Biometrika (1995), **82,** 4, *pp.* 733–46
Printed in Great Britain

On conditional and intrinsic autoregressions

BY JULIAN BESAG AND CHARLES KOOPERBERG
Department of Statistics, University of Washington, Box 354322, Seattle, Washington 98195-4322, U.S.A.

SUMMARY

Gaussian conditional autoregressions have been widely used in spatial statistics and Bayesian image analysis, where they are intended to describe interactions between random variables at fixed sites in Euclidean space. The main appeal of these distributions is in the Markovian interpretation of their full conditionals. Intrinsic autoregressions are limiting forms that retain the Markov property. Despite being improper, they can have advantages over the standard autoregressions, both conceptually and in practice. For example, they often avoid difficulties in parameter estimation, without apparent loss, or exhibit appealing invariances, as in texture analysis. However, on small arrays and in nonlattice applications, both forms of autoregression can lead to undesirable second-order characteristics, either in the variables themselves or in contrasts among them. This paper discusses standard and intrinsic autoregressions and describes how the problems that arise can be alleviated using Dempster's (1972) algorithm or an appropriate modification. The approach represents a partial synthesis of standard geostatistical and Gaussian Markov random field formulations. Some nonspatial applications are also mentioned.

Some key words: Agricultural experiments; Bayesian image analysis; Conditional autoregression; Dempster's algorithm; Geographical epidemiology; Geostatistics; Intrinsic autoregression; Multi-way table; Prior distribution; Spatial statistics; Surface reconstruction; Texture analysis.

1. INTRODUCTION

Let $X=(X_1, \ldots, X_n)^{\mathrm{T}}$ denote a random vector for which a joint distribution $p(x)$ is to be constructed. There are many practical applications in which each component X_i is located at a fixed site i, and where the relative arrangement of sites conveys useful information. Such considerations apply in spatial statistics and image analysis, where the sites represent regularly or irregularly distributed points or regions in Euclidean space, but they also arise in many other contexts: for example, in multi-way tables with several factors at equally-spaced levels, both for the model describing dependent responses and for the prior distribution for factor effects or interactions.

In the Markov random field approach to such modelling, $p(x)$ is constructed via the n full conditional distributions, or local characteristics, $p(x_i|x_{-i})$, where $x_{-i}:=\{x_j: j \neq i\}$. If $p(x_i|x_{-i})$ depends on x_j, then j is called a neighbour of site i. The usual first step is to postulate a relatively small set of neighbours ∂i for each site i, and then to choose corresponding mutually compatible full conditional distributions. Under a positivity condition, which can be relaxed somewhat, these determine $p(x)$ through the Brook expansion (Besag, 1974, eqn (2.2)). There are strong links here with conditions for the validity of single component Gibbs samplers for $p(x)$ (Besag, 1994).

For the remainder of the paper, we shall be concerned only with Gaussian specifications.

Also we assume that the graph induced by the neighbour relation is connected, since otherwise the system can be broken down into its connected subsystems. In § 2, we first provide a summary of the usual conditional autoregressions of spatial statistics, which focus on specifying the precision matrix rather than the dispersion matrix of X. However, these models typically produce quite different marginal variances, which is generally undesirable. We show that the problem can be alleviated using an algorithm given by Dempster (1972). The section concludes by discussing regular lattice systems and provides a corresponding numerical example of the algorithm.

Another common disadvantage of conditional autoregressions is that appreciable correlations between the X_i's at neighbouring sites require parameter values extremely close to a particular boundary of the parameter space. In § 3, we turn this to advantage by considering intrinsic limits of conditional autoregressions. Although $p(x)$ is then improper, the distributions of certain, usually all, contrasts among the X_i's are well defined. In the latter case, we show that Dempster's algorithm can be modified so as to ensure that $X_i - X_j$ has constant variance when i and j are neighbours. Again we discuss some aspects of the problem for regular arrays and give examples.

In geostatistics, e.g. Cressie (1991, Ch. 2), specifications are usually obtained by direct modelling of the dispersion matrix or, when this does not exist, the semivariogram. Our approach represents a partial synthesis of the geostatistical and Markov random field formulations.

2. GAUSSIAN CONDITIONAL AUTOREGRESSIONS

2·1. *General formulation*

Suppose that the random vector $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ has density

$$p(x) \propto e^{-\frac{1}{2}x^{\mathrm{T}}Qx}, \quad x \in R^n, \tag{2·1}$$

where Q is a $n \times n$ positive definite symmetric matrix. Then

$$X_i | x_{-i} \sim N\left(\sum_j \beta_{ij} x_j, \kappa_i\right), \tag{2·2}$$

where $\beta_{ii} = 0$, $\beta_{ij} = -Q_{ij}/Q_{ii}$ $(i \neq j)$ and $\kappa_i = 1/Q_{ii} > 0$. The symmetry of Q requires that

$$\beta_{ij}\kappa_j = \beta_{ji}\kappa_i. \tag{2·3}$$

Note that i and j are neighbours if and only if $\beta_{ij} \neq 0$, in which case we write $i \sim j$, and that $\mathrm{sgn}(\beta_{ij})\sqrt{(\beta_{ij}\beta_{ji})}$ is the partial correlation coefficient between X_i and X_j. Of course, x on the right-hand side of (2·1) can be replaced by $x - \mu$, where μ is an arbitrary real n-vector, with corresponding adjustment to (2·2).

The reverse route from (2·2) to (2·1), with given β_{ij}'s and κ_i's satisfying (2·3), is less obvious but follows from the Brook expansion for $p(x)/p(0)$. Positive definiteness of Q may need to be checked on an individual basis but the identity,

$$x^{\mathrm{T}}Qx \equiv \sum_i Q_{i+}x_i^2 - \sum_{i<j} Q_{ij}(x_i - x_j)^2, \tag{2·4}$$

where subscripts $+$ denote summation over replaced indices, implies that a sufficient condition is that the β_{ij}'s are all nonnegative and $\beta_{i+} \leqslant 1$ for all i, with strict inequality for at least one i. When the specification of $p(x)$ is based on (2·2) and hence on the precision matrix Q, rather than on the dispersion matrix $V = Q^{-1}$, it is usually referred to as a conditional autoregressive or auto-Normal formulation (Besag, 1974).

