

## **Intrinsic autoregressions and related models on the two-dimensional lattice**

BY HANS R. KÜNSCH

*Eidgenössische Technische Hochschule, Seminar für Statistik, 8092 Zürich, Switzerland*

### SUMMARY

Stationary autoregressions on a two-dimensional lattice are generalized to intrinsic models where only increments are assumed to be stationary. Prediction formulae and the asymptotic behaviour of the semivariogram are derived. For parameter estimation we propose an approximate maximum likelihood estimator, a generalization of Whittle's estimator; it is derived also for general intrinsic models.

*Some key words:* Approximate maximum likelihood; Autoregression on  $Z^2$ ; Intrinsic model; Mean free prediction; Semivariogram.

### 1. INTRODUCTION

Most models proposed in the literature for planar lattice data  $X_i$ , for  $i = (i_1, i_2) \in Z^2$ , assume stationarity. For applications, however, this is often not very realistic. Locally the assumption of stationarity is reasonable, but over larger areas the level and sometimes also the gradient change substantially. For instance, Wilkinson et al. (1983) use a smooth trend plus independent errors model for field trials. In time series in such a situation one might model the first or second differences of the data. Yet in two dimensions this approach is not very satisfactory since differences can be taken in many directions. A more detailed discussion of the disadvantages of differencing is given in § 2.

We pursue in this paper the use of the so-called intrinsic models of geostatistics (Matheron, 1973) where all possible spatial differences are modelled simultaneously. However, whereas geostatistics uses only some small classes with a simple semivariogram  $\gamma(k) = \frac{1}{2}E\{(X_i - X_{i+k})^2\}$  ( $k \in Z^2$ ), we introduce here a class of models with simple prediction formulae by which we can approximate almost any intrinsic model. This class, called intrinsic autoregressions, is a generalization of the conditional autoregressions of Besag (1974) to the nonstationary case. They have a simple spectrum, but it does not seem possible to express the semivariogram in closed form. Section 2 gives their properties.

In § 3 we develop an approximation to the log likelihood of intrinsic autoregressions which is very similar to the approximation of Whittle (1954) in the stationary case. The corresponding estimator uses only the empirical semivariogram for small lags. If there is a monotonic trend in the data, this is much more reliable than the empirical covariances used to fit a stationary autoregression. We show also how the approximation can be extended to other intrinsic models. The resulting estimator can be applied whenever the spectrum can be calculated, for instance in the case of a linear semivariogram. Section 4 contains an example using Landsat data.

Often the sum of the estimated coefficients of a stationary autoregression is very close to one. Since the asymptotic normality of these estimators breaks down when the sum

of the coefficients tends to one, it is then difficult to make inferences. Our intrinsic autoregressions can be considered as the limit of stationary autoregressions as the sum of the coefficients tends to one. Using intrinsic autoregressions the above difficulty disappears because here the asymptotic normality of the parameter estimators continues to hold. This is another attractive feature of our models.

2. INTRINSIC AUTOREGRESSIONS

First recall the definition and the basic properties of the intrinsic model of Matheron (1973).

*Definition 2.1.* A finite linear combination  $\sum \lambda_i X_i$  is called an increment of order  $d$  if  $\sum \lambda_i i_1^\alpha i_2^{\alpha'} = 0$  for all  $\alpha \geq 0, \alpha' \geq 0, 0 \leq \alpha + \alpha' \leq d$ .

Thus an increment of order  $d$  filters out polynomials of order  $d$ . Increments of order zero are also called contrasts. Note that in field experiments all estimators of treatment effects are contrasts. All subsequent coefficients  $\lambda_i$  are such that  $\sum \lambda_i X_i$  is an increment. A process  $Y_j(\lambda) = \sum \lambda_i X_{i+j}$  ( $j \in Z^2$ ) is called an increment process.

*Definition 2.2.* The process  $\{X_i\}$  for  $i \in Z^2$  is called intrinsic of order  $d$  if all increment processes of order  $d$  are stationary.

