ON STATIONARY PROCESSES IN THE PLANE

BY P. WHITTLE

Applied Mathematics Laboratory, New Zealand Department of Scientific and Industrial Research

The sampling theory of stationary processes in space is not completely analogous to that of stationary time series, due to the fact that the variate of a time series is influenced only by past values, while for a spatial process dependence extends in all directions. This point is elaborated in §§2–4. The estimation and test theory developed in §7 is applied in §8 to uniformity data for wheat and oranges. The final section is devoted to an examination of some particular two-dimensional processes.

1. INTRODUCTION

The disturbing effect of topographic correlation on the results of field experiments, forest and crop surveys, sampling surveys of populated areas, etc., is well known, and it is recognized that we have here examples of two-dimensional stochastic processes. The physicists also encounter higher dimensional processes (for instance, in the studies of turbulence and of systems of particles) and have indeed been the principal investigators of the subject.

The processes we mentioned can only as a first approximation be regarded as stationary, if they can be so regarded at all. However, the approximation is satisfactory sufficiently often to make the study of the stationary type of process worth while.

Much of the two-dimensional theory is no more than a formal extension of that used in the study of time series, so we shall treat these aspects as briefly as possible. There is one interesting new feature, however, which is mentioned in the summary above, and we shall consider it at length.

The difficulty of examining particular models is in general greater in two dimensions than in one, simply because there are more mathematical obstacles of a technical nature. One seems to have left the domain of the elementary functions entirely. We shall indicate these obstacles, but take advantage of the fact that for many purposes they may be avoided.

For many applications it is sufficient to consider only purely non-deterministic processes, and we shall restrict our attention to processes of this type, more particularly to linear autoregressions.

2. THE LINE TRANSECT

It is useful to begin with a discussion of the simple *line transect* (i.e. a straight line laid over an area, along which observations are taken equidistantly). The observations of the transect may be regarded as being generated by a one-dimensional process, just as are the terms of a time series. However, there is an important difference between the two cases. At any instant in a time series we have the natural distinction of past and future, and the value of the observation at that instant depends only upon *past* values. That is, the dependence extends only in one direction: backwards. In the case of the transect, however, there is no such day and night distinction between the two directions, and dependence will extend both ways. We can consider an example for the more general two-dimensional case of a field: a dab of fertilizer applied at any point in the field will ultimately affect soil fertility in *all* directions. (Exceptions are of course possible; the field may have so strong a slope that only the area on the downward side of the spot will be affected.)

Denoting now observation and 'error' variates by ξ_t, ϵ_t respectively ($t = \ldots, -2, -1, 0, 1, 2, \ldots$), the simplest realistic time series model is perhaps the first-order autoregression

$$\xi_t = a\xi_{t-1} + \epsilon_t. \tag{1}$$

For the transect, however, model (1) would have to be regarded as constituting a degenerate case, in which dependence extended only in one direction. The simplest non-degenerate transect model would be

$$\xi_t = a\xi_{t-1} + b\xi_{t+1} + \epsilon_t, \tag{2}$$

where it is intuitively clear that a and b must not be too large. We shall term (2) a bilateral autoregression, in distinction to (1) which is a unilateral autoregression.

Consequences of the admission of the bilateral type of scheme become apparent when one considers the estimation of parameters. Whereas the parameter a in (1) may be estimated consistently by minimizing the residual sum of squares $\sum_t (\xi_t - a\xi_{t-1})^2$, an attempt to estimate a and b in (2) simply by minimizing

$$U = \sum_t (\xi_t - a\xi_{t-1} - b\xi_{t+1})^2, \tag{3}$$

leads to nonsensical results. One could explain the breakdown by saying that it is not legitimate to include ξ_{t+1} in the conditional mean of ξ_t when the value of ξ_{t+1} itself depends upon ξ_t. Formally, we may say that the Jacobian of the transformation from the ϵ_t to the ξ_t is not unity for relation (2) as it would be for a unilateral relation such as (1). It may be shown, however (and in § 7 will be proved for the more general two dimensional case), that the correct equations for the least-square estimates are obtained by minimizing kU, where k is a certain function of the parameters given by

$$\log k = -\frac{1}{2\pi}\int_0^{2\pi} \log (a\,e^{i\omega} - 1 + b\,e^{-i\omega})\,(a\,e^{-i\omega} - 1 + b\,e^{i\omega})\,d\omega. \tag{4}$$

3. Formal properties of the bilateral scheme

We shall briefly examine the properties of the transect model as a preparation for the discussion of the two-dimensional case. Let us consider the general bilateral linear autoregression

$$L(T)\,\xi_t = \epsilon_t, \tag{5}$$

where $L(T) = \Sigma a_j T^j$ and T is the translation operator

$$T\xi_t = \xi_{t+1}. \tag{6}$$

Equation (5) has the solution $$\xi_t = \frac{\epsilon_t}{L(T)} = \Sigma b_j \epsilon_{t+j}, \tag{7}$$

where b_j is the coefficient of $e^{ij\omega}$ in the Fourier expansion of $[L(e^{i\omega})]^{-1}$ (see Bartlett, 1946, p. 60). Using Fourier transforms it may similarly be shown that the spectral function of scheme (5) is

$$F(\omega) = \frac{\sigma^2(\epsilon)}{L(e^{i\omega})\,\overline{L(e^{-i\omega})}} \tag{8}$$