Gaussian conditional autoregressions have been used in a wide range of applications: human geography, e.g. Cliff & Ord (1975; 1981, Ch. 4); agricultural field experiments, e.g. Bartlett (1978 including the Discussion), Kempton & Howes (1981), Martin (1990); geographical epidemiology, e.g. Clayton & Kaldor (1987), A. Mollié in the University of Paris Ph. D. thesis 'Représentation géographique des taux de mortalité: modélisation spatiale et méthodes Bayésiennes', Cressie (1991, Ch. 7), Marshall (1991), Mollié & Richardson (1991), Bernardinelli & Montomoli (1992); astronomy, e.g. Molina & Ripley (1989), Ripley (1991); texture analysis e.g. Chellappa & Kashyap (1985), Cohen, Fan & Patel (1991), Cohen & Patel (1991); and other forms of image processing, e.g. Jinchi & Chellappa (1986), Cohen & Cooper (1987), Simchony, Chellappa & Lichtenstein (1989), Zerubia & Chellappa (1990). Here, we consider the simplest case of practical interest.

Example 2·1. Given the neighbours ∂i of each site i, suppose that the conditional mean in (2·2) is $\lambda\bar{x}_i$, where $\lambda \in (0, 1)$ and $\bar{x}_i$ is the mean of the x_j's, $j \in \partial i$. Connectedness and (2·3) together imply that $\mathrm{var}(X_i \mid x_{-i}) = \kappa/n_i$ for some $\kappa > 0$, where n_i is the cardinality of ∂i.

Even in this very simple example, an unsatisfactory feature is present, for it is not possible for the X_i's all to have the same marginal variance, nor for all neighbour pairs to have the same covariance, unless the corresponding graph has a very special structure. This suggests an alternative strategy. We first choose a neighbourhood criterion, as before, but then fix, perhaps empirically, the V_{ij}'s for $i = j$ and for $i \sim j$. Thus, we seek a positive definite matrix Q such that, for each i and j, either $Q_{ij} = 0$ or the value of $(Q^{-1})_{ij}$ is specified.

2·2. *Dempster's algorithm*

As one aspect of a very wide-ranging paper, Dempster (1972) proves that, if the above matrix Q exists, it is unique and can be found using a Newton–Raphson algorithm, with possible step reduction. We give a numerical example in § 2·3.

THEOREM 2·1 (*Dempster*, 1972). *Define I_0 to be the set of all $\frac{1}{2}n(n+1)$ index pairs (i, j) for $1 \leqslant i \leqslant j \leqslant n$, and let (I_1, I_2) be particular ordered partition of I_0, with I_1 having m elements, say. Let V^* and Q^* be fixed $n \times n$ matrices. If there exist symmetric positive definite matrices Q and V such that:*

(*i*) $Q = V^{-1}$;

(*ii*) $V_{ij} = V^*_{ij}$ *for* $(i, j) \in I_1$;

(*iii*) $Q_{ij} = Q^*_{ij}$ *for* $(i, j) \in I_2$;

then Q and V are unique.

Algorithm 2·1 (*Dempster*, 1972). Suppose Q and V exist. For any symmetric $n \times n$ matrix B, let $\theta_1(B)$ be the row vector whose kth element is B_{ij} if $i \neq j$, or $\frac{1}{2}B_{ii}$ if $i = j$, where (i, j) is the kth element of I_1. Define $\theta_2(B)$ correspondingly with respect to I_2 and write $\theta(B) = (\theta_1(B), \theta_2(B))$. Let $Q^{(0)}$ be any symmetric positive definite matrix having property (iii) above. Then the following steps for $l = 0, 1, \ldots$ will converge to Q and V:

(a) set $V^{(l)} = (Q^{(l)})^{-1}$;

(b) form the $m \times m$ matrix $C^{(l)}$ with (k_0, k_1) element

$$C^{(l)}_{k_0k_1} = \begin{cases} \frac{1}{4}(V^{(l)}_{i_0i_1} V^{(l)}_{j_0j_1} + V^{(l)}_{i_0j_1} V^{(l)}_{j_0i_1}) & \text{if } i_0 = j_0,\ i_1 = j_1, \\ \frac{1}{2}(V^{(l)}_{i_0i_1} V^{(l)}_{j_0j_1} + V^{(l)}_{i_0j_1} V^{(l)}_{j_0i_1}) & \text{if } i_0 = j_0,\ i_1 \neq j_1 \text{ or } i_0 \neq j_0,\ i_1 = j_1, \\ V^{(l)}_{i_0i_1} V^{(l)}_{j_0j_1} + V^{(l)}_{i_0j_1} V^{(l)}_{j_0i_1} & \text{if } i_0 \neq j_0,\ i_1 \neq j_1, \end{cases}$$

where (i_0, j_0) and (i_1, j_1) are the k_0th and k_1th elements of I_1;

(c) obtain $Q^{(l+1)}$ by

$$\theta_1(Q^{(l+1)}) = \theta_1(Q^{(l)}) + \{\theta_1(V^*) - \theta_1(V^{(l)})\}(C^{(l)})^{-1},$$
$$\theta_2(Q^{(l+1)}) = \theta_2(Q^{(l)}).$$

2·3. *Regular arrays*

Practical applications of Gaussian conditional autoregressions often involve random variables distributed on a regular lattice. Examples include image analysis, where sites represent pixels, and crop experiments, where they equate with plots in the field. A two-way table with responses that are the aggregate of independent Gaussian row effects, column effects and noise is an autoregression with conditional mean at each site that depends on the average responses at sites in the same row, at sites in the same column, and at sites elsewhere; in some applications, more localised autoregressive formulations would be of interest.

The second-order properties of lattice processes are often interpreted in terms of finite restrictions of stationary autoregressions on corresponding infinite arrays. The latter are most elegantly studied through their spectral densities, or equivalently their autocovariance generating functions. Here we provide a brief review for the two-dimensional rectangular lattice; see Lévy (1948), Whittle (1954), Rosanov (1967), Moran (1973), Besag (1974), Künsch (1987), Cressie (1991, Ch. 6) and Guyon (1992, Ch. 1) for further details. Generalisations to multivariate site variables are discussed by Kittler & Föglein (1984) and Mardia (1988) in the context of multi-spectral imaging.

There is a simple relationship between the conditional formulation of a stationary autoregression and its autocovariance generating function. Thus, let $i = (u, v)$, for $u, v = 0, \pm 1, \ldots$, denote the sites of an infinite rectangular lattice, and suppose that $\{X_{uv}\}$ is a stationary Gaussian process with conditional moments, as in (2·2),

$$E(X_{uv}|\ldots) = \sum_k \sum_l \eta_{kl} x_{u-k,v-l}, \quad \text{var}(X_{uv}|\ldots) = \kappa > 0, \tag{2·5}$$

with the conditioning being on all other variables, where

(i) $\eta_{00} = 0$,
(ii) the number of nonzero η_{kl}'s is finite,
(iii) $\eta_{kl} = \eta_{-k-l}$,
(iv) $\sum\sum \eta_{kl} \cos(\omega_1 k + \omega_2 l) < 1$ for all ω_1 and ω_2.

