was published. However, this credit is somewhat diminished by the fact that I was not made aware of the reply (Griffith, 1988), nor given an opportunity to comment on it before or after publication. My third response (Martin, 1989a) also was a purely statistical work; it attempted to give the theory behind some numerical results obtained by geographers. This work was interesting in that it would never have occurred to me to obtain the results if it had not been for the other paper. Indeed, I still doubt that the results have any practical significance (see my discussion in Section 5). This present paper constitutes an invited fourth response, which I have used to develop my previous arguments.

I thus have four papers that may be of interest to geographers, but I still do not see myself as being, or wanting to be, a statistician interested in geographical research. I have reacted to geographical research, and may continue to do so. However, I have long thought that I would have nowhere near enough time, even if it were my only aim, to keep up with, and correct, statistical publications by geographers in the area of my research interest.

Perhaps fortunately, I am not kept in touch with current geographical research in this area, so that I only find out about it on the rare occasions that I am asked to referee a paper, or a paper appears in a journal that I notice. In connection with this, Bennett and Haining (1985) note, on geographic modelling, that I appeared unaware of "an extensive methodological discussion of these points ... in the social sciences and geography." I am happy to acknowledge my unawareness of this discussion. I feel that if geographers wish statisticians to become involved in their research, then the onus is on them to help make their research accessible to statisticians. It is quite unreasonable to expect a statistician to keep abreast of all the research in all the areas that use or abuse statistics.

My second point in Martin (1985) was that research purporting to contain statistical advances should be submitted to the scrutiny of statisticians. My hope was that better refereeing would result in better papers, and that more statisticians would become aware of the research. I am pleased to see that there are now submissions to statistical journals; for example, Haining (1987, 1988), and Haining, Griffith and Bennett (1989). The drawback is that the statistical community must now share the blame for any criticisms of these papers. That I do have criticisms can be seen from my comments in Martin (1989a), and in this paper. As an additional example, I will mention that the reference to Matérn's lower bound of −0.403 (Haining, 1987, p. 464) is incorrect. This bound is for an isotropic process in continuous space, and has no relevance in discrete space. Even the concept of "isotropy" has little meaning or relevance for regional data, such as pixel measurements.

However, worse things are still occurring in quantitative geographical journals. For instance, Griffith (1987) gives (p. 72) an 11-line derivation of the simple result $E(\mathbf{x}) = (1-\rho)^{-1}\mu_\xi \mathbf{1}$ when $\mathbf{x} = (1-\rho C)^{-1}\xi$, $E(\xi) = \mu_\xi \mathbf{1}$, and C has row sums equal to one (and does not achieve this result). The paper contains several errors, admittedly relatively minor once the necessary assumptions have been deduced. Also, the torus limit of $\rho_{\mathrm{g,h}}$ must equal the planar value (see Martin, 1986).

Poor published research does not represent a step forward, but several steps backwards. It sets a standard for subsequent publications, and deters those who might have worthwhile contributions to make. I wonder what the reaction of the geographic community would be if a statistician published, in a statistical journal, articles suggesting the present state of, and future research necessary in, geography. I wonder why people wishing to do research in statistics do not liase with statisticians who are expert in that area of research. I wonder why geographers do not concentrate on the many interesting geographic problems that are amenable to a sensible use of statistics.

As examples, I would like to see geographers investigating, by examining many relevant data sets, what models are reasonable for the sorts of data that arise in geographical applications. I would like to see investigations of different methods of predicting "missing values" in geographical situations in which an answer is actually required.

I also would like to see geographers who meet a statistical problem actively seeking the views and help of statisticians well before the stage of seeking publication. I am confident that many statisticians would be interested in such investigations and ready to help, when asked, with any necessary theoretical developments.

4. Boundary effects

There have been several papers discussing the "problem" of boundary values, and possible solutions to this "problem" (see Griffith, 1980, 1983, 1985, 1987; Griffith and Amrhein, 1983; Ord, 1981). In Martin (1987), written before I had seen Griffith (1987), I examined the "problem" and came to the conclusion that the published research was unsuccessful because the problem had not been sufficiently well defined, and the research had not considered problems that might be of interest. Some of the discussion is worth elaborating on here.