(cf. Doob, 1944; Daniell, 1946). The autocovariances

$$\phi(j) = \mathrm{cov}\,(\xi_t, \xi_{t+j}) \quad (j = 0, \pm 1, \pm 2, \ldots) \tag{9}$$

are generated as the coefficients in the Fourier expansion of $F(\omega)$. Thus, writing for convenience $e^{i\omega} = z$, $\sigma^2(\epsilon) = v$, we have

$$F(\omega) = \frac{v}{L(z)\,L(z^{-1})}. \tag{10}$$

It is as well to consider the conditions under which expressions (7) and (8) are valid, since a number of essential points are involved. If (5) is to represent a unilateral scheme, expressing 'dependence only on the past', then the Laurent expansion of $[L(z)]^{-1}$ on $|z| = 1$ must not involve positive powers of z, and we are led to the usual condition for the stability of a time-series autoregression: that all roots of $L(z) = 0$ fall *inside* the unit circle (Wold, 1938). If, however, one is prepared to admit bilateral schemes, then all that is required is that $[L(z)]^{-1}$ possess a convergent Laurent expansion—coefficients otherwise unrestricted. The stationarity condition is in this case greatly weakened; all that is required is that no root of $L(z) = 0$ fall *on* the unit circle.

While this is the only condition which need be imposed upon a particular $L(z)$ to secure stationarity, it is desirable that $L(z)$ be such that in relation (5) it is ξ_t which is the dependent variate (not, for example, ξ_{t+1}, as would be the case if relation (1) were modified to $a\xi_t = \xi_{t+1} - \epsilon_t$). This may be achieved by multiplying $L(z)$ by some integral power of z (or, what is the same thing, redefining the ϵ sequence by translation so that ϵ_t comes into correspondence with ξ_t). For, suppose that as z describes the unit circle once in a positive direction $L(z)$ encircles the origin in the complex plane r times (where r is necessarily integral). The normalization will then be achieved by translating the ϵ sequence r steps backwards, so that $L(T)$ is replaced by the normalized operator $T^{-r}L(T)$, or $L^*(T)$, say. Since the origin is not a branch point for $\log L^*(z)$ (as it was for $\log L(z)$ if r was not zero) then $\log L^*(z)$ may be expanded in a Laurent series in z, and the operator $L^*(T)$ may be represented

$$L^*(T) = e^{\Sigma c_j T^j}. \tag{11}$$

It will be assumed in the sequel that this normalization has been performed, so that the representation (11) is always possible.

4. Indeterminacies of estimation

If one is prepared to admit bilateral schemes, then the autoregression exhibits the same kind of ambiguity that Wold (1938) has shown to obtain for the moving average. Thus, suppose that one is given a set of autocovariances which could have been generated by an autoregression of order p, and that one is set the task of determining the autoregression, i.e. of determining the polynomial $L(T)$ of (5). It is thus necessary to find an $L(z)$ which will satisfy equation (10), where $F(\omega)$ is determined by the given autocovariances. Such an $L(z)$ may be chosen in anything up to 2^p ways, since if a particular $L(z)$ has roots $\alpha_1, \alpha_2, \ldots, \alpha_p$, then there are 2^p possible finite autoregressions, corresponding to the 2^p possible representations of $F(\omega)$:

$$F(\omega) = \frac{\text{const.}}{|(z^{\pm 1} - \alpha_1)(z^{\pm 1} - \alpha_2)\ldots(z^{\pm 1} - \alpha_p)|^2}. \tag{12}$$

Two of these schemes are unilateral, and differ only in that the 'time' axis runs in opposite directions.

A similar argument applies for the least square fitting of a pth order autoregression to a set of data, since here again one works from the autocovariances, which could as easily have stemmed from one of the 2^p fitted autoregressions as from another.

For the sake of example, let us examine once more the bilateral scheme (2). If the scheme has been normalized according to the last paragraph of § 3 then the equation

$$a - z + bz^2 = 0 \tag{13}$$

will have one root inside the unit circle and one outside. Let us denote the roots α, β^{-1}, $[|\alpha| < 1, |\beta| < 1]$. Defining coefficients A and B by the relation

$$(z-\alpha)(z-\beta) = z^2 + Az + B, \tag{14}$$

we can construct an autoregression

$$\xi_t + A\xi_{t-1} + B\xi_{t-2} = \epsilon_t, \tag{15}$$

which is unilateral and yet generates the same autocorrelations as (2).

Evaluating integral (4) we find that a and b will be estimated by minimizing the quantity

$$[1+\sqrt{(1-4ab)}]^{-2}\,\Sigma(\xi_t - a\xi_{t-1} - b\xi_{t-2})^2$$
$$\approx [1+\sqrt{(1-4ab)}]^{-2}\,[(1+a^2+b^2)\,C_0 - 2(a+b)\,C_1 + 2abC_2], \tag{16}$$

where C_s is the observed autocovariance of lag s. Transforming to parameters A and B we find expression (16) proportional to

$$(1+A^2+B^2)\,C_0 + 2(A+AB)\,C_1 + 2BC_2 \approx \text{const.}\,\Sigma(\xi_t + A\xi_{t-1} + B\xi_{t-2})^2, \tag{17}$$

which is another indication of the equivalence of schemes (2) and (15). The practical significance of this equivalence is that the minimization of the rather awkward expression (16) may be replaced by that of the much simpler expression (17). The estimates of a and b may then be calculated from those of A and B by means of relation (14).

It may seem unnecessary to introduce the bilateral type of scheme when any such scheme may effectively be reduced to a unilateral one. Indeed, the step appears positively undesirable when we consider how much more complicated and indeterminate the parameter estimates are for a bilateral model. We shall see, however, that in the two-dimensional case the reduction to a unilateral scheme complicates matters very much, in contrast to the one-dimensional case we have just considered. There is thus no escaping the explicit introduction of a dependence in all directions. A further point is relevant: it is the multilateral scheme which in general corresponds to reality, even in those cases for which the formal work of estimation, etc., is simplest performed using an equivalent unilateral model.