Condition (iii) replaces (2·3), and (iv) that of positive definiteness.

Now define $\gamma_{rs} = E(X_{00}X_{rs})$ and $\rho_{rs} = \gamma_{rs}/\gamma_{00}$ to be the autocovariance and autocorrelation of lag(r, s), respectively. The equation (2·5) implies that

$$\gamma_{rs} = \kappa\delta_{rs} + \sum_k \sum_l \eta_{kl}\gamma_{r+k,s+l} \quad (r, s = 0, \pm 1, \ldots), \tag{2·6}$$

where $\delta_{00} = 1$ and $\delta_{rs} = 0$ otherwise, and the autocovariance generating function of $\{X_{uv}\}$ is

$$\Gamma(z_1, z_2) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \gamma_{rs} z_1^r z_2^s = \kappa \Big/ \Big(1 - \sum_k \sum_l \eta_{kl} z_1^k z_2^l\Big), \tag{2·7}$$

from (2·6), establishing the simple relationship between (2·5) and $\Gamma(z_1, z_2)$. The spectral density of $\{X_{uv}\}$ is proportional to $\Gamma(e^{-i\omega_1}, e^{-i\omega_2})$, and yields the inversion formula for

the autocovariances:

$$\gamma_{rs} = \frac{\kappa}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\cos(\omega_1 r + \omega_2 s)\, d\omega_1\, d\omega_2}{1 - \sum\sum \eta_{kl} \cos(\omega_1 k + \omega_2 l)}; \tag{2·8}$$

integrability is ensured by (iv). Knowledge of the spectral density also underlies a simple method of simulating a stationary autoregression over an arbitrary finite set of sites, e.g. Cressie (1991, p. 203). Results for processes on higher-dimensional lattices are similar to those above, except that strict inequality in the analogue of (iv) can be relaxed.

In order to obtain the restriction of a particular infinite lattice autoregression to a finite array, it remains to identify the conditional means and variances at its boundary B; that is, at sites that have missing neighbours with respect to the infinite system. Unfortunately, except in special cases such as the separable processes of Martin (1979), the neighbours of any boundary site include all other sites in B. This unwieldiness aside, the problem can be solved in principle by using (2·8) to calculate all γ_{rs} relevant to the finite array and then inverting the corresponding V to obtain Q. However, numerical integration of (2·8) is usually exceedingly delicate because the moderate or substantial autocorrelations that are typical in practical applications occur only when η_{++} is extremely close to unity. We shall turn this to advantage in § 3 but here it remains problematical, with published results available only for the first-order autoregression, for which $\eta_{10} = \eta_{-10}$ and $\eta_{01} = \eta_{0-1}$ are the sole nonzero coefficients in (2·8): see Besag (1981) and, for extensions to second-order processes, unpublished work by R. Roberts and J. Besag.

The above difficulties have led to the use of various boundary approximations. It should be emphasised that, in large-scale applications where the region of interest is well removed from the boundary, the choice of an approximation is of secondary importance and one should usually settle for the most convenient one. However, as we illustrate below, this is unlikely to be the case for small arrays where all the data need to be used and, for example, conditioning on the boundary values is inappropriate.

Example 2·2. The simplest version of (2·5) is the symmetric first-order autoregression, for which

$$E(X_{uv}|\ldots) = \tfrac{1}{4}\lambda(x_{u-1v} + x_{u+1v} + x_{uv-1} + x_{uv+1}),$$

with $|\lambda| < 1$. Here, we take $\gamma_{00} = 1$ and neighbour autocorrelations $\rho_{10} = \rho_{01} = 0{\cdot}75$, which requires $\lambda \doteq 0{\cdot}999972$. Suppose we are concerned with the restriction of the process to a 10×10 array. The $(\rho_{rs},\ 0 \leqslant r \leqslant s \leqslant 9)$ needed to construct the correct Q are given in the middle rows of Table 1: compare with Table 3 of Besag (1981). In the approximations below, λ and κ are chosen so that at least the variables in the central 2×2 block of the array have variance unity and neighbour correlation 0·75.

One common boundary approximation is to replace each missing variable in (2·5) by some 'typical value', which here would be zero. This shrinks the conditional expectations at boundary sites towards zero and hence decreases the corresponding marginal variances. In the example, we find $\lambda \doteq 1{\cdot}0463$; variances range from 0·31 to 1, and neighbour covariances and correlations from 0·12 to 0·75 and from 0·35 to 0·75, respectively. Note that the value of λ is allowable here, although not of course on the infinite lattice; the original value gives correlations that are much too small.

Another suggestion is to ignore the missing values and to re-scale the coefficients at boundary sites in some appropriate fashion. Here, this is unambiguous and leads to the autoregression in Example 2·1, with $\lambda \doteq 0{\cdot}9954$. The corresponding variances range from

Table 1. *Example 2·2: Minimum (bottom) and maximum (top) correlations of* lag(r, s), *for* $r, s = 0, 1, \ldots 9$, *when Dempster's algorithm is used to approximate the symmetric first-order autoregression, with autocorrelations* ρ_{rs} *(middle) and* $\rho_{01} = \rho_{10} = 0{\cdot}75$