Usually it is assumed that all increments have zero means, and we shall do so in the following. Clearly, if  $\{X_i\}$  is intrinsic and second moments of the increments exist, each increment process  $\{Y_i(\lambda)\}$  has a spectrum  $F_{Y(\lambda)}(d\omega)$ , for  $\omega \in (-\pi, \pi]^2$ . Moreover, for different coefficients  $\lambda$ , these spectra are related and can be expressed by means of a single measure. We state the result only for the absolutely continuous case. The general case as well as proofs are given by Matheron (1973) and Gelfand & Vilenkin (1964, § III, 5.2). Let  $j\omega$  ( $j \in Z^2, \omega \in (-\pi, \pi]^2$ ) denote the inner product  $j_1\omega_1 + j_2\omega_2$ .

**THEOREM 2.1.** *If  $\{X_i\}$  is intrinsic and all increment processes have an absolutely continuous spectrum, then there exists a function  $f(\omega)$  ( $\omega \in (-\pi, \pi]^2$ ) with  $f(\omega) \geq 0, f(-\omega) = f(\omega)$  and  $\int \|\omega\|^{2d+2} f(\omega) d\omega < \infty$  such that*

$$E(\sum \lambda_j X_j \sum \lambda'_k X_k) = (2\pi)^{-2} \int \sum \lambda_j e^{ij\omega} \sum \lambda'_k e^{-ik\omega} f(\omega) d\omega. \tag{2.1}$$

Then  $f(\omega)$  is called the spectral density of  $\{X_i\}$ . It is similar to the spectral density of a stationary process except that it can have a nonintegrable singularity at  $\omega = 0$  and thus may be used to calculate only covariances between increments. In particular, the spectral density of an increment process  $\{Y_i(\lambda)\}$  is, by (2.1),

$$f_{Y(\lambda)}(\omega) = |\sum \lambda_j e^{ij\omega}|^2 f(\omega). \tag{2.2}$$

Any function  $f$  with the properties of Theorem 2.1 is the spectral density of some intrinsic process  $\{X_i\}$ . However, in contrast to the stationary case, even a Gaussian  $\{X_i\}$  is not completely determined by its spectral density since one can always add a polynomial of degree  $\leq d$ . One can thus choose the values of  $X_i$  at  $\frac{1}{2}(d+1)(d+2)$  points freely. An intrinsic model should be regarded as a whole equivalence class of distributions of  $\{X_i\}$ .

Among stationary models, the conditional autoregressions of Besag (1974) form a basic class of parametric models. They are defined by

$$E(X_i|X_j, j \neq i) = \left(1 - \sum_{k \in N} a_k\right) E(X_i) + \sum_{k \in N} a_k X_{i+k}, \tag{2.3}$$

where  $N$  is a finite symmetric neighbourhood of the origin and  $a_k = a_{-k}$ . This means that only the values close by are used in the predictor of a missing value. With increasing  $N$  we get models of increasingly complex interactions. Assuming that  $\{X_i\}$  has a spectral density  $f$ , it can easily be shown (Rosanov, 1967) that

$$f(\omega) = \sigma^2 \left\{1 - \sum_{k \in N} a_k \cos(k\omega)\right\}^{-1}, \tag{2.4}$$

where  $\sigma^2$  is the mean square prediction error  $E[\{X_i - E(X_i|X_j, j \neq i)\}^2]$ . Since  $f(\omega)$  is integrable, we must have  $\sum a_k \cos(k\omega) < 1$  for all  $\omega$ . In particular,  $\sum a_k < 1$ , so the interpolation always uses the mean. If  $\{X_i\}$  is Gaussian, (2.4) implies (2.3), and in any case the right-hand side of (2.3) is the best linear predictor.