5. GENERALITIES ON THE TWO-DIMENSIONAL PROCESS

One can consider continuous processes in which the variate has a value at every point of the plane, or discrete processes, which are usually such that the variate is observed only at the points of a rectangular plane lattice. We shall for the most part consider the discrete case, as being simpler mathematically and of greater practical interest. The observed and error variates will be denoted ξ_{st} and ϵ_{st} respectively ($s, t = \ldots, -2, -1, 0, 1, 2, \ldots$). When discussing continuous processes this notation will be modified to $\xi(x.y)$ and $\epsilon(x, y)$ (x and y assuming all real values).

The particular model which we shall consider almost exclusively is the two-dimensional linear autoregression, which we shall write

$$L(T_s, T_t)\,\xi_{st} = \epsilon_{st}, \tag{18}$$

where T_s and T_t are translation operators defined by

$$T_s\xi_{st} = \xi_{s+1,t}, \quad T_t\xi_{st} = \xi_{s,t+1} \tag{19}$$

and

$$L(T_s, T_t) = \sum_j \sum_k a_{jk} T_s^j T_t^k. \tag{20}$$

Corresponding to equations (7)–(10) we have

$$\xi_{st} = \frac{\epsilon_{st}}{L(T_s, T_t)} = \sum_j \sum_k b_{jk} \epsilon_{s+j,t+k}, \tag{21}$$

$$F(\omega_1, \omega_2) = \frac{v}{L(z_1, z_2)\, L(z_1^{-1}, z_2^{-1})} \tag{22}$$

$$= \sum_j \sum_k \phi(j,k) z_1^j z_1^k, \tag{23}$$

where b_{jk} is the coefficient of $z_1^j z_2^k$ in the Fourier expansion of $[L(z_1, z_2)]^{-1}$, $(z_1 = e^{i\omega_1}, z_2 = e^{i\omega_2})$, v is the variance of the ϵ's, $\phi(j,k)$ the covariance of ξ_{st} and $\xi_{s+j,t+k}$, and $F(\omega_1, \omega_2)$ the corresponding spectral function, as defined by (23).

As before, it is necessary and sufficient for the validity of (21) and (22) that $L(z_1, z_2)$ be not zero for any z_1 and z_2 which simultaneously satisfy $|z_1| = 1$, $|z_2| = 1$ (at least when the autoregression (18) is finite). We shall assume that the operator L has been so normalized that $L(z_1, z_2)$ will not circle the origin in the complex plane as either z_1 or z_2 moves around the unit circle. (If L does not circle the origin as z_1 traces out the unit circle and z_2 holds a particular value, neither will it do so for any other value of z_2. For if it should do so there would be some intermediate value of z_2 for which L would move *through* the origin, against hypothesis.)

6. Unilateral representation of the two-dimensional process

We shall prove, by a type of argument which is now well known (due to Wiener; see, for example, Wiener, 1949, p. 78), that a unique process may be found which will generate a given set of autocorrelations and in which ξ_{st} is expressed as an autoregression upon ξ_{su} $(u > t)$ and ξ_{vw} $(v > s, w$ unrestricted). That is, in the lattice of Fig. 1 the value at the white dot may be expressed in terms of the values at the black dots. The given autocorrelogram may obviously not be completely arbitrary, but the necessary conditions are very mild.

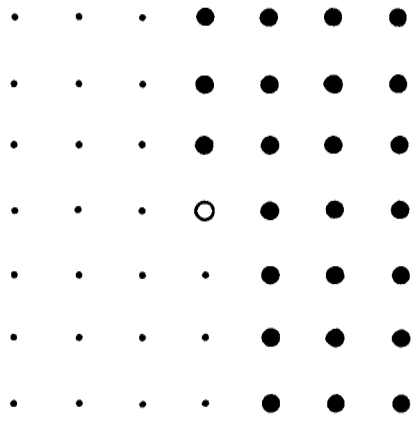

Fig. 1

The first condition is that the autocorrelations be such that they could have been generated by a purely non-deterministic process, which implies among other things that $\log F(\omega_1, \omega_2)$ has a Fourier expansion, so that $F(\omega_1, \omega_2)$ may be represented

$$F(\omega_1, \omega_2) = \exp\left[\sum_j \sum_k \alpha_{jk} z_1^j z_2^k\right]. \tag{24}$$

Define now the function

$$P(z_1, z_2) = \exp -\left[\frac{\alpha_{00}}{2} + \sum_{k=1}^{\infty} \alpha_{0k} z_2^k + \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} \alpha_{jk} z_1^j z_2^k\right]. \tag{25}$$

Since

$$F(\omega_1, \omega_2) = \frac{1}{P(z_1, z_2)\, P(z_1^{-1}, z_2^{-1})}, \tag{26}$$

the autoregression

$$P(T_s, T_t)\, \xi_{st} = \epsilon'_{st} \quad (\sigma^2(\epsilon') = 1) \tag{27}$$

will have spectral function $F(\omega_1, \omega_2)$, as required. Further, the expansion of $P(T_s, T_t)$ will involve the same powers of T_s and T_t as occur of z_1 and z_2 in expression (25), so that the autoregression is of the required form. To ensure that (27) may validly be written out as an autoregression we include a second condition: that $P(e^{i\omega_1}, e^{i\omega_2})$ possess a Fourier expansion.