r	$s=0$	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$	$s=8$	$s=9$
0	1·000	0·750	0·637	0·562	0·510	0·467	0·429	0·396	0·365	0·337
	1·000	0·750	0·637	0·570	0·523	0·487	0·458	0·434	0·413	0·394
	1·000	0·750	0·629	0·546	0·478	0·423	0·378	0·340	0·306	0·275
1	0·750	0·681	0·610	0·551	0·503	0·463	0·426	0·394	0·364	0·336
	0·750	0·682	0·613	0·560	0·518	0·484	0·456	0·432	0·412	0·393
	0·750	0·676	0·604	0·536	0·478	0·429	0·387	0·350	0·318	0·288
2	0·637	0·610	0·569	0·527	0·487	0·451	0·417	0·386	0·357	0·330
	0·637	0·613	0·576	0·538	0·504	0·475	0·450	0·428	0·408	0·390
	0·629	0·604	0·559	0·510	0·465	0·423	0·386	0·353	0·323	0·295
3	0·562	0·551	0·527	0·497	0·465	0·434	0·404	0·375	0·347	0·322
	0·570	0·560	0·538	0·512	0·486	0·462	0·440	0·420	0·402	0·386
	0·546	0·536	0·510	0·478	0·443	0·410	0·378	0·349	0·322	0·296
4	0·510	0·503	0·487	0·465	0·439	0·413	0·386	0·360	0·334	0·309
	0·523	0·518	0·504	0·486	0·467	0·447	0·429	0·411	0·395	0·380
	0·478	0·478	0·465	0·443	0·418	0·392	0·366	0·340	0·316	0·292
5	0·467	0·463	0·451	0·434	0·413	0·391	0·367	0·343	0·319	0·296
	0·487	0·484	0·475	0·462	0·447	0·432	0·416	0·401	0·387	0·373
	0·423	0·429	0·423	0·410	0·392	0·371	0·350	0·328	0·306	0·284
6	0·429	0·426	0·417	0·404	0·386	0·367	0·345	0·324	0·301	0·279
	0·458	0·456	0·450	0·440	0·429	0·416	0·403	0·390	0·378	0·365
	0·378	0·387	0·386	0·378	0·366	0·350	0·332	0·313	0·294	0·273
7	0·396	0·394	0·386	0·375	0·360	0·343	0·324	0·304	0·282	0·262
	0·434	0·432	0·428	0·420	0·411	0·401	0·390	0·379	0·368	0·357
	0·340	0·350	0·353	0·349	0·340	0·328	0·313	0·297	0·279	0·260
8	0·365	0·364	0·357	0·347	0·334	0·319	0·301	0·282	0·263	0·243
	0·413	0·412	0·408	0·402	0·395	0·387	0·378	0·368	0·358	0·349
	0·306	0·318	0·323	0·322	0·316	0·306	0·294	0·279	0·262	0·243
9	0·337	0·336	0·330	0·322	0·309	0·296	0·279	0·262	0·243	0·226
	0·394	0·393	0·390	0·386	0·380	0·373	0·365	0·357	0·349	0·340
	0·275	0·288	0·295	0·296	0·292	0·284	0·273	0·260	0·243	0·226

1 to 1·82, and the neighbour covariances and correlations from 0·75 to 1·33 and from 0·75 to 0·81, respectively. This approximation has the opposite effect to the previous one and is therefore perhaps more satisfactory.

A third option is to impose periodic boundary conditions, identifying opposite edges of the finite lattice. This has the advantage that the awkward inversion formula (2·8) is replaced by a finite summation (Besag & Moran, 1975; Besag, 1977). We now find that $\lambda \doteq 0{\cdot}9957$, with all variances and neighbour covariances being correctly matched. However, the conditional distributions for variables at the boundary are highly erroneous and lead to X_{uv}'s at opposite ends of rows and columns also having correlation 0·75. Because of their computational simplicity, periodic boundary conditions have been especially popular in large-scale image analysis, but they are inappropriate for small arrays, such as those usually met in agricultural experiments.

Lastly, we examine how our new strategy fares. We maintain the simple neighbourhoods

for interior sites but, for each boundary site, also include sites on the same edge that are distance two away. Thus, corner and next-to-corner sites also have four neighbours but the remaining boundary sites have five. The construction via Dempster's algorithm ensures that all variances and neighbour correlations are matched correctly. Otherwise, the worst discrepancies are shown in Table 1, where the top and bottom rows provide the maximum and minimum correlations for each lag. If required, an improved approximation could be obtained by adding further to the neighbourhoods of the boundary sites.

When empirical autocovariances are used, the data tapers of Dahlhaus & Künsch (1987) ensure positive definiteness, without sacrificing the consistency of unbiased estimation (Guyon, 1982), although consistency is of questionable relevance in the present small-sample context. Several authors have commented on the close agreement between the numerical estimates of autoregressive parameters given by Besag (1974) and their own, technically more correct, versions. The explanation is that, somewhat fortuitously, Besag adopted less fashionable, but here more appropriate, unbiased autocovariances in his calculations. For detailed discussion of parameter estimation and more on boundary effects, see also Besag (1981), Künsch (1983), Ripley (1988, Ch. 2), Cressie (1991, Ch. 6, 7) and Guyon (1992, Ch. 4).

3. Gaussian intrinsic autoregressions

3·1. *General formulation*

It follows from (2·4) that, for the conditional autoregression (2·2), there is a boundary in the parameter space where $Q_{i+} = 0$ or, equivalently, $\beta_{i+} = 1$, for all i. For the stationary autoregression (2·5), the corresponding boundary is defined by $\eta_{++} = 1$. Moreover, it is very often the case that no appreciable correlations occur unless the parameter values are extremely close to these edges. We saw this for a simple lattice process in Example 2·2 and in the discussion that followed it. In image analysis, Cohen et al. (1991) use the method of maximum likelihood to estimate the parameters of nine different textures on a torus lattice and obtain values of β_{i+} between 0·99484 and 0·99991. Indeed, alternative estimators that do not take account of the determinant in the constant of proportionality in (2·1) frequently produce invalid results. This holds both for coding (Besag, 1974) and for pseudolikelihood (Besag, 1975) estimators, despite their asymptotic properties (Besag & Moran, 1975; Besag, 1977; Geman & Graffigne, 1987).

Such considerations suggest we should examine limiting forms of standard Gaussian conditional autoregressions, in which Q is well defined but $Q1 = 0$, where 1 and 0 denote appropriate vectors of 1's and 0's. We refer to such processes as intrinsic autoregressions, following the infinite lattice terminology of Künsch (1987), which itself borrows from closely related ideas in the geostatistics literature, e.g. Matheron (1973). Equations (2·2) and (2·3) remain valid but positive definiteness of Q must now be replaced by positive semi-definiteness and, of course, Q^{-1} no longer exists. Formally, we can write a pairwise-difference 'density' for X, 'proportional' to

$$\exp\left\{\frac{1}{2}\sum_{i<j} Q_{ij}(x_i - x_j)^2\right\}. \tag{3·1}$$

This is required in Bayesian applications, where (3·1) may approximate prior beliefs about a true x, and leads to a proper posterior distribution, subject to the usual care in hierarchical formulations; see the comment following Example 3·1. In the Bayesian context, an equivalent representation is obtained if X_i is replaced by $\mu + X_i$, where μ has a 'uniform'

vague prior and the X_i's have density (3·1), constrained to have mean $\bar{X} = 0$. The latter is a proper $(n-1)$-dimensional density provided Q has rank $n-1$. Note that equation (2·2) becomes a stochastic interpolation rule, which in many practical contexts has independent appeal. We return to this interpretation in § 3·3.