Much of the previous discussion on the topic of the "boundary value problem" appeared to assume that the problem was a well-defined one, and that a statistical solution to the problem was possible. I suspect that some of the confusion was due to the use of the term "boundary value," which has certain connotations in Applied Mathematics. In solutions to differential or difference equations, a general solution is found that depends on certain initial conditions. Once these initial conditions are specified, the solution is unique. Also, a geographical boundary between sites is different from the boundary sites at which "boundary values" may occur.

However, the spatial statistical models used by geographers are of a different kind. I already have stated my view that these models are only descriptive of the covariance of the data, and have no meaning as generative models of the data. This still applies even when the model can be expressed in a "generative" or "causal" form. Thus the fact that these "generative" models include, for some "boundary" sites, dependence on unobserved sites, is irrelevant. Also irrelevant is any attempt to predict these boundary values in order to estimate parameters (Martin, 1987). Even the definition of what are boundary sites is unclear. One definition was used here in Section 2, but many others are possible.

A simple example is given by the first–order autoregression in one dimension, namely $x_i = \alpha x_{i-1} + \varepsilon_i$, where $\{\varepsilon_i\}$ is a sequence of uncorrelated random variables with zero mean and constant variance. With finite data $\{x_i\}$, $i = 1, 2, \ldots, n$, it appears that an assumption about x_0 needs to be made. However, for spatial data we could equally well "assume" that the data were "generated" from the right, with the model $x_i = \alpha x_{i+1} + \varepsilon_i$, so that now it appears that we need an assumption on x_{n+1}. If we write the model in conditional form, then $E(x_i|\cdot) = \beta(x_{i-1} + x_{i+1})$, where $E(x_i|\cdot)$ denotes the mean of x_i given all other xs and $\beta = \alpha/(1+\alpha^2)$, so that in this formulation we need assumptions about both x_0 and x_{n+1}. All these requirements concern what I term "exterior boundary" values; but it also is possible to formulate the problem in such a way that we need assumptions about x_1 and/or x_n—what I term "interior boundary" values.

A preferable way of considering the problem is through the covariance structure of $x_1, x_2, \ldots, x_n$. If the covariance matrix is V, then the only elements that change under the different assumptions on the first–order autoregression are the $(1,1)$ and/or the (n,n) elements of V^{-1}. We therefore essentially require assumptions about these. It also is possible to define the covariance structure of a larger set of xs, for instance $x_0, x_1, \ldots, x_{n+1}$, and then derive V from this. This is discussed in Martin (1987), and eight different forms that have been suggested for V^{-1} are given in Kunert and Martin (1987). Note that it is quite unnecessary to believe that the data were generated temporarily from an infinite past, or spatially from an infinite space, in order for V to have the stationary form. Thus the use of a finite geographic region does not, of itself, rule out the use of the stationary V. However, doubtlessly it is true that when the region considered has natural boundaries, it may be reasonable to expect those sites on the geographic boundary to have different properties from those sites in the geographic interior.

Note that we can produce identical effects by including different assumptions on the variances of some of the "innovations." For instance, the assumptions that $x_0 = 0$ and $\mathrm{var}(\varepsilon_1) = \sigma_\varepsilon^2/(1-\alpha^2)$ lead to the stationary form for V^{-1}.

Thus the first possible boundary effect is that for a given model different "boundary" assumptions lead to different dispersion matrices. Since it is unlikely that even a large data set would allow statistical differentiation between mildly different "boundary" assumptions, the choice is largely a matter of convenience, unless there are strong prior arguments for one form over all other forms. The specification of a reasonable model for the "interior" sites usually will be more important than the specification of the precise form of the model to be used, although attempts should always be made to incorporate good prior information.

In more that one dimension there is another problem. The stationarity assumption for many statistical models does not lead to a V or a V^{-1} that can easily be numerically calculated. This means that exact likelihood is not feasible. Since it is, at least for the finite conditional or simultaneous schemes, the "boundary" sites that cause problems for V^{-1}, we have a second possible boundary effect, which is that some stationary models cannot, at present, sensibly be fitted by exact likelihood. This, together with results reported by Guyon (1982) on approximate likelihood not necessarily being $\sqrt{n}$-consistent, suggest that we should not attempt to fit the stationary form, but one of the other forms that is associated with different "boundary" assumptions.