Representation (27) corresponds to the unilateral representation in the case of a transect. It is by no means as useful, however, and for such purposes as estimation one is in general better advised to work with the original model. For one thing, it is not true in two dimensions as it is in one, that the unilateral representation of a finite autoregression is also a finite autoregression. This is confirmed by the following simple example: the finite autoregression

$$(1 + \beta^2)\, \xi_{st} = \beta(\xi_{s+1,t} + \xi_{s,t+1} + \xi_{s,t-1}) + \epsilon_{st} \tag{28}$$

has a unilateral representation which is infinite:

$$\xi_{st} = 2\beta\xi_{s,t+1} - \beta^2\xi_{s,t+2} - \beta^2\xi_{s+1,t+1} + \beta(1 - \beta^2) \sum_0^{\infty} \beta^j \xi_{s+1,t-j} + \epsilon'_{st}. \tag{29}$$

Further, the real usefulness of the unilateral representation is that it suggests a simplifying change of parameters (although the same parameter transformation is usually suggested in the evaluation of integral (4) or its two-dimensional equivalent). For most two-dimensional models, however, the appropriate transformation, even if evident, is so complicated that nothing is gained by performing it. For example, to calculate the unilateral representation of the model

$$\xi_{st} = \alpha(\xi_{s+1,t} + \xi_{s-1,t} + \xi_{s,t+1} + \xi_{s,t-1}) + \epsilon_{st} \tag{30}$$

one must evaluate the coefficients in the Fourier expansion of $\log[1 - \alpha(z_1 + z_1^{-1} + z_2 + z_2^{-1})]$, and these are not expressible in terms of anything simpler than elliptic integrals.

7. SAMPLING THEORY

In this section we shall consider the sampling theory of the discrete process, deriving general estimation equations, tests of fit, and formulae for the asymptotic variances and covariances of parameter estimates. The pattern of development is similar to that of previous articles on time-series analysis (Whittle, 1951–3), apart from the fact that special attention must be paid to the effect of multilaterality. The particular processes considered are those which are purely non-deterministic and for which $[F(\omega_1, \omega_2)]^{-1}$ is not zero for any real ω_1, ω_2, so that the process is representable as an autoregression. We shall also in general assume the

variates normally distributed. Estimation equations thus obtained on the maximum-likelihood criterion can be regarded as least-square estimation equations when the variate distribution is not normal, but the extent to which remaining results are valid in such a case is a matter for further inquiry.

Let us suppose that we have a series of mn observations ξ_{st} ($s = 1, 2, \ldots, m$; $t = 1, 2, \ldots, n$), so that the empirical covariance of lag j, k is

$$C_{jk} = \frac{1}{mn} \sum_{s=1}^{m-j} \sum_{t=1}^{n-k} \xi_{st} \xi_{s+j, t+k}. \tag{31}$$

We shall uniformly neglect end-effects, and so have divided in (31) by mn instead of $(m-j)(n-k)$. Let us further define the quantity

$$\begin{aligned} f(\omega_1, \omega_2) &= \frac{1}{mn} \{[\Sigma\Sigma \xi_{st} \cos(s\omega_1 + t\omega_2)]^2 + [\Sigma\Sigma \xi_{st} \sin(s\omega_1 + t\omega_2)]^2\} \\ &= \sum_{-m}^{m} \sum_{-n}^{n} C_{jk} \cos(j\omega_1 + k\omega_2) \\ &= \sum_{-m}^{m} \sum_{-n}^{n} C_{jk} e^{i(j\omega_1 + k\omega_2)}, \end{aligned} \tag{32}$$

which we see on comparison with (23) to be the *empirical spectral function*, and in fact the appropriate extension of the Schuster periodogram to two dimensions.

At the end of the section we shall prove the following basic result: that if the variates ξ_{st} are normally distributed with zero mean, and are generated by a stationary process of the type considered with spectral function $F(\omega_1, \omega_2)$, then their joint likelihood is given, apart from end-effects, by the expression

$$p(\xi) = \frac{1}{(2\pi V)^{\frac{1}{2}(mn)}} \exp\left[-\frac{mn}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \frac{f}{F} d\omega_1 d\omega_2\right], \tag{33}$$

where

$$V = \exp\left[\frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \log F \, d\omega_1 d\omega_2\right]. \tag{34}$$

Taking the logarithm of (33) we see that the maximum-likelihood estimates of the parameters of F are obtained by minimizing

$$\mathfrak{L} = \frac{1}{4\pi^2} \iint \left[\log F + \frac{f}{F}\right] d\omega_1 d\omega_2. \tag{35}$$

Even if the variates are not considered to be normally distributed, expression (35) will upon minimization yield the least-square estimates (see Whittle, 1953, p. 132).

Expression (35) actually does not lead to the most convenient form of the estimation equations, this is obtained by expressing $\mathfrak{L}$ in terms of the autocovariances rather than the periodogram

$$\mathfrak{L} = \frac{1}{4\pi^2} \iint \log F \, d\omega_1 d\omega_2 + \Sigma\Sigma c_{jk} C_{jk}, \tag{36}$$

where the coefficients c are given by

$$\Sigma\Sigma c_{jk} z_1^j z_2^k = \frac{1}{F(\omega_1, \omega_2)}. \tag{37}$$

Consider now the special case of a process generated by a stochastic difference equation of type (18) so that

$$F = \frac{v}{L(z_1, z_2)\, L(z_1^{-1}, z_2^{-1})}. \tag{38}$$

Inserting this relation in (36) we obtain

$$\mathfrak{L} = \log v + \log k + U/mnv, \tag{39}$$

where

$$U = \sum_1^m \sum_1^n \epsilon_{st}^2 = \sum_1^m \sum_1^n [L(T_s, T_t)\, \xi_{st}]^2 = \sum_j \sum_k c_{jk} C_{jk} \tag{40}$$

and

$$\log k = -\frac{1}{4\pi^2} \iint \log [L(z_1, z_2)\, L(z_1^{-1}, z_2^{-1})]\, d\omega_1 d\omega_2. \tag{41}$$

The minimized value of (39) with respect to the irrelevant parameter v is $\log(kU/mn)$, so that the quantity to be minimized with respect to the parameters of L is kU. This is the result which is of most use in investigations of the present type: that the least-square estimates are yielded by the minimization of the usual 'residual sum of squares' times a function of the parameters, k. We see from (41) that $-\frac{1}{2}\log k$ may be interpreted as the absolute term in the double Fourier expansion of $L(e^{i\omega_1}, e^{i\omega_2})$, a fact that may sometimes ease its evaluation.