Marginally, intrinsic autoregressions have undefined means and infinite variances, but usually all contrasts $c^{\mathrm{T}}X$, where c is nonnull and $c^{\mathrm{T}}1 = 0$, have proper distributions. The simplest example is that of a one-dimensional random walk, with arbitrary level and independent, identically distributed, Gaussian increments. This process has been used for one-dimensional fertility adjustment in the analysis of agricultural field experiments, from both frequentist and Bayesian perspectives (Besag & Kempton, 1986; Besag & Higdon, 1993; Besag et al., 1995). However, there is not generally an interpretation of intrinsic autoregressions in terms of independent increments. See Künsch (1987) for a thorough discussion of such issues with regard to infinite lattices. Below, we collect some useful facts for finite n, in the form of a lemma and two corollaries.

LEMMA 3·1. *Let $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ denote a Gaussian vector with positive semi-definite precision matrix Q of rank k, where $1 \leqslant k \leqslant n-1$. Let A denote a $k \times n$ matrix whose rows span the same k-dimensional subspace of R^n as the rows of Q. Define $Y = AX$. Then*

(*i*) *Y has a nonsingular Gaussian distribution with precision matrix $Q_Y = \bar{A}^{\mathrm{T}} Q \bar{A}$, where $\bar{A}$ is any generalised inverse of A;*

(*ii*) *$Q = A^{\mathrm{T}} Q_Y A$;*

(*iii*) *$V = \bar{A} Q_Y^{-1} \bar{A}^{\mathrm{T}}$ is a generalised inverse of Q.*

The proof is standard matrix algebra.

COROLLARY 3·1. *If CX has a proper Gaussian distribution, then its density coincides with that of $C\bar{A}Y$, so that the well-determined second-order properties of X match those of the vector $\bar{A}Y$.*

When k is not too large for the applicability of matrix methods, the corollary provides a simple method of simulation, particularly useful if the only indeterminacy is the overall level, as below.

COROLLARY 3·2. *Let $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ denote a Gaussian intrinsic autoregression with precision matrix Q of rank $n-1$, so that the sole redundancy in Q is $Q1 = 0$. Then any linearly independent set of contrasts among the X_i's has a proper Gaussian distribution*

Proof. The method is again standard but we supply explicit results for later use. Let $Y_i = X_i - X_n$ $(i = 1, \ldots, n-1)$ so that $A = (I \mid -1)$ and we can take $\bar{A} = (I \mid 0)^{\mathrm{T}}$. Then Q_Y is the upper left $(n-1) \times (n-1)$ submatrix of Q, so Y, and hence any linearly independent set of contrasts among the X_i's, has a proper Gaussian distribution. □

We note that the Y_i's in the proof satisfy

$$E(Y_i \mid y_{-i}) = \sum_{j=1}^{n-1} \beta_{ij} y_j, \quad \mathrm{var}(Y_i \mid y_{-i}) = \kappa_i.$$

Indeed, this follows directly from the conditional means and variances in (2·2) and does not require the X_i's to be conditionally Gaussian.

Example 3·1. Let $\lambda = 1$ in Example 2·1, so that $E(X_i \mid x_{-i}) = \bar{x}_i$, $\mathrm{var}(X_i \mid x_{-i}) = \kappa/n_i$, and Q has rank $n-1$.

The above distribution has been used in geographical epidemiology to represent prior

beliefs about the spatial component of log relative risk from a rare disease in each of n contiguous regions (Besag, York & Mollié 1991; Clayton & Bernardinelli, 1992). In that context, κ needs to be estimated and the extra term $\kappa^{-n/2}$ must be included in (3·1). The conventional vague prior for κ must be avoided, since it leads to an improper posterior; this problem is not related to the impropriety of (3·1), but occurs in the simplest of hierarchical formulations.

The distribution in Example 3·1 is somewhat unsatisfactory in that it does not arise as the limit of an autoregression with equal variances and the same neighbour covariances; or equivalently, equal V_{ii}'s and the same W_{ij}'s for $i \sim j$, where $W_{ij} = \text{var}(X_i - X_j)$. As a general approach, one might prefer to construct intrinsic autoregressions by specifying a neighbourhood criterion and the W_{ij}'s for $i \sim j$. Such a strategy again fixes the correct number of parameters, with $Q1 = 0$ ensuring infinite variances and the W_{ij}'s taking the place of specified V_{ij}'s. Below we describe a modification of the Dempster (1972) algorithm which uses the new inputs to identify the β_{ij}'s and κ_i's in (2·1) and (2·2), provided such a rank $n-1$ autoregression exists. We first note that equating the W_{ij}'s for $i \sim j$ fixes Q apart from scale, and that a corresponding re-analysis of the epidemiological data sets given by Besag et al. (1991) produces negligible changes in the point and interval estimates of relative risk. Indeed, any other conclusion would have been very unsatisfactory in this particular case.

3·2. *Modified Dempster's algorithm*

THEOREM 3·1. *Define J_0 to be the set of all $\frac{1}{2}n(n-1)$ index pairs (i, j), for $1 \leqslant i < j \leqslant n$, and let (J_1, J_2) be a particular ordered partition of J_0, with J_1 having m elements, say. Let W^* and Q^* be fixed $n \times n$ matrices. Let A and $\bar{A}$ be as in the proof of Corollary 3·2. If there exist symmetric $n \times n$ matrices Q and W such that:*

(i) there is a positive definite $(n-1) \times (n-1)$ matrix V_Y, with

$$V_Y = (\bar{A}^{\mathrm{T}} Q \bar{A})^{-1}, \quad (V_Y)_{ii} + (V_Y)_{jj} - 2(V_Y)_{ij} = W_{ij} \quad (i, j < n), \quad (V_Y)_{ii} = W_{in} \quad (i < n),$$

(ii) $W_{ij} = W^*_{ij}$ for $(i, j) \in J_1$, $W_{nn} = 0$,

(iii) $Q_{ij} = Q^*_{ij}$ for $(i, j) \in J_2$,

(iv) $Q1 = 0$,

then Q and W are unique.

The argument of Dempster (1972) extends to the current intrinsic case and leads to the following Newton–Raphson algorithm for Q and W.