The third possible boundary effect is that estimators may be biased, and that different "boundary" assumptions may reduce this bias. Whilst this is undoubtedly true, it is not at all clear in what way some boundary assumptions reduce the bias for estimators of parameters of V for the same, or other, forms; nor is it clear whether changing the "boundary" assumptions is a good way to reduce estimator bias, or even whether the bias is large enough to cause concern.

The geographical discussion of "boundary effects" is greatly complicated by the lack of clear definitions of what are the effects that are causing concern, and what are the attempted solutions to them. It is difficult to comment on ambiguous or unclear work, since there is always the possibility that there are hidden assumptions that make the analysis correct. A step forward in research would be for all assumptions and aims to be clearly stated.

5. Missing Values

Although I had looked at the theory for estimation of the parameters of a spatial model when observations at some sites are not available as early as 1978, it was not until 1983, when I was told that geographers were working on a special case of the problem and were encountering difficulties, that I completed and wrote up the work (Martin, 1984). The results were circulated earlier, and mentioned in Martin (1983). This work covered in full generality the estimation of parameters using exact maximum likelihood for a Gaussian process. The work was referred to in several subsequent publications by geographers—see for example Bennett, Haining and Griffith (1984), Griffith, Haining, and Bennett (1985), and Haining, Griffith and Bennett (1984, 1989), although not always correctly, as I pointed out in Martin (1987).

There are two aspects to 'missing values'. One is the ability to use, with possibly minor modifications, known estimation methods on a given configuration of sites, usually a regular rectangular lattice. The results are of most usefulness when the covariance matrix, or its inverse, is of a known simple form on a given configuration, and m, the number of unobserved sites, is small. The other aspect is the prediction of the unobserved values.

Much of the original impetus for the interest of geographers appears to have been as a possible 'solution' to the 'boundary problem'. As I discussed in Martin (1987), and have commented again above, missing value techniques are quite irrelevant to the 'boundary problem'. Since then, more realistic problems have been proposed in which missing value techniques may be valuable. One is in the area of analysis of remotely-sensed data. For such data, it is possible to have unobserved sites for several reasons. Two possibilities are cloud cover when a passive sensor is used, and instrument malfunction. The former will result in unobserved data in all bands over a region on the ground, whilst the latter may result in the loss of data on individual pixels or lines of pixels, and may only affect one band. Another situation in which 'missing data' arises is when the data contain possible outliers, that is observations that appear unusual amongst the others, or influential observations, that is data which have a large effect on the analysis. It may then be desirable to perform any analysis with such observations omitted. Some general theory on influence and residuals for known V is in Martin (1989b). It may even be sensible to routinely calculate such 'leave-k-out statistics' as a diagnostic procedure—see the time series case in Bruce and Martin (1989). Procedures for dealing with an unknown mean and an unknown dispersion matrix require further investigation (Martin, 1989c).

Although the application to remotely-sensed data has been mentioned recently (Haining, Griffith and Bennett, 1989), the example given is unsatisfactory, in that there is little indication that the chosen covariance structure, the one-parameter first-order conditional process, is an adequate representation.

The main purpose of that paper appears to be to advance statistical theory on the loss in information (here meaning the Fisher information) when some sites are unobserved. The interest appeared to be on how the loss varies over different spatial configurations of the spatial sites. Results were obtained numerically for the special case of the one-parameter conditional process on a rectangular lattice.

The paper does not explain what the purpose of obtaining these results is. Since in any application the unobserved sites are given, and are not in the control of the investigator, it is difficult to see what the point is in comparing different configurations and different numbers of unobserved sites. However, if we assume that the results are of interest, it is easy to obtain theoretically much more powerful results. I have given the appropriate theoretical results in Martin (1989a). Special cases can easily be found—all the cases considered by Haining, Griffith and Bennett (1989) are also given in Martin (1989a). Many other special cases can also be considered, although it is only for the one-parameter conditional process that the formulæ are at all simple. Mrs. T. Krug at Sheffield has obtained formulæ for more sites and for the one-parameter first-order simultaneous model.

Assuming that there is an interest in these results, I shall outline some of them, elaborate on some of the details omitted in Martin (1989a), and include some new results. Assume that the n-vector of observations (strictly the random variable) is $\mathbf{u}$, with dispersion matrix $\mathrm{var}(\mathbf{u}) = V\sigma^2$, where σ^2 is a scale parameter and V depends on β. Although it is possible to allow the mean to include trend and other fixed effects, I shall just discuss here the case of a constant mean, so that $E(\mathbf{u}) = \mu \mathbf{1}_n$, where $\mathbf{1}_n$ is an n-vector of ones. Also, it is easy to generalize to the case that V is a function of the q-vector λ.