However, it is not quite obvious how to generalize these autoregressions to the class of intrinsic models. By analogy to the ARIMA, autoregressive integrated moving average, models of Box & Jenkins (1970) one might ask that a certain increment process is a stationary autoregression. This approach fails in two dimensions. If  $\{X_i\}$  is intrinsic with spectrum  $f$  and the increment process  $\{\sum \lambda_{i+j} X_j\}$  is a stationary autoregression, then, by (2.2) and (2.4),  $\sigma^2 \{1 - \sum a_k \cos(k\omega)\}^{-1} = |\sum \lambda_j e^{ij\omega}|^2 f(\omega)$ . Because  $\lambda_j$  are the coefficients of an increment,  $|\sum \lambda_j e^{ij\omega}|^2 = (\sum \lambda_j j\omega)^2 + O(\|\omega\|^4)$  as  $\omega \rightarrow 0$  and hence  $\|\omega\|^2 |\sum \lambda_j e^{ij\omega}|^{-2}$  is not integrable at zero. This contradicts  $\int \|\omega\|^2 f(\omega) d\omega < \infty$  which should hold for  $d = 0$ . We thus have seen that for  $d = 0$  no increment process of an intrinsic model can be a stationary autoregression. Further, whereas for  $d = 1$  there are intrinsic models  $\{X_i\}$  such that

$$X_i - \beta(X_{i+(1,0)} + X_{i-(1,0)}) - (\frac{1}{2} - \beta)(X_{i+(0,1)} + X_{i-(0,1)})$$

with  $0 < \beta < \frac{1}{2}$  is a stationary autoregression, it is not clear how one can estimate  $\beta$  and the model class obtained in this way is much smaller than the one we propose next.

We base our generalization on the form of the spectral density (2.4).

*Definition 2.3.* Let  $a_k$  ( $k \in N$ ) be coefficients with  $a_k = a_{-k}$  and  $P(\omega) = 1 - \sum a_k \cos(k\omega) \geq \text{const } \|\omega\|^{2d+2}$  for all  $\omega \in (-\pi, \pi]^2$ . Then the intrinsic model with spectral density  $f(\omega) = \sigma^2 / P(\omega)$  is called an intrinsic autoregression with neighbourhood  $N$ . The order is the smallest integer  $d$  such that  $P$  satisfies the inequality above.

In the one-dimensional case, if an even trigonometric polynomial  $P(\omega)$  has a zero of order  $2d + 2$  at  $\omega = 0$ , then  $P(\omega)$  contains the factor  $(1 - \cos \omega)^{d+1}$ . Hence a time series  $\{X_i\}$  is an intrinsic autoregression if and only if its  $(d + 1)$ th difference is a stationary autoregression. In more than one dimension such a factorization of polynomials is no longer possible. For this reason our class is more general than the one obtained by the approach described in the last paragraph.

What is the analogue of (2.3) for our intrinsic autoregressions? Here a problem occurs since the spectrum no longer determines all the covariances of  $\{X_i\}$ . It is, therefore, not possible to calculate from the spectrum the best linear predictor of  $X_i$  given  $X_j$  ( $j \neq i$ ). But the rationale behind the intrinsic model is to eliminate polynomial trends, so the prediction and the inference should also be independent of such trends. Hence we restrict the predictions considered as follows.

**Definition 2.4.** Let  $V \subset Z^2$  be finite. A finite linear combination  $\sum \mu_j X_j$  is called an intrinsic predictor of  $X_i$  ( $i \in V$ ) given  $X_j$  ( $j \notin V$ ) if  $\mu_j = 0$  for all  $j \in V$  and  $X_i - \sum \mu_j X_j$  is an increment. It is called the best intrinsic predictor if

$$E\{(X_i - \sum \mu_j X_j)^2\} \leq E\{(X_i - \sum \mu'_j X_j)^2\}$$

for any other intrinsic predictor  $\sum \mu'_j X_j$ .

Then we have the following theorem.