Algorithm 3·1. For any symmetric $n \times n$ matrix B, let $\varphi_1(B)$ be the row vector whose kth element is B_{ij}, where (i, j) is the kth element of J_1. Define $\varphi_2(B)$ correspondingly with respect to J_2 and write $\varphi(B) = (\varphi_1(B), \varphi_2(B))$. Let $Q^{(0)}$ be any symmetric positive definite matrix having properties (iii) and (iv) above. Then the following steps for $l = 0, 1, \ldots$ will converge to Q and W:

(a) set $V_Y^{(l)} = (\bar{A}^{\mathrm{T}} Q^{(l)} \bar{A})^{-1}$;

(b) form the $n \times n$ symmetric matrix $W^{(l)}$ with (i, j) element

$$W_{ij}^{(l)} = \begin{cases} (V_Y^{(l)})_{ii} + (V_Y^{(l)})_{jj} - 2(V_Y^{(l)})_{ij} & \text{if } i, j < n, \\ (V_Y^{(l)})_{ii} & \text{if } i < n,\ j = n, \\ 0 & \text{if } i = n,\ j = n; \end{cases}$$

(c) form the $m \times m$ matrix $D^{(l)}$ with (k_0, k_1) element

$$D^{(l)}_{k_0 k_1} = \tfrac{1}{2}(W^{(l)}_{i_0 i_1} + W^{(l)}_{j_0 j_1} - W^{(l)}_{i_0 j_1} - W^{(l)}_{j_0 i_1})^2,$$

where (i_0, j_0) and (i_1, j_1) are the k_0th and k_1th elements of J_1;

(d) obtain $Q^{(l+1)}$, defined by

$$\varphi_1(Q^{(l+1)}) = \varphi_1(Q^{(l)}) + \{\varphi_1(W^*) - \varphi_1(W^{(l)})\}(D^{(l)})^{-1}, \quad \varphi_2(Q^{(l+1)}) = \varphi_2(Q^{(l)}),$$
$$Q^{(l+1)}1 = 0.$$

3·3. *Regular arrays*

As in § 2·3, we consider Gaussian variables satisfying (2·5) on the doubly-infinite rectangular lattice but now suppose that (iv) gives equality when $\omega_1 = \omega_2 = 0$. The generalised spectral density of this intrinsic process again follows from (2·7) and its behaviour in the neighbourhood of the origin determines which contrast processes have proper stationary distributions. Maximum likelihood estimation from a partial realisation is described by Künsch (1987). The estimates are invariant to the addition of any constant to the realisation, a property of particular interest in texture analysis.

When all contrasts in an intrinsic autoregression have proper distributions, the semivariogram (Matheron, 1973; Cressie, 1991, Ch. 2),

$$v_{rs} = \tfrac{1}{2}\operatorname{var}(X_{rs} - X_{00}) \quad (r, s = 0, \pm 1, \ldots)$$

is well defined and satisfies, compare with (2·6),

$$v_{rs} = -\kappa\delta_{rs} + \sum_k \sum_l \eta_{kl} v_{r+k,s+l}.$$

Partial realisations, apart from an arbitrary level, can be generated by the spectral method referred to in § 2·3.

Example 3·2. Consider the asymmetric version of Example 2·2,

$$E(X_{uv}|\ldots) = \eta_{10}(x_{u-1v} + x_{u+1v}) + \eta_{01}(x_{uv-1} + x_{uv+1}), \tag{3·2}$$

but with $\eta_{10} + \eta_{01} = \tfrac{1}{2}$. Then,

$$v_{rs} = -\kappa\delta_{rs} + \eta_{10}(v_{r-1s} + v_{r+1s}) + \eta_{01}(v_{rs-1} + v_{rs+1}), \tag{3·3}$$

and also

$$(2r-1)v_{r-1r-1} + (2r+1)v_{r+1r+1} = 4rv_{rr} \quad (r = 0, 1, \ldots);$$

compare with equation (3·4) of Besag (1981) inserting a missing factor of four on the right-hand side.

It can be shown that,

$$v_{10} = \frac{\kappa}{\pi\eta_{10}} \tan^{-1}\left(\frac{\eta_{10}}{\eta_{01}}\right)^{\frac{1}{2}}, \quad v_{01} = \frac{\kappa}{\pi\eta_{01}} \tan^{-1}\left(\frac{\eta_{01}}{\eta_{10}}\right)^{\frac{1}{2}}, \quad v_{11} = \frac{\kappa}{\pi(\eta_{10}\eta_{01})^{\frac{1}{2}}} = v_{-11}. \tag{3·4}$$

Equations (3·3) also occur in two-dimensional simple random walk and it is well known (Spitzer, 1976, p. 148) that, in the symmetric case, $\eta_{10} = \eta_{01} = \frac{1}{4}$,

$$v_{10} = v_{01} = \kappa, \quad v_{rr} = v_{-rr} = \frac{4\kappa}{\pi}\left(1 + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2r-1}\right) \quad (r = 1, 2, \ldots)$$

determining v_{rs} for all r and s via (3·3).

The formulae for v_{10} and v_{01} in (3·4), in conjunction with the modified Dempster's algorithm, are used by Besag & Higdon (1993, § 4) to carry out two-dimensional fertility adjustment, including estimation of κ and η_{10}, in a fully Bayesian analysis of a variety trial on spring wheat.

Here we consider a more complicated spatial model, in which diagonally adjacent sites are included as neighbours, although in the simpler practical context of plots without treatments.

Example 3·3. Kempton & Howes (1981, Table 2) provide the yields from a 28 × 7 uniformity trial on spring barley, carried out in 1979 at the Plant Breeding Institute, Cambridge, U.K. The authors fit a four-neighbour auto-normal scheme by ordinary least squares but note the invalidity of the resulting estimates for an infinite lattice model. Alternatively, asymptotic maximum likelihood estimation in (2·5) gives $\eta_{10} = 0{\cdot}4848$ and $\eta_{01} = 0{\cdot}0132$, so that $\eta_{++} = 0{\cdot}9960$.

In this example, we might prefer an intrinsic autoregression, with coefficients chosen to match the empirical v_{rs} for those r and s that correspond to neighbour pairs, reflecting the asymptotic maximum likelihood theory of Künsch (1987). We consider an eight-neighbour model, for which the relevant empirical values are $v_{10} = 0{\cdot}3516$, $v_{01} = 1{\cdot}1735$, $v_{11} = 1{\cdot}3395$, $v_{-11} = 1{\cdot}1752$; the small value for v_{10} reflects the particularly strong within-column associations. We chose to average the two diagonal values and equate the corresponding coefficients, although this does not simplify the computations. Table 2 shows the resulting nonzero Q_{ij}'s for each plot $i = (u, v)$ in selected rows and the first four columns. Thus, in plot $i = (14, 4)$, the β_{ij}'s are obtained on division by $-5{\cdot}7631$ and are 0·4829 for column neighbours, 0·2039 for row neighbours and $-0{\cdot}0934$ for diagonal neighbours, although equality for left and right neighbours, for example, is not exact in general. The negative diagonal terms suggest a curvature effect.