Assume also that data are unavailable at m of the sites, and that $\mathbf{u}$ is permuted into $\mathbf{x}$, where the first $n-m$ elements of $\mathbf{x}$ are $\mathbf{y}$ and correspond to the observed sites, while the last m elements are $\mathbf{z}$ and correspond to the unobserved sites. Similarly, let $\mathrm{var}(\mathbf{x})/\sigma^2 = V_{xx}$ be partitioned as

$$\begin{pmatrix} V_{yy} & V_{yz} \\ V_{zy} & V_{zz} \end{pmatrix}$$

and V_{xx}^{-1} as

$$\begin{pmatrix} V^{yy} & V^{yz} \\ V^{zy} & V^{zz} \end{pmatrix} = \begin{pmatrix} V^{yx} \\ V^{zx} \end{pmatrix}$$

Then, in general, the loss in information on μ when m sites are not observed is $\mathbf{c}'(V^{zz})^{-1}\mathbf{c}\sigma^2$, where $\mathbf{c} = V^{z\bar{z}}\mathbf{1}_m$. For the one-parameter conditional process with $V^{-1} = I - \beta W$, we find that $V^{zz} = I - \beta W_{zz}$, where W is partitioned similarly to V. Also, for any interior site of the first-order process on a rectangular lattice, $\mathbf{c} = \alpha\mathbf{1}_m$, where $\alpha = 1 - 4\beta$. Thus the information loss then becomes $\mathbf{1}_m'(I - \beta W_{zz})^{-1}\mathbf{1}_m$ times $\alpha^2\sigma^2$.

Exact formulæ can be obtained for this situation. Some general results, plus particular formulæ for the different configurations when $m = 1, 2, 3, 4$ are given in Martin (1989a). Note that when $m = 4$, one of the configurations was omitted by Haining, Griffith and Bennett (1989). For this case one configuration is

XXX
 X

and the loss is $\frac{4+6\beta}{1-3\beta^2}$ times $\alpha^2\sigma^2$. This is intermediate in its loss between cases 3(d) and 3(e) of Haining, Griffith and Bennett (1989). For the values of β they consider, 0.075, 0.150, and 0.225, the loss is 2.218, 0.840, and 0.063 respectively.

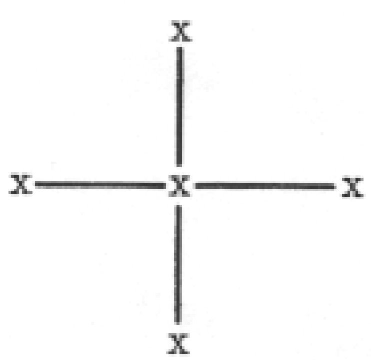
Exact results can be obtained for greater values of m, although the number of essentially different configurations increases rapidly with m, as does the difficulty in general in obtaining the formulæ for the elements of $(V^{zz})^{-1}$. The recursion given below is often useful. However, good approximations are also possible. Provided that $|\beta|$ is not too large, the information loss on β for m missing sites is approximately $\{m + 2m_1\beta + 2(m_1 + m_2)\beta^2\}$ times $\alpha^2\sigma^2$, where m_1 is the number of 'links' of length one among the missing sites, and m_2 is the number of 'links' of length two. These links are found using the usual city-block metric.

As an example, consider the case $m = 5$. There are several cases in which not all the sites are joined, but I will only consider those four cases in which all sites are connected.

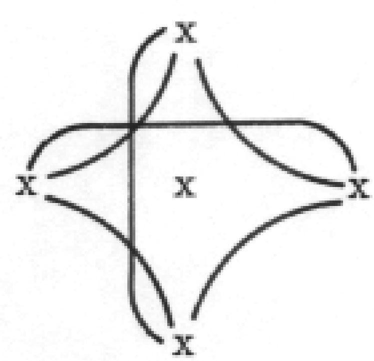
Case 1	*Case 2*	*Case 3*	*Case 4*
X XXX X	XX XXX	X XXXX	XXXXX
(4,6)	(5,6)	(4,4)	(4,3)

The pair of numbers associated with each configuration are (m_1, m_2). The figure below shows how these are obtained for Case 1.