**THEOREM 2.2.** *If  $\{X_i\}$  is an intrinsic autoregression of order  $\geq 0$ , then  $\sum a_k X_{i+k}$  is the best intrinsic predictor of  $X_i$  given  $X_j$  ( $j \neq i$ ). Conversely, if  $\{X_i\}$  is intrinsic with spectral density  $f$  and  $\sum \mu_k X_{i+k}$  is the best intrinsic predictor of  $X_i$  given  $X_j$  ( $j \neq i$ ), then*

$$\mu_k = \mu_{-k}, \quad f(\omega) = E\{(X_i - \sum \mu_k X_{i+k})^2\} / \{1 - \sum \mu_k \cos(k\omega)\};$$

that is  $\{X_i\}$  is an intrinsic autoregression.

*Proof.* For any increment  $\sum \lambda_l X_l$  we have, by (2.1),

$$E\{(X_j - \sum a_k X_{j+k}) \sum \lambda_l X_l\} = (2\pi)^{-2} \sigma^2 \int P(\omega) e^{ij\omega} \sum \lambda_l e^{-il\omega} P(\omega)^{-1} d\omega = \lambda_j \sigma^2.$$

If  $\sum \mu_l X_l$  is an intrinsic predictor of  $X_j$ ,  $\mu_j = 0$  and  $\sum a_k X_{j+k} - \sum \mu_l X_l$  is an increment. Hence

$$E\{(X_j - \sum \mu_l X_l)^2\} = E\{(X_j - \sum a_k X_{j+k})^2\} + E\{(\sum a_k X_{j+k} - \sum \mu_l X_l)^2\},$$

and the first part follows. For the converse we first note that for  $d = 0$  we must have

$$E\{(X_j - \sum \mu_k X_{j+k})(X_l - X_{l-(1,0)})\} = 0 \quad (l \neq j, j + (1, 0)),$$

otherwise  $\sum \mu_k X_{j+k} + \varepsilon(X_l - X_{l-(1,0)})$  would be better for some  $\varepsilon$ . Hence we obtain from (2.1)

$$f(\omega)(1 - \sum \mu_k e^{ik\omega})(1 - e^{i\omega_1}) = c_1 + c_2 e^{i\omega_1}$$

except on a set of Lebesgue measure zero. In other words  $f(\omega)(1 - \sum \mu_k e^{ik\omega})$  is a function of  $\omega_1$  alone. The same argument with  $X_l - X_{l-(0,1)}$  shows that  $f(\omega)(1 - \sum \mu_k e^{ik\omega})$  is a function of  $\omega_2$  alone, and hence it is a constant. For  $d > 0$  the arguments are similar.  $\square$

For the formulation of the next result it is convenient to define  $a_k$  for all  $k \in Z^2$  by putting  $a_k = 0$  if  $k \notin N$ .

**THEOREM 2.3.** *Let  $\{X_i\}$  be an intrinsic autoregression. Then for any finite  $V$ , the matrix  $(\delta_{ij} - a_{i-j})_{i,j \in V}$  has an inverse denoted by  $g^V$ . Furthermore, with  $h_{ij}^V = \sum g_{ik}^V a_{k-j}$  ( $i \in V, j \notin V$ ), the best intrinsic interpolation of  $X_i$  ( $i \in V$ ) given  $X_j$  ( $j \notin V$ ) is  $\sum h_{ij}^V X_j$  and*

$$E\{(X_i - \sum h_{ij}^V X_j)(X_{i'} - \sum h_{i'j'}^V X_{j'})\} = \sigma^2 g_{ii'}^V.$$

*Proof.* We give the proof for  $d = 0$ , the other cases being similar. Choose  $i_0 \notin V$  and put  $Y_i = X_i - X_{i_0}$ . In the proof of Theorem 2.2 it was shown that  $Y_i - \sum a_k Y_{i+k} = X_i - \sum a_k X_{i+k}$  is uncorrelated with  $Y_j$  for  $j \neq i$ , so in the Gaussian case

$$E(Y_i | Y_j, j \neq i) = \sum a_k Y_{i+k} \quad (i \neq i_0), \quad E\{(Y_i - \sum a_k Y_{i+k})^2\} > 0.$$