Table 2. *Example* 3·3: *Q coefficients obtained using the modified Dempster's algorithm to fit an eight-neighbour intrinsic autoregression to the Kempton & Howes* (1981) 28 × 7 *uniformity data*

u	$v=1$			$v=2$			$v=3$			$v=4$		
1	3·26	−0·96	−0·96	3·63	−0·95	−0·95	3·62	−0·95	−0·95	3·62	−0·95	
	−2·77	0·48	0·53	−2·76	0·51	0·52	−2·76	0·52	0·52	−2·77	0·52	
2	−2·77	0·53	0·48	−2·76	0·52	0·51	−2·76	0·52	0·52	−2·77	0·52	
	5·65	−1·12	−1·12	5·75	−1·14	−1·14	5·75	−1·15	−1·15	5·75	−1·15	
	−2·77	0·49	0·52	−2·76	0·51	0·53	−2·77	0·52	0·53	−2·77	0·53	
3	−2·77	0·52	0·49	−2·76	0·53	0·51	−2·77	0·53	0·52	−2·77	0·52	
	5·64	−1·12	−1·12	5·75	−1·15	−1·15	5·75	−1·16	−1·16	5·75	−1·16	
	−2·77	0·50	0·52	−2·77	0·52	0·53	−2·77	0·53	0·53	−2·77	0·53	
4	−2·77	0·52	0·50	−2·77	0·53	0·52	−2·77	0·53	0·53	−2·77	0·53	
	5·64	−1·13	−1·13	5·75	−1·16	−1·16	5·76	−1·16	−1·16	5·76	−1·16	
	−2·77	0·50	0·52	−2·77	0·52	0·53	−2·77	0·53	0·53	−2·78	0·53	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14	−2·77	0·52	0·50	−2·77	0·53	0·52	−2·77	0·53	0·53	−2·78	0·53	
	5·62	−1·14	−1·14	5·76	−1·17	−1·17	5·76	−1·18	−1·18	5·76	−1·18	
	−2·76	0·52	0·52	−2·78	0·53	0·53	−2·78	0·54	0·54	−2·78	0·54	

3·4. *Examples of prior distributions on regular arrays*

When $\eta_{10}=\eta_{01}=\frac{1}{4}$, the interpolation (3·2) can be thought of as the least squares fit of a plane to the values at the four neighbours of (u, v). When dealing with large arrays, as for example in image analysis, one would often prefer to use a larger neighbourhood and a more sophisticated local representation of the underlying true surface. In particular, we assume here that the surface is perceived to be locally quadratic, although the ideas are not restricted to polynomial representations. We now encounter intrinsic autoregressions with more than a single deficiency in rank. In one dimension a locally quadratic prior corresponds to independent second differences and is used for ordinal factors in logistic regression by Berzuini, Clayton & Bernardinelli (1993) and by Besag et al. (1995).

Example 3·4. Define the neighbours of each $i=(u, v)$ on the doubly-infinite rectangular lattice to be its eight nearest sites. Then the least-squares locally quadratic fit generates the intrinsic autoregression for which

$$E(X_{uv}|\ldots)=\tfrac{1}{2}(x_{u-1v}+x_{u+1v}+x_{uv-1}+x_{uv+1})$$
$$-\tfrac{1}{4}(x_{u-1v-1}+x_{u+1v+1}+x_{u-1v+1}+x_{u+1v-1}). \tag{3·5}$$

The corresponding generalised spectral density is inversely proportional to

$$(1-\cos\omega_1)(1-\cos\omega_2),$$

so that simple differences no longer have proper distributions and it is necessary instead to consider genuine two-dimensional contrasts, such as

$$Y_{uv}=X_{uv}-X_{u+1v}-X_{uv+1}+X_{u+1v+1}. \tag{3·6}$$

It is easily checked that the Y_{uv}'s in (3·6) are independent $N(0, 4\kappa)$ random variables and that the notional density (3·1) is invariant to the addition of constants to any rows and columns.

The above degeneracy suggests that (3·5) may be useful in two-way tables, as a prior distribution that accommodates arbitrary row and column effects. Furthermore, the restriction of the process to a finite array, $\{(u, v): 0\leqslant u\leqslant p,\ 0\leqslant v\leqslant q\}$, is very easy to handle exactly. For example,

$$E(X_{00}|\ldots)=x_{10}+x_{01}-x_{11},\quad \mathrm{var}(X_{00}|\ldots)=4\kappa,$$
$$E(X_{u0}|\ldots)=x_{u1}+\tfrac{1}{2}(x_{u-10}+x_{u+10}-x_{u-11}-x_{u+11}),\quad \mathrm{var}(X_{u0}|\ldots)=2\kappa\quad(0<u<p),$$

with corresponding results at other corners and edges. In terms of Lemma 3·1, Q has a deficiency in rank of order $p+q-1$ and the Y_i's can be taken as in (3·6). In frequentist terms, (3·6) resembles a fertility model adopted by Cullis & Gleeson (1991), in which first-differencing in each direction produces independence, and can be interpreted as the limiting case of the separable stationary autoregressions advocated by Martin (1979, 1990) in the same context.

However, the degeneracy with regard to row and column effects make (3·5) unsuitable as a prior for a slowly varying surface. This defect can be remedied by expanding the system of neighbours for each site.

Example 3·5. Define the neighbours of each $i=(u, v)$ on the doubly-infinite rectangular lattice to be its twelve nearest sites. Then the least-squares locally quadratic fit generates the intrinsic autoregression for which

$$E(X_{uv}|\ldots)=\tfrac{1}{4}(x_{u-1v}+x_{u+1v}+x_{uv-1}+x_{uv+1})+\tfrac{1}{8}(x_{u-1v-1}+x_{u+1v+1}+x_{u-1v+1}+x_{u+1v-1})$$
$$-\tfrac{1}{8}(x_{u-2v}+x_{u+2v}+x_{uv-2}+x_{uv+2}).$$

The generalised spectrum is now inversely proportional to

$$(1-\cos\omega_1)(1-\cos\omega_2)+(\cos\omega_1-\cos\omega_2)^2.$$

Again, simple differences do not have proper distributions but now any second difference process of the form $Y_{uv}=X_{uv}-2X_{u+k,v+l}+X_{u+2k,v+2l}$, where k and l are positive integers, is stationary, as is any genuinely two dimensional contrast process, such as that defined in (3·5). Note that, for example, $X_{uv}-2X_{u+1v}+X_{u+1v+1}$ does not have a proper distribution, because of invariance to the addition of a plane, although not to the addition of arbitrary constants to rows and columns.

We do not envisage that such a detailed model is likely to be useful on small arrays but it has been applied successfully to larger problems in surface reconstruction, where a slowly varying image is contaminated by blur and/or noise. For an example involving noise, see Kooperberg (1993).

Acknowledgement

We express thanks to David Madigan for the reference to Dempster (1972) and to David Downham for pointing out the typographical error in the paper by Besag (1981). The help of the referees is gratefully acknowledged. This work was supported in part by grants from the National Science Foundation.