The 4 links of length 1



The 6 links of length 2



In general, the exact result requires the inversion of V^{zz}. However, using results on partitioned matrices, it is possible to obtain recursively formulæ for the information loss.

Partition the m sites into $m-1$ and 1, so that

$$V^{zz} = \begin{pmatrix} A & \mathbf{b} \\ \mathbf{b}' & d \end{pmatrix}$$

where A is an $m-1$ square matrix, $\mathbf{b}$ is an $m-1$ vector and d is a scalar. Then

$$(V^{zz})^{-1} = \begin{pmatrix} A^{-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix} + \frac{1}{d - \mathbf{b}'A^{-1}\mathbf{b}} \begin{pmatrix} -A^{-1}\mathbf{b} \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -A^{-1}\mathbf{b} \\ 1 \end{pmatrix}'.$$

Thus, if $\mathbf{c} = \alpha \mathbf{1}_m$, which is so for interior points for conditional and simultaneous schemes, then the loss in information on μ is the loss for the $m-1$ sites,

$$(\mathbf{1}'_{m-1} A^{-1} \mathbf{1}_{m-1}) \alpha^2 \sigma^2,$$

plus $\alpha^2\sigma^2$ times $\frac{(1-\mathbf{b}'A^{-1}\mathbf{1}_{m-1})^2}{d-\mathbf{b}'A^{-1}\mathbf{b}}$. This latter additional term can be very easy to calculate if the extra point is carefully selected. For instance, for the one–parameter first–order conditional process, if in Case 1 above the centre point is chosen, then $A = I_4$ and $\mathbf{b} = -\beta \mathbf{1}_4$. Thus the loss is $\alpha^2\sigma^2$ times $4 + \frac{(1+4\beta)^2}{1-4\beta^2}$.

Note that the above result can easily be extended to m sites being partitioned into $m - m'$ and m'.

Results for the one–parameter first–order simultaneous model can be obtained, but are nowhere near as simple. There are several reasons for this. Firstly, $(V^{zz})^{-1}$ does not have the form $(I - \beta W_{zz})'(I - \beta W_{zz})$. Secondly, because V^{-1} has non-zero terms for $(1,1)$ and $(2,0)$ lags, there are more cases to consider. For example, when $m = 2$ there are 3 different configurations, and when $m = 3$ there are 12. The numbers for the conditional process are 2 and 3 respectively.

Thus, when $m = 2$ the four cases are:

Case 1	*Case 2*	*Case 3*	*Case 4*
immediate neighbours	lag 2 neighbours	diagonal neighbours	all other configurations
xx	x·x	x· ·x	

For the conditional process Cases 2 and 3 are included with Case 4. An interesting point with the simultaneous process is that when $\beta > 0$, the smallest loss is not associated with Case 4, but with Case 3. This follows from the element of V^{-1} associated with diagonal neighbours being $2\beta^2$, which is positive. The next smallest is for Case 2, as the element of V^{-1} associated with lag 2 neighbours is β^2, which is also positive.

So far I have considered the easier case of the loss of information on μ. The loss of information on β was also considered in Haining, Griffith and Bennett (1989), and Martin (1989a). This is more complicated for several reasons. Firstly, there are more configurations to consider, and secondly, the formulæ involve both V^{zz} and $(V^{zz})^{-1}$. Because of the second point, the formulæ depend not just on the configuration of sites, but also on the actual positions of the sites. However, provided attention is restricted to interior points of the *stationary* process, then the result only depends on the configuration. Although Haining, Griffith and Bennett (1989) do evaluate their results for the stationary process, it appears that they are also assuming $V^{-1} = I - \beta W$.

Note that if the loss of information on β is being considered because of an interest in $\mathrm{var}(\hat{\beta})$, then the information required is that for β conditional on σ^2, which was considered in Section 2.

The formulæ for the loss of information on β are most easily obtained by using the missing information principle of Orchard and Woodbury (1972). I take their principle to be their equations (2.13) and (2.15); that is, the use of the expectation with respect to x of the conditional likelihood of $z|y$. Setting the mean μ to 0, since its value does not affect the information on β, the distribution of $z|y$ is Normal with mean $-(V^{zz})^{-1}V^{zy}\mathbf{y}$ and dispersion matrix $(V^{zz})^{-1}\sigma^2$.