Suppose that F contains the unknown parameters $\theta_1, \theta_2, \ldots, \theta_p$ (v being included among these, if, as is usual, its value is unknown). Then, following a familar line of reasoning (see Whittle, 1953, p. 135, for an application similar to the present one) that the asymptotic covariance matrix of the least square estimates of $\theta_1, \theta_2, \ldots, \theta_p$ is

$$\frac{2}{mn}\left[\frac{1}{4\pi^2}\iint \frac{\partial \log F}{\partial \theta_j}\frac{\partial \log F}{\partial \theta_k}\, d\omega_1 d\omega_2\right]^{-1}. \tag{42}$$

Suppose again that when the p parameters have been fitted the minimized value of kU is $(kU)_p$, and that when an additional q parameters have been fitted the quantity falls to $(kU)_{p+q}$. Then we can test the improvement attained by the introduction of the extra q parameters by using the fact that, if the initial p-parameter hypothesis were correct, then

$$\psi^2 = (mn - p - q) \log \frac{(kU)_p}{(kU)_{p+q}} \tag{43}$$

would be asymptotically distributed as χ^2 with q degrees of freedom. The argument is again an old one (cf. Whittle, 1952, 1953).

Proof of relation (33). Let the process under consideration have the unilateral representation (27). The joint frequency function of the mn residuals ϵ'_{st} ($s = 1, 2, \ldots, m; t = 1, 2, \ldots, n$) is

$$p(\epsilon') = \frac{1}{(2\pi)^{\frac{1}{2}(mn)}} \exp\left[-\frac{1}{2}\sum_1^m \sum_1^n (\epsilon'_{st})^2\right]. \tag{44}$$

Performing the linear transformation to which (27) is equivalent, we obtain (neglecting end effects) the following expression for the frequency function of the ξ_{st}:

$$\begin{aligned} p(\xi) &\approx \frac{A^{mn}}{(2\pi)^{\frac{1}{2}(mn)}} \exp\left[-\frac{1}{2}\sum_1^m \sum_1^n [P(T_s, T_t)\, \xi_{st}]^2\right] \\ &\approx \frac{A^{mn}}{(2\pi)^{\frac{1}{2}(mn)}} \exp\left[-\frac{mn}{2}\Sigma\Sigma c_{jk} C_{jk}\right] \\ &\approx \frac{A^{mn}}{(2\pi)^{\frac{1}{2}(mn)}} \exp\left[-\frac{mn}{8\pi^2}\iint \frac{f}{F}\, d\omega_1 d\omega_2\right], \end{aligned} \tag{45}$$

where A is the coefficient of ξ_{st} in (27). By (24) and (25)

$$A = e^{-\frac{1}{2}\alpha_{\infty}} = \exp\left[-\frac{1}{8\pi^2}\iint \log F \, d\omega_1 d\omega_2\right], \tag{46}$$

which on insertion in (45) leads to (33).

The nature of the approximate equality in (45) may be specified more exactly. If the right-hand member in (45) is denoted p', then the relation obtaining is that $\log p$ and $\log p'$ are asymptotically equal (neglect of end-effects leading to neglect of terms in the exponent). This kind of approximation is quite satisfactory, since the addition of a term of relative order m^{-1} or n^{-1} to the logarithm of the likelihood will for large m and n have little effect on either the values of the maximum-likelihood estimates, or the significance points of likelihood ratios.

8. Numerical examples

Neither of the two sets of numerical data to be examined provides a picture book example of the particular models we have been studying, but they are perhaps the more valuable for just that reason.

The first set of data concern a uniformity trial on wheat (yield of grain) conducted by Mercer & Hall (1911). It involves 500 plots, each 11 ft by 10·82 ft., arranged in a 20×25 rectangle, plot totals constituting the observations. The first and fourth quadrants of the correlation field are given in Table 1; the second and third quadrants can be filled in by means of the relation

$$\phi(-j, -k) = \phi(j, k).$$

We note that correlations along the north-south (s) axis are considerably stronger than those along the east-west (t) axis, at least part of the explanation presumably being that the plots are not square. Correlations are generally higher in the north-east direction than in the north-west, indicating a definite directional effect. Another fact worthy of note is that correlations do not decrease monotonically as one moves out from the origin, instead they dip to a minimum and then rise again. One could imagine competition between neighbouring plants producing such an effect if the distances were smaller, but the more likely explanation in the present case is that there are 'waves of fertility' of the kind often remarked in a ploughed field (see, for example, Neyman, 1952, p. 75).

In Table 2 we summarize the various schemes considered, giving the fitted coefficients and the corresponding values of k, U and kU.