The existence of an inverse follows now from Künsch (1979, Th. 4.1, form. (4.5)). Furthermore

$$\sum_{j \in V} h_{ij}^V = \sum_{k \in V} g_{ik}^V \sum_{j \in V} a_{k-j} = \sum_{k \in V} g_{ik}^V - \sum_{k \in V} g_{ik}^V \sum_{j \in V} a_{k-j} = 1,$$

so that  $\sum h_{ij}^\vee X_j$  is an intrinsic predictor and  $X_i - \sum h_{ij}^\vee X_j = Y_i - \sum h_{ij}^\vee Y_j$ . But this last random variable is by the same theorem of Künsch uncorrelated with  $Y_j$  ( $j \notin V$ ), and therefore also with all increments  $\sum \lambda_j X_j$  such that  $\lambda_j = 0$  for all  $j \in V$ . By the usual orthogonality argument the proof is completed.  $\square$

According to Matheron (1973, Th. 2.4), an intrinsic model of order zero is stationary if and only if the semivariogram is bounded. We next show that the semivariogram of intrinsic autoregressions in two dimensions increases logarithmically. Such a behaviour is known in geostatistics in the so-called de Wijsian scheme and indicates a rather mild departure from stationarity.

**THEOREM 2.4.** *Let  $\gamma(k) = \frac{1}{2}E\{(X_{i+k} - X_i)^2\}$  be the semivariogram of an intrinsic autoregression of order zero with coefficients  $a_k$  ( $k \in N$ ). As  $\|k\| \rightarrow \infty$ ,  $\gamma(k)/\log \|k\|$  converges to  $\sigma^2/\{2\pi(\det M)^{\frac{1}{2}}\}$ , where  $M$  is a  $2 \times 2$  matrix with elements  $M_{ij} = \sum a_k k_i k_j$ . Moreover  $\gamma(k) - \sigma^2/\{2\pi(\det M)^{\frac{1}{2}}\} \log \|k\|$  converges to a constant which depends on  $\lim k/\|k\|$ . Here  $\|k\|^2 = k_1^2 + k_2^2$ .*

When all coefficients  $a_k$  are positive, the semivariogram is proportional to the Green's function of the random walk with transition probabilities  $a_k$  (Spitzer, 1964, eqn (12.3)). In this case the result is well known; see Spitzer (1964, Prop. 12.3) for a proof in the isotropic case. The proof for the general case goes along the same lines and is left to the reader.

If we want to treat more general intrinsic models, we have to consider limits of intrinsic predictors. We then can extend Theorems 2.2 and 2.3 to a large class of intrinsic models by putting  $a_k = -\alpha_k/\alpha_0$ , where

$$\alpha_k = (2\pi)^{-2} \int f(\omega)^{-1} \cos(k\omega) d\omega. \tag{2.5}$$

The only condition needed is  $\sum |\alpha_k| < \infty$ . If  $f(\omega) \sim \text{const } \|\omega\|^{-2}$  as  $\omega \rightarrow 0$ , the asymptotic growth of the semivariogram is logarithmic, whereas if  $f(\omega) \sim \text{const } \|\omega\|^{-3}$  it is linear.

For higher dimensional lattices  $Z^\nu$ , the order of the zero of  $P(\omega)$  needed to get intrinsic models depends on  $\nu$ . For instance, if  $\nu = 3$  and  $P(\omega) \sim \text{const } \|\omega\|^2$ , then  $P(\omega)^{-1}$  is integrable and thus leads to a stationary model, but its correlations are not summable. For an intrinsic autoregression of order 0 we need, for  $\nu = 3$ ,  $P(\omega) \sim \text{const } \|\omega\|^4$ . Apart from these differences, our results can be extended to  $Z^\nu$ .