Since the second differential with respect to β of both $V^{zy} = -\beta W_{zy}$ and $V^{zz} = I - \beta W_{zz}$ is 0, the second differential of the conditional log likelihood becomes

$$-\frac{1}{2}\frac{\partial^2 \ln|V^{zz}|}{\partial\beta^2} + \frac{1}{2\sigma^2}\frac{\partial^2\{\mathbf{y}'V^{yz}(V^{zz})^{-1}V^{zy}\mathbf{y}\}}{\partial\beta^2}.$$

The first term can be evaluated as before. Taking the expectation over $\mathbf{y}$ of the second term gives

$$\frac{1}{2}\mathrm{trace}\left[V_{yy}\frac{\partial^2\{V^{yz}(V^{zz})^{-1}V^{zy}\}}{\partial\beta^2}\right].$$

Now, $V^{yz}(V^{zz})^{-1}V^{zy} = \beta^2 W_{yz}(I - \beta W_{zz})^{-1}W_{zy}$, and so its second differential with respect to β is $2W_{yz}(I - \beta W_{zz})^{-3}W_{zy}$ {compare this with the second differential with respect to x of $x^2/(1-ax)$, which is $2/(1-ax)^3$ }. Then using $(V^{zz})^{-1}V^{zy}V_{yy} = -V_{zy}$ [see Martin (1984)] and $V_{zy}V^{yz} + V_{zz}V^{zz} = I$, it follows that this expectation becomes

$$\beta^{-1}\,\mathrm{trace}\{(V^{zz})^{-2}V_{zy}W_{yz}\} = \beta^{-2}\,\mathrm{trace}\{(V^{zz})^{-1}V_{zz} - (V^{zz})^{-2}\}.$$

The second term here can be evaluated as before, using the sum of squares of the elements of $(V^{zz})^{-1}$ for small m. The first term involves V_{zz}, as stated above. For small m, exact formulæ can be found (Martin, 1989a). Again, approximate formulæ can be derived—see Martin (1989a).

Also, these formulæ can be extended to larger m, and to other processes. Although the mathematics is interesting, I feel that further theory should be justified by practical needs. Which models are reasonable for a given application needs to be discovered, as well as why it is of interest to know the loss in information.

6. Conclusion

I have given a personal view of some of the spatial statistical models used in geography, and of some of the publications concerning these models. I hope that the papers in this volume will lead to an improvement in modelling, and in published research. If geographers stimulate statisticians by presenting problems of practical interest, then valuable joint research should result.

If my comments have been unduly negative, I should say that I have been heartened by the apparent willingness with which geographers accept criticism of their mistakes, although I would prefer that the mistakes were not made. I should also emphasize that similar comments could be made about workers in other disciplines, or even within the statistical community. I have tried to ensure there are no mathematical or statistical errors in this paper, and will endeavour to correct any that I notice subsequently or that are bought to my attention.

7. References

Bennett, R. J. and Haining, R. P. (1985) Spatial structure and spatial interaction: modelling approaches to the statistical analysis of geographical data (with discussion). *Journal of the Royal Statistical Society A* **148**: 1-36.

Bennett, R. J., Haining, R. P. and Griffith, D. A. (1984) The problem of missing data on spatial surfaces. *Annals of the Association of American Geographers* **74**: 138-56.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B* **36**: 192-236.

Besag, J. (1975) Statistical analysis of non–lattice data. *The Statistician* **24**: 179-95.

Besag, J. (1977a) Errors–in–variables estimation for Gaussian lattice schemes. *Journal of the Royal Statistical Society B* **39**: 73-8.

Besag, J. (1977b) Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika* **64**: 616-8.

Besag, J. (1985) Contribution to the discussion of a paper by R. J. Bennett and R. P. Haining. *Journal of the Royal Statistical Society A* **148**: 31.

Besag, J. and Moran, P. A. P. (1975) On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrika* **62**: 555-62.

Brandsma, A. S. and Ketellapper, R. H. (1979) A biparametric approach to spatial autocorrelation. *Environment and Planning A* **11**: 51-8.

Bruce, A. G. and Martin, R. D. (1989) Leave–k–out diagnostics for time series (with discussion). *Journal of the Royal Statistical Society B* **51**: 363-424.