Since all kU are near to 0·7 we can say roughly that if the introduction of a new parameter depresses kU by Δ, then by equation (43) Δ is significant at the 5 % level if

$$3{\cdot}841 < 500 \log\left(\frac{0{\cdot}7 + \Delta}{0{\cdot}7}\right)$$

or $\Delta > 0{\cdot}0054$. The corresponding limits if two or three parameters have been introduced are 0·0084 and 0·0110. These limits provide useful rules of thumb for judging significances of differences in kU in the table. Thus, hypotheses 1 and 3 do not differ significantly in degree of fit, while 1 and 4 very definitely do. Hypotheses such as 1 and 5 cannot strictly be compared in this manner, since neither hypothesis is a special instance of the other. However, one would be very much inclined to say that the fit of 1 is superior to that of 5, and this conclusion is borne out if one compares each in turn with the more general hypothesis 7.

The surprising feature of the results is that the simple unilateral scheme 1 fits the data so much better than the symmetric second order schemes 5 and 6. Thus, although the theme of this article has largely been that spatial processes will in general be multilateral, our first example appears to be dominantly unilateral. The reason for this becomes apparent upon a re-examination of the correlations of Table 1. These correlations fall steeply as soon as a lag is introduced (to 0·52 on the s axis and 0·29 on the t axis), but decrease only slowly as the lag is increased (the corresponding figures for a lag of two units are thus 0·41 and 0·15). However, it will be seen in § 9 that the correlogram of a scheme such as 5 or 6 decays smoothly right from the origin, having in fact zero derivative there (cf. Fig. 2), so that neither 5 nor 6 can possibly fit the observations well. There are at least two possible explanations of this behaviour of the observed correlogram. The underlying scheme may in fact be a unilateral

Table 1. *Autocorrelations for the wheat data*, $s = 0$ *to* 4, $t = -3$ to $+3$

t	$s=0$	$s=1$	$s=2$	$s=3$	$s=4$
−3	0·1880	0·1602	0·1509	0·1276	0·1352
−2	0·1510	0·0234	0·0020	−0·0137	−0·1039
−1	0·2923	0·1853	0·1349	0·0788	0·0878
0	1·0000	0·5252	0·4055	0·3639	0·3561
1	0·2923	0·2354	0·1799	0·1205	0·1399
2	0·1510	0·1285	0·0999	0·0749	0·0859
3	0·1880	0·1935	0·2483	0·2415	0·2284

Table 2. *Details of the models fitted*

Model no.	$L(T_s, T_t)$	k	U	kU
1	$1-0{\cdot}488T_s-0{\cdot}202T_t$	1	0·6848	0·6848
2	$1-0{\cdot}483T_s-0{\cdot}179T_t^{-1}$	1	0·6940	0·6940
3	$1-0{\cdot}492T_s-0{\cdot}211T_t+0{\cdot}019T_sT_t$	1	0·6845	0·6845
4	$1-0{\cdot}402T_s-0{\cdot}168T_t-0{\cdot}172T_s^2-0{\cdot}092T_t^2$	1	0·6564	0·6564
5	$1-0{\cdot}159(T_s+T_s^{-1}+T_t+T_t^{-1})$	1·1240	0·6508	0·7314
6	$1-0{\cdot}213(T_s+T_s^{-1})-0{\cdot}102(T_t+T_t^{-1})$	1·1332	0·6217	0·7045
7	$1-0{\cdot}488T_s+0{\cdot}030T_s^{-1}-0{\cdot}202T_t-0{\cdot}034T_t^{-1}$	0·9843	0·6816	0·6709

one, due to e.g. the presence of a slope in the ground or a prevailing wind. Alternatively, we must recall that the observations are not *point* observations of growth, but *integrated* observations of the growth over an area. An integration such as this will enhance the autocovariance of zero lag relative to the others, and the correlogram of a scheme such as 6, when thus distorted, would not be dissimilar to the observed correlogram. However, the consideration of such an effect would lead us too far for present purposes, and we shall take the observations at their face value.

Scheme 1 is not completely satisfactory, in that it does not explain the dip and rise in the correlation for increasing distance. The inclusion of second-order terms in the same direction as the first-order terms improves the fit considerably, see scheme 4.

Unilateral schemes such as 1–4 can be fitted directly by least square regression methods, since for them $k = 1$. The fitting of multilateral schemes such as 5–7 is more difficult, however. The evaluation of U according to (40) is direct, it is the evaluation of k from (41) which presents new problems. For the present calculations the pedestrian methods of series expansion and numerical integration have been employed, although it is likely that better methods exist. Thus, for a scheme

$$\xi_{st} = \alpha\xi_{s+1,t} + \beta\xi_{s-1,t} + \gamma\xi_{s,t+1} + \delta\xi_{s,t-1} + \epsilon_{st}, \tag{47}$$

we have $\log(k) =$ minus twice absolute term in $\log[1 - \alpha z_1 - \beta z_1^{-1} - \gamma z_2 - \delta z_2^{-1}]$

$$= \sum_{j=1}^{\infty} \sum_{k=0}^{j} \frac{(2j)!}{j[k!\,(j-k)!]^2} (\alpha\beta)^k (\gamma\delta)^{j-k}. \tag{48}$$

For $\alpha\beta = \gamma\delta$ this reduces to

$$\log(k) = \sum_{j=1}^{\infty} \frac{1}{j} \binom{2j}{j}^2 (\alpha\beta)^j. \tag{49}$$

Expansions (48) and (49) are useful as long as $\alpha\beta$ and $\gamma\delta$ are very small, but convergence becomes slow as these quantities approach their maximum value of $\frac{1}{16}$. In Table 3 the value of $\log k$ in (49) is given for some specimen values of $\theta = \sqrt{(\alpha\beta)}$; these values have been obtained by numerical integration, and vary slowly enough to permit quite a fair graphical interpolation.