### 3. PARAMETER ESTIMATION

Assume that we have observed an intrinsic autoregression  $\{X_i\}$  in an  $n \times m$  rectangle  $V$ , and that we want to estimate the parameters  $a_k$  ( $k \in N$ ) and  $\sigma^2$ . We restrict ourselves to the case  $d = 0$ . Since  $a_k = a_{-k}$  and  $\sum a_k = 1$ , the number of linearly independent parameters is  $\frac{1}{2}|N|$ . We take as our parameterization  $\alpha_k = a_k \sigma^{-2}$  ( $k \in N$ ). Denoting the empirical semivariogram  $\frac{1}{2}(n - |k_1|)^{-1}(m - |k_2|)^{-1} \sum (X_i - X_{i+k})^2$  by  $\hat{\gamma}(k)$ , we propose to estimate  $\alpha_k$  by minimizing

$$L(\alpha) = -(2\pi)^{-2} \int \log [\sum \alpha_k \{1 - \cos(k\omega)\} / (1 - \frac{1}{2} \cos \omega_1 - \frac{1}{2} \cos \omega_2)] d\omega + \sum \alpha_k \hat{\gamma}(k). \tag{3.1}$$

Taking derivatives we get

$$(2\pi)^{-2} \int \{1 - \cos(j\omega)\} [\sum \hat{\alpha}_k \{1 - \cos(k\omega)\}]^{-1} d\omega = \hat{\gamma}(j) \quad (j \in N); \quad (3.2)$$

i.e. we have to choose the parameters in such a way that the theoretical and empirical semivariograms agree for lags close to zero. This is completely analogous to the case of stationary autoregressions (Besag, 1974, § 6.3). The Hessian of  $L$  is easily seen to be strictly positive-definite for all  $(\alpha_k)$ , so the solution of (3.2) is unique and determines a minimum of  $L$ . Furthermore, consistency and asymptotic normality of  $\hat{\alpha}_k$  follow from the law of large numbers and the central limit theorem for  $\hat{\gamma}_k$ . At least for Gaussian  $\{X_i\}$  these can be proved by standard arguments. Note however that for the central limit theorem it is important to have the denominator  $(n - |k_1|)(m - |k_2|)$  instead of  $nm$  in the definition of  $\hat{\gamma}(k)$  (Guyon, 1982). Expression (3.1) is an approximation to the Gaussian log likelihood similar to the one by Whittle (1954) in the stationary case. Its derivation follows. Denote by  $W$  the interior of  $V$ , so  $i \in W$  if and only if  $i + k \in V$  for all  $k \in N$ . First condition on the boundary  $\partial W = V \setminus W$  using the best intrinsic predictors and their covariances from Theorem 2.3. Then minus twice the Gaussian conditional log likelihood is

$$\text{const} + |W| \log \sigma^2 + \log (\det g^W) + \sigma^{-2} \sum (\delta_{ij} - a_{i-j})(x_i - \sum h_{ik}^W x_k)(x_j - \sum h_{jl}^W x_l).$$

An algebraic manipulation of the quadratic form using the definitions of  $h^W$  and  $g^W$  leads to  $\sum \alpha_k \hat{\gamma}(k)$  plus boundary terms. A rather long computation shows that the expectation and variance of the boundary terms are  $O(|\partial W|)$ . Finally we need an asymptotic approximation to  $\log (\det g^W) = -\log \{ \det (\delta_{ij} - a_{i-j})_{i,j \in W} \}$ . Since  $P(\omega)$  is the spectrum of a stationary random field and  $\int \log P(\omega) d\omega > -\infty$ , we could invoke the two-dimensional version of Szegő's theorem. However we take a different approach and consider the ratio of two determinants.

**THEOREM 3.1.** *Let  $\{a_k\}_{k \in N}$  be the coefficients of an intrinsic autoregression of order 0 and put  $b_k = \frac{1}{4}$  if  $|k_1| + |k_2| = 1$ ,  $b_k = 0$  otherwise. Then*

$$nm^{-1} \log \{ \det (\delta_{ij} - a_{i-j})_{i,j \in V} / \det (\delta_{ij} - b_{i-j})_{i,j \in V} \}$$

converges to

$$(2\pi)^{-2} \int \log \{ \sum a_k (1 - \cos(k\omega)) / (1 - \frac{1}{2} \cos \omega_1 - \frac{1}{2} \cos \omega_2) \} d\omega$$

as  $n, m \rightarrow \infty$ .