Cliff A. D. and Ord, J. K. (1975) Model building and the analysis of spatial pattern in human geography (with discussion). *Journal of the Royal Statistical Society B* **37**: 297-348.

Cliff A. D. and Ord, J. K. (1981) *Spatial Processes*. Pion: London.

Gasim, A. A. (1988) First–order autoregressive models: a method for obtaining eigenvalues for weighting matrices. *Journal of Statistical Planning and Inference* **18**: 391-8.

Griffith, D. A. (1980) Towards a theory of spatial statistics. *Geographical Analysis* **12**: 325-39.

Griffith, D. A. (1983) The boundary value problem in spatial statistical analysis. *Journal of Regional Science* **23**: 377-87.

Griffith, D. A. (1985) An evaluation of correction techniques for boundary effects in spatial statistical analysis: contemporary methods. *Geographical Analysis* **17**: 81-8.

Griffith, D. A. (1987) Towards a theory of spatial statistics: another step forward. *Geographical Analysis* **19**: 69-82.

Griffith, D. A. (1988) A reply to R. Martin's "Some comments on correction techniques for boundary effects and missing value techniques". *Geographical Analysis* **20**: 70-5. Correction (1989) **21**: 359.

Griffith, D. A. and Amrhein, C. G. (1983) An evaluation of correction techniques for boundary effects in spatial statistical analysis: traditional methods. *Geographical Analysis* **15**: 352-60.

Griffith, D. A., Haining, R. P., and Bennett R. J. (1985) Estimating missing values in space–time data series. In *Time Series Analysis: Theory and Practice 6*, Eds. O. D. Anderson, J. K. Ord and E. A. Robinson, pp. 273-82. North Holland : Amsterdam.

Guyon, X. (1982) Parameter estimation for a stationary process on a d–dimensional lattice. *Biometrika* **69**: 95-105.

Haining, R. (1987) Trend–surface models with regional and local scales of variation with an application to aerial survey data. *Technometrics* **29**: 461-9.

Haining, R. (1988) Estimating spatial means with an application to remotely sensed data. *Communications in Statistics—Theory and Methods* **17**: 573-97.

Haining, R., Griffith, D. A. and Bennett R. J. (1983) Simulating two–dimensional autocorrelated surfaces. *Geographical Analysis* **15**: 247-55.

Haining, R., Griffith, D. A. and Bennett R. J. (1984) A statistical approach to the problem of missing spatial data using a first–order Markov model. *Professional Geographer* **36**: 338-45.

Haining, R., Griffith, D. A. and Bennett R. J. (1989) Maximum likelihood estimation with missing spatial data and with an application to remotely sensed data. *Communications in Statistics—Theory and Methods* **18**: 1875-94.

Journel, A. G. and Huijbregts, Ch. (1978) *Mining Geostatistics*. Academic Press: London.

Kunert, J. and Martin, R. J. (1987) Some results on optimal design under a first–order autoregression and on finite Williams type II designs. *Communications in Statistics—Theory and Methods* **16**: 1901-22.

Künsch, H. (1987) Intrinsic autoregressions and related models on the lattice Z^2. *Biometrika* **74**: 517-24.

Mardia, K. V. and Marshall, R. J. (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**: 135-46.

Martin, R. J. (1983) A review of continuous valued spatial processes. Paper presented at the Institute of British Geographers Quantitative Methods Study Group Conference on Advances in Applied Spatial Analysis, Sheffield.

Martin, R. J. (1984) Exact maximum likelihood for incomplete data from a correlated Gaussian process. *Communications in Statistics—Theory and Methods* **13**: 1275-88.

Martin, R. J. (1985) Contribution to the discussion of a paper by R. J. Bennett and R. P. Haining. *Journal of the Royal Statistical Society A* **148**: 29-30.

Martin, R. J. (1986) A note on the asymptotic eigenvalues and eigenvectors of the dispersion matrix of a second–order stationary process on a d–dimensional lattice. *Journal of Applied Probability* **23**: 529-35.

Martin, R. J. (1987) Some comments on correction techniques for boundary effects and missing value techniques. *Geographical Analysis* **19**: 273-82.

Martin, R. J. (1989a) Information loss due to incomplete data from a spatial Gaussian one-parameter first-order conditional process. *Communications in Statistics—Theory and Methods* **18**: 4631-45.