Table 3. *Specimen values of the correcting factor* $\log(k)$

θ	0·00	0·05	0·10	0·15	0·20	0·22	0·25
$\log(k)$	0·0000	0·0076	0·0420	0·1010	0·2028	0·2656	0·4406

Table 4. *Autocorrelations for the orange data,* $s = 0$ to 9, $t = -4$ to 4.

t	Values of s									
	0	1	2	3	4	5	6	7	8	9
−4	0·3912	0·3902	0·3771	0·3581	0·3710	0·2923	0·2549	0·2546	0·2440	0·2131
−3	·4609	·3956	·3930	·3696	·3484	·3353	·3137	·2834	·2894	·2870
−2	·4667	·4159	·3982	·3470	·3243	·3225	·3250	·2801	·2721	·2752
−1	·5462	·4669	·4336	·3854	·3880	·3761	·3495	·2914	·2567	·2606
0	1·0000	·5403	·5052	·4840	·4585	·4233	·3960	·3246	·3210	·3150
1	0·5462	·4458	·4280	·3751	·3914	·3622	·3716	·2948	·2958	·2925
2	·4667	·3858	·3799	·3459	·3603	·3475	·3401	·2905	·2766	·2230
3	·4609	·4190	·3716	·3764	·3509	·3336	·3369	·2765	·2433	·2741
4	·3912	·3543	·3882	·3392	·3537	·3446	·3305	·2893	·2685	·2321

The models of Table 2 were fitted simply by inserting trial values of the parameters and calculating the corresponding kU. Improved values were then found by approximating kU in the neighbourhood of its minimum by a paraboloid, and then locating this approximate minimum, whereupon the process was repeated. The work is tedious, but not unreasonably so if one considers the effort which the data themselves represent.

Our second example is derived from a uniformity trial of 1000 orange trees by Batchelor & Reed (1924), the trees being arranged in a 20×50 rectangular lattice. The correlation field $\rho(s,t)$ is given in Table 4, and the correlations along the s axis plotted in Fig. 2. The correlations in the two quadrants follow a very similar pattern, indicating a high degree of symmetry in the original experimental field. A feature of interest is, that while the correlations $\rho(s,t)$ for the most part change smoothly in value as one moves through adjacent values of s and t, there is a sharp discontinuity at the origin. The correlation surface thus consists of a smoothly rounded broad dome with a spike at the apex. This suggests that the yield of an individual tree, Y_{st}, may be represented

$$Y_{st} = \xi_{st} + \eta_{st}, \tag{50}$$

where ξ_{st} obeys a process with smooth correlogram, while distinct η_{st} are uncorrelated with one another (or with the ξ_{st}) so that the η correlogram consists simply of a 'spike' at the origin. A natural interpretation of (50) springs to mind: that ξ represents the fertility of the

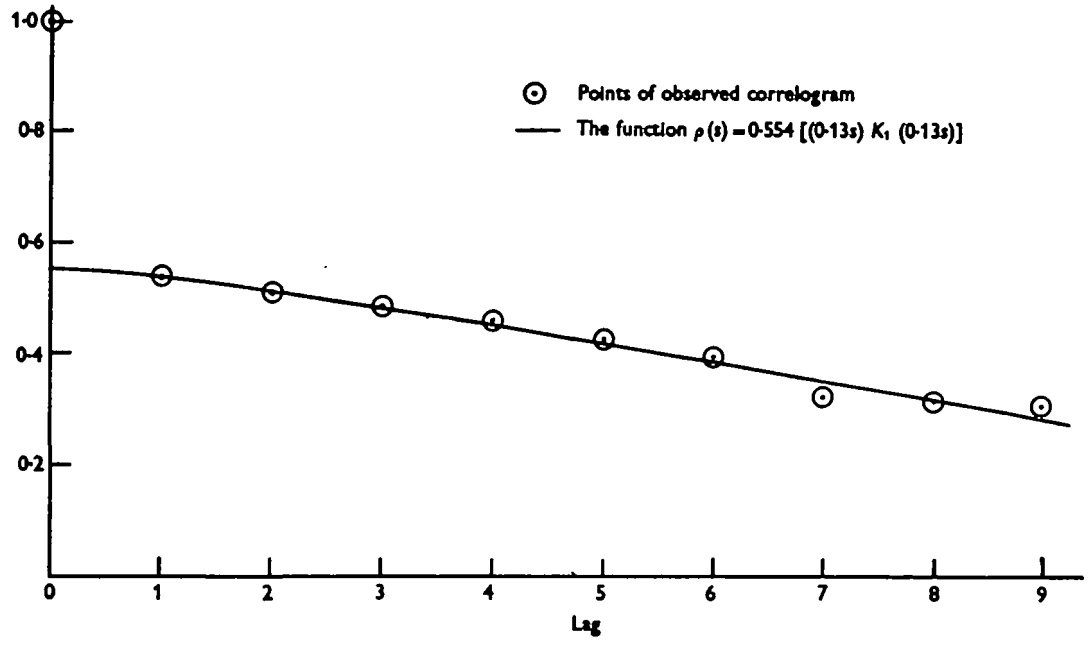

Fig. 2

soil around the tree, and presumably obeys some simple symmetric process such as (30), while η represents the intrinsic fruitfulness of the individual tree. Looking at the correlogram, one would say that soil variability and tree variability contribute roughly 56 and 44 % of the total variation respectively. However, one should not forget the possibility that the tree may 'integrate' the fertility of the surrounding soil as in the previous example.