The proof is similar to that of Theorem 2.5 of Künsch (1981). We consider coefficients  $a_k(\theta) = \theta a_k + (1 - \theta)b_k$  and then show first that  $(nm)^{-1} d \log \det \{ \delta_{ij} - a_{i-j}(\theta) \}_{i,j \in V} / d\theta$  converges to  $(2\pi)^{-2} \int \sum (b_k - a_k) \cos(k\omega) / \{ 1 - \sum a_k(\theta) \cos(k\omega) \} d\omega$ . Details are left to the reader.

Omitting all terms which are of smaller order or independent of  $\alpha_k$  we arrive at (3.1). In the stationary case where similar approximations occur, it seems difficult to improve upon the resulting estimators even for quite small  $n$  and  $m$  (Künsch, 1983).

The integrands in (3.1) and (3.2) are bounded, but they have a discontinuity at zero. The most successful method for the numerical integration seems to be the substitution  $\omega_i = \pi \{ 1 - \tanh(u_i) \}$  ( $i = 1, 2$ ), which takes the discontinuity to infinity and produces an

analytic integrand decaying exponentially. The integral can then be well approximated by a trapezoidal sum since the error is of smaller order than any power of the step size by the Poisson Summation formula (Henrici, 1977, eqn (10.6-21)).

For the minimization of (3.1) we used the IMSL routine ZXCGR. With nearest neighbours the convergence was quite fast, but with more neighbours a method adapted to the particular situation might be more efficient. As a starting value we used the so-called least-squares estimator which minimizes  $\sum (X_i - \sum a_k X_{i+k})^2$  under the side condition  $\sum a_k = 1$ . Since it often gives estimates  $\hat{a}_k$  for which  $1 - \sum \hat{a}_k \cos(k\omega)$  becomes negative, some modification is needed in these cases.

The approximation (3.1) of  $-2 \log$  likelihood can be extended to a parametric family of intrinsic models with spectrum  $f(\omega, \theta)$  ( $\theta \in \Theta$ ). We assume that there is a spectrum  $g(\omega)$  such that  $f(\omega, \theta)/g(\omega)$  is continuous and bounded away from zero. Then the analogue of (3.1) is

$$L(\theta) = (2\pi)^{-2} \int \log \{f(\omega, \theta)/g(\omega)\} d\omega + \sum \alpha_k(\theta) \hat{\gamma}(k),$$

where  $\alpha_k$  has been defined in (2.5). In the case of an unknown scale,  $\gamma(k, \theta) = \theta \gamma_0(k)$  and  $f(\omega, \theta) = \theta f_0(\omega)$ , which has also been discussed by Switzer (1984), we obtain  $\hat{\theta} = \sum \alpha_{0,k} \hat{\gamma}(k)$ .

#### 4. AN EXAMPLE

In this section, we will apply our techniques to a set of Landsat data. These data are measurements of the reflected energy in four spectral bands for each of 400 pixels arranged in a  $16 \times 25$  rectangular grid.

These data have been used by Switzer (1980). The area consists of three rock types, so the assumption of stationarity is doubtful. The simple model  $X_i$  is a function of rocktype at site  $i$  plus white noise is not adequate either. There seems to be some blurring, and other features of the ground like texture and orientation contribute also to the signal, so that the values in two regions of the same rocktype differ considerably. For this reason an intrinsic model looks plausible.