Martin, R. J. (1989b) Leverage, influence and residuals when errors are correlated. Research report no. 326/89, Department of Probability and Statistics, University of Sheffield.

Martin, R. J. (1989c) Contribution to the discussion of a paper by A. G. Bruce and R. D. Martin. *Journal of the Royal Statistical Society B* **51**: 414.

Martin, R. J. (1990) The use of time-series models in the analysis of agricultural field trials. *Communications in Statistics—Theory and Methods* **19**: 55-81.

Orchard, T. and Woodbury, M. A. (1972) A missing information principle: theory and applications. In *Proceedings of the 6th Berkeley Symposium 1*, Eds. L. M. Le Cam, J. Neyman and E. L. Scott, pp. 697-715. University of California Press: Berkeley.

Ord, J. K. (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* **70**: 120-6.

Ord, J. K. (1981) Towards a theory of spatial statistics: a comment. *Geographical Analysis* **13**: 86-93.

Patankar, V. N. (1954) The goodness of fit of frequency distributions obtained from stochastic processes. *Biometrika* **41**: 450-62.

Ripley, B. D. (1988) *Statistical inference for spatial processes.* Cambridge University Press: Cambridge.

Whittle, P. (1954) On stationary processes in the plane. *Biometrika* **41**: 434-49.

DISCUSSION

"The role of spatial statistical processes in geographic modelling"

by R. J. Martin

Statistical models of spatial dependence have been used quite commonly in geographic research. In his presentation, the author both reviews and comments on their use. He further takes up the topics of boundary effects and missing values, attempting to clarify the former and giving some new results on the latter.

The paper starts with a substantial section (Section 2) on models that includes mathematical details on their fitting as well as the author's view on how a modelling exercise should be justified. Computations for fitting first–order models are thoroughly discussed, and the author gives convenient, simple forms for the Fisher information matrix of the parameters, both for the conditional and the simultaneous versions. Then restrictions of first–order models are developed, leading to a review of selected extensions, still using contiguity matrices, which would allow some form of non–isotropy for the dependence or an increase of its range.

The author omits from his review a class of models where the covariance between sites $\mathbf{i}$ and $\mathbf{j}$ is not modelled through arbitrarily defined contiguity matrices, but rather has a parametrised functional form. This class of models rarely has been used in geographical studies, although it has received attention in the statistical, epidemiological and geostatistical literature (Ripley, 1988; Cook and Pocock, 1983; Mardia and Marshall, 1984; Vecchia, 1988); it would be interesting to see applications of this model in geography.

Section 2 starts and ends with some methodological considerations about justification and comparison of models. This is certainly an important area that, until now, has not received enough attention, and the author's emphasis and suggestions are most welcome. I would add that the strategy used to justify or compare different models depends upon whether the aim of the modelling exercise is explanatory, for forecasting purposes, or to be used in a generalised regression framework.

Section 3 recounts some of the arguments that have arisen between the author and geographers concerning the application of statistics. Although part of this section may be difficult to follow for a reader who does not have all of the quoted papers on hand, the author develops a convincing case on the desirability of constructive discussions between statisticians and geographers that should benefit both professions. It is in everyone's interest to avoid incorrect uses of statistics. Discussions of this kind often stimulate new research.

Section 4 is of a general nature and argues for a precise definition of what is called the boundary value problem, whether it influences the dispersion matrices or the bias in estimators. In contrast, the final section gives some results on the loss of information due to missing values on the mean μ and the parameter β of the first–order conditional or simultaneous process. Since the author wanted to expand on some new results, this section is the least self–contained. Useful approximations for the loss of information are given when the number of missing sites becomes large.

Richardson on Martin

In this paper the author presents original and thoughtful considerations on the use of spatial statistics in geography, emphasising throughout the need to link theoretical developments (like those arising for missing values) to relevant examples, and to relate models to geographical problems.

References

Cook, D., and S. Pocock. (1983) Multiple regression in geographic mortality studies with allowance for spatially correlated errors. *Biometrics*, **39**: 361-371.

Mardia, K., and R. Marshall. (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **71**: 135-146.

Ripley, B. (1988) *Statistical inference for spatial processes*. Cambridge: Cambridge University Press.

Vecchia, A. (1988) Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society*, **50B**: 297-312.

Sylvia Richardson, INSERM