References

Bartlett, M. S. (1978). Nearest neighbour models in the analysis of field experiments (with discussion). *J. R. Statist. Soc.* B **40**, 147–74.

Bernardinelli, L. & Montomoli, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statist. Med.* **11**, 983–1007.

Berzuini, C., Clayton, D. G. & Bernardinelli, L. (1993). Bayesian inference on the Lexis diagram. *Bull. Int. Statist. Inst.* **55** (Bk. 1), 149–64.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. R. Statist. Soc.* B **36**, 192–236.

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

Besag, J. (1977). Errors-in-variables estimation for Gaussian lattice schemes. *J. R. Statist. Soc.* B **39**, 73–8.

Besag, J. (1981). On a system of two-dimensional recurrence equations. *J. R. Statist. Soc.* B **43**, 302–9.

Besag, J. (1994). Discussion of paper by L.-J. Tierney. *Ann. Statist.* **22**, 1734–41.

Besag, J., Green, P. J., Higdon, D. M. & Mengersen, K. L. (1995). Bayesian computation and stochastic systems (with Discussion). *Statist. Sci.* **10**, 3–66.

Besag, J. & Higdon, D. M. (1993). Bayesian inference for agricultural field experiments. *Bull. Int. Statist. Inst.* **55** (Bk. 1), 121–36.

Besag, J. & Kempton, R. A. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **78**, 301–4.

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

Besag, J., York, J. C. & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with Discussion). *Ann. Inst. Statist. Math.* **43**, 1–59.

Chellappa, R. & Kashyap, R. L. (1985). Texture synthesis using 2-d noncausal autoregressive models. *IEEE Trans. Acoust. Speech Sig. Proces.* **33**, 30–9.

Clayton, D. G. & Bernardinelli, L. (1992). Bayesian methods for mapping disease risk. In *Small Area Studies in Geographical and Environmental Epidemiology*, Ed. J. Cuzick and P. Elliott, pp. 205–20. Oxford University Press.

Clayton, D. G. & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–81.

Cliff, A. D. & Ord, J. K. (1975). Model building and the analysis of spatial pattern in human geography (with Discussion). *J. R. Statist. Soc.* B **37**, 297–348.

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

COHEN, F. S. & COOPER, D. B. (1987). Simple parallel hierarchical and relaxation algorithms for segmenting non-causal Markov random fields. *IEEE Trans. Pat. Anal. Mach. Intel.* **9**, 195–219.

COHEN, F. S., FAN, Z. & PATEL, M. A. (1991). Classification of rotated and scaled textured images using Gaussian Markov random field models. *IEEE Trans. Pat. Anal. Mach. Intel.* **13**, 192–202.

COHEN, F. S. & PATEL, M. A. (1991). Modeling and synthesis of images of 3D textured surfaces. *Comp. Vis. Graph. Image Proces.: Graph. Mod. Image Proces.* **53**, 501–10.

CRESSIE, N. A. C. (1991). *Statistics for Spatial Data.* New York: Wiley.

CULLIS, B. R. & GLEESON, A. C. (1991). Spatial analysis of field experiments—an extension to two dimensions. *Biometrics* **47**, 1449–60.

DAHLHAUS, R. & KÜNSCH, H. R. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74**, 877–82.

DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–75.

GEMAN, S. & GRAFFIGNE, C. (1987). Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians, 1986*, Ed. A. M. Gleason, pp. 721–41. Providence: Am. Math. Soc.

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

GUYON, X. (1992). *Champs Aléatoires sur un Réseau.* Paris. Masson.

JINCHI, H. & CHELLAPPA, R. (1986). Restoration of blurred and noisy images using Gaussian Markov random field models. In *Proc. Princeton Conf. on Information Sciences and Systems.*

KEMPTON, R. A. & HOWES, C. W. (1981). The use of neighbouring plot values in the analysis of variety trials. *Appl. Statist.* **30**, 59–70.

KITTLER, J. & FÖGLEIN, J. (1984). Contextual classification of multispectral pixel data. *Image Vis. Comp.* **2**, 13–29.

KOOPERBERG, C. (1993). Discussion of the meeting on the Gibbs sampler and other Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B **55**, 79–81.

KÜNSCH, H. R. (1983). Approximations to the likelihood equations for some Gaussian random fields. *Scand. J. Statist* **10**, 239–46.

KÜNSCH, H. R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika* **74**, 517–24.

LÉVY, P. (1948). Chaînes doubles de Markoff et functions aléatoires de deux variables. *C. R. Acad. Sci., Paris* **226**, 53–5.

MARDIA, K. V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J. Mult. Anal.* **24**, 265–84.

MARSHALL, R. J. (1991). A review of methods for the statistical analysis of spatial patterns of disease. *J. R. Statist. Soc.* A **154**, 421–41.

MARTIN, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika* **66**, 209–17.

MARTIN, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Commun. Statist.* A **19**, 55–81.

MATHERON, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Prob.* **5**, 439–68.

MOLINA, R. & RIPLEY, B. D. (1989). Using spatial models as priors in astronomical image analysis. *J. Appl. Statist.* **16**, 193–206.

MOLLIÉ, A. & RICHARDSON, S. (1991). Empirical Bayes estimates of cancer mortality rates using spatial models. *Statist. Med.* **10**, 95–112.

MORAN, P. A. P. (1973). A Gaussian Markovian process on a square lattice. *J. Appl. Prob.* **10**, 54–63.

RIPLEY, B. D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

RIPLEY, B. D. (1991). The use of spatial models as image priors. In *Spatial Statistics and Imaging*, Ed. A. Possolo, pp. 309–40. Hayward: Institute of Mathematical Statistics.

ROSANOV, Y. A. (1967). On the Gaussian homogeneous field with given conditional distributions. *Theory Prob. Applic.* **12**, 381–91.

SIMCHONY, T., CHELLAPPA, R. & LICHTENSTEIN, Z. (1989). Graduated non-convexity algorithm for image estimation using compound Gauss-Markov random fields. In *Proc. ICASSP '89, IEEE Conf. Acoust., Speech Sig. Proces.* **3**, pp. 1417–20. New York: IEEE.

SPITZER, F. (1976). *Principles of Random Walk*, 2nd ed. New York: Springer.

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **49**, 305–14.

ZERUBIA, J. & CHELLAPPA, R. (1990). Mean field approximation using compound Gauss-Markov random fields for edge detection and image estimation. In *Proc. ICASSP '90, IEEE Conf. Acoust., Speech Sig. Proces.* **4**, pp. 2193–6. New York: IEEE.

[*Received March* 1993. *Revised April* 1995]