Since we deal only with univariate observations, we considered one of the four spectral bands. The estimated coefficients for three different neighbourhoods are given in Table 1. Since the expression  $L$  of formula (3.1) is an approximation to  $-2 \log$  likelihood per observation, we can select a model by minimizing  $400 L$  plus  $C$  times the number of parameters with  $C = 2$  (Akaike), or  $C = 2 \log \log 400 = 3.58$  (Hannan) or  $C = \log 400 = 5.99$  (Rissanen). It thus seems by all these criteria worthwhile to include more than just the four nearest neighbours although their coefficients are clearly the most important ones. There is also a clear anisotropy in the data. The difference between  $a_{1,1}$  and  $a_{1,-1}$  displaces the ridges in the spectrum somewhat from  $\omega_1 = 0$  and  $\omega_2 = 0$ , which seems rather peculiar.

Table 1. Estimated parameters of intrinsic autoregressions for Landsat data

Number of neighbours	$\hat{\sigma}^2$	$\hat{a}_k$ , for specified $k$						$400L(\hat{a}, \hat{\sigma}^2)$
		(0, 1)	(0, 2)	(1, -1)	(1, 0)	(1, 1)	(2, 0)	
4	7.35	0.346			0.154			931.7
8	6.05	0.385		-0.014	0.267	-0.138		905.0
12	5.74	0.442	-0.057	-0.002	0.206	-0.104	0.015	897.8

Finally, let us see how well these models fit the data and how they compare with corresponding stationary models. For this we fitted also a stationary autoregression with 12 neighbours. The sum of the coefficients is 0.4977; that is we obtain almost an intrinsic model. We then calculated numerically the theoretical semivariogram of the intrinsic and of the stationary autoregression for lags  $k$  with  $|k_1| \leq 6$ ,  $|k_2| \leq 6$ . For both models the agreement between theoretical and empirical semivariogram is surprisingly good, but for larger lags the autoregressive model gives always too small values. This indicates that the intrinsic model is to be preferred and gives a better description of the data.

#### ACKNOWLEDGEMENT

I should like to express my gratitude to Paul Switzer for discussions and the hospitality at Stanford in Summer 1985 where parts of this report were written. I am also indebted to Jörg Waldvogel from ETH Zürich for his advice on the numerical integration.

#### REFERENCES

- BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B* **36**, 192-236.
- BOX, G. E. P. & JENKINS, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day.
- GELFAND, I. M. & VILENKIN, N. YA. (1964). *Generalized Functions, 4, Applications of Harmonic Analysis*. New York and London: Academic Press.
- GUYON, X. (1982). Parameter estimation for a stationary process on a  $d$ -dimensional lattice. *Biometrika* **69**, 95-105.
- HENRICI, P. (1977). *Applied and Computational Complex Analysis, 2*. New York: Wiley.
- KÜNSCH, H. R. (1979). Gaussian Markov random fields. *J. Fac. Sci. Univ. Tokyo, Sec. IA*, **26**, 53-73.
- KÜNSCH, H. R. (1981). Thermodynamics and statistical analysis of Gaussian random fields. *Z. Wahr. verw. Geb.* **58**, 407-21.
- KÜNSCH, H. R. (1983). Approximations to the maximum likelihood equations for some Gaussian random fields. *Scand. J. Statist.* **10**, 239-46.
- MATHERON, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Prob.* **5**, 439-68.
- ROSANOV, Y. A. (1967). On Gaussian fields with given conditional distributions. *Theory Prob. Applic.* **12**, 381-91.
- SPITZER, F. (1964). *Principles of Random Walk*. Princeton: Van Nostrand.
- SWITZER, P. (1980). Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *J. Int. Assoc. Math. Geology* **12**, 367-76.
- SWITZER, P. (1984). Inference for spatial autocorrelation functions. In *Geostatistics for Natural Resources Characterization*, Ed. M. David and G. Verly, pp. 127-40. Dordrecht: Reidel.
- WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434-49.
- WILKINSON, G. N., ECKERT, S. R., HANCOCK, T. W. & MAYO, O. (1983). Nearest neighbour analysis of field experiments (with discussion). *J. R. Statist. Soc. B* **45**, 151-211.

[Received April 1986. Revised December 1986]