It is a much more tedious matter to fit a compound scheme such as (50) plus (30) than to fit a simple autoregression such as (30). Supposing var $(\eta) = A$ and var $(\epsilon) = B$ we have

$$F(\omega_1, \omega_2) = A + \frac{B}{[1 - \alpha(z_1 + z_1^{-1} + z_2 + z_2^{-1})]^2}. \tag{51}$$

The quantities required for estimation purposes are $\log V$ and the coefficients in the expansion of F^{-1} (see (36)). We find

$$\begin{aligned}\log V = {} & +\frac{1}{4\pi^2}\iint \log F\, d\omega_1 d\omega_2 = \log(A+B) + \sum_{j=1}^{\infty} \frac{1}{j}\binom{2j}{j}\alpha^{2j}t \\ & + 4(\alpha\beta)^2[1-2\beta] \\ & + 36(\alpha\beta)^4[-\tfrac{1}{2}+4\beta-4\beta^2] \\ & + 400(\alpha\beta)^6[\tfrac{1}{3}-6\beta+16\beta^2-\tfrac{32}{3}\beta^3] \\ & + 4900(\alpha\beta)^8[-\tfrac{1}{4}+8\beta-40\beta^2+60\beta^3-32\beta^4] \\ & + 63504(\alpha\beta)^{10}[\tfrac{1}{5}-10\beta+80\beta^2-224\beta^3+256\beta^4-\tfrac{1024}{10}\beta^5] \\ & + \ldots; \end{aligned} \tag{52}$$

and

$$\begin{aligned} F^{-1} &= \Sigma\Sigma c_{jk} z_1^j z_2^k \\ &= \frac{1}{A+B}[1+(2\beta-2)S+(1-5\beta+4\beta^2)S^2+(4\beta-12\beta^2+8\beta^3)S^3 \\ &\qquad +(-\beta+13\beta^2-28\beta^3+16\beta^4)S^4+\ldots], \end{aligned} \tag{53}$$

where $$\beta = \frac{A}{A+B}, \quad S = \alpha(z_1+z_1^{-1}+z_2+z_2^{-1}).$$

Since the scheme is not a pure autoregression, F^{-1} is not a polynomial and the coefficients c_{jk} do not terminate, so that in the present case we find that we should have to include correlations of up to about lag 10 in order to obtain a sufficient approximation to the sum $\Sigma\Sigma c_{jk}C_{jk}$ of (36). However, correlations are only available up to $t = 4$, to continue would be a very arduous matter.

In general, if the scheme is not purely an autoregressive one it will be possible to apply least square methods only if the observed correlogram decays quite quickly. In a case such as the present one it becomes necessary to use more primitive methods of fitting: e.g. equation of observed and theoretical autocorrelation coefficients.

On the other hand, a slow rate of decay has the advantage that the discrete model may be approximated by a continuous one, which is in most cases a simplification. Thus, if in scheme (30) α has a value approaching 0·25 (corresponding to a very slow decay of dependence) then it is shown in § 9 that the scheme may be approximated by a continuous one with autocovariance function

$$\rho(r) = \text{const.}\, rK_1(\kappa r), \tag{54}$$

where r is the distance between the two points considered, K_1 a modified Bessel function, and

$$\kappa = \sqrt{\left(\frac{1}{\alpha}-4\right)}. \tag{55}$$

The thin line of the correlogram (Fig. 2) indicates the values of function (55) when the constants have been adjusted for coincidence at r_1 and r_8. It is found that $\kappa = 0{\cdot}13$, whence $\alpha = 0{\cdot}2489$.

The agreement looks impressive, but must be discounted considerably, since almost any monotone decreasing function would fit the observed curve reasonably well if only the end-points were arrange to coincide. However, if one fits, for example, an exponential curve in the same fashion, agreement is not at all as good, the exponential curve sagging too much in the middle.

9. SPECIAL PROCESSES

Strictly, one does not need to be acquainted with the properties of a model in order to fit it to a set of data (cf. the examples of the previous section). However, a knowledge of these properties does help one to arrive at a good model quickly, to say the least. In the present context 'properties' is usually synonymous with 'correlogram', since it is most often the correlogram one uses to characterize a given set of data. We shall now examine a few special models and their correlograms, but in a rather unsystematic fashion, since few results of an explicit nature are obtainable.

The simplest model of interest is the degenerate autoregression

$$\xi_{st} = \alpha\xi_{s+1,t} + \beta\xi_{s,t+1} + \epsilon_{st}, \tag{56}$$

with 'solution'

$$\xi_{st} = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{j+k}{j}\alpha^j\beta^k\epsilon_{s+j,t+k}. \tag{57}$$

For the process to be stationary, ξ_{st} being taken as the dependent variate, one must have $|\alpha|+|\beta|<1$ (see § 5). The autocovariances are generated by

$$(1-\alpha z_1-\beta z_2)^{-1}(1-\alpha z_1^{-1}-\beta z_2^{-1})^{-1},$$

and yield correlations

$$\rho(s,0)=A^s, \quad \rho(0,t)=B^t,$$

$$\rho(s,t)=\left[\sum_{j=0}^{s}\binom{s+t-j}{t}\alpha^{s-j}\beta^t A^j+\sum_{k=0}^{t}\binom{s+t-k}{s}\alpha^s\beta^{t-k}B^k\right], \tag{58}$$

where

$$\left.\begin{aligned} A &= \frac{1+\alpha^2-\beta^2-\Delta}{2\alpha}, \quad B=\frac{1+\beta^2-\alpha^2-\Delta}{2\beta}, \\ \Delta &= \sqrt{\{(1+\alpha+\beta)(1+\alpha-\beta)(1-\alpha+\beta)(1-\alpha-\beta)\}}. \end{aligned}\right\} \tag{59}$$

Formula (58) holds for $s, t>0$.

The continuous analogue of scheme (56) would be the first-order stochastic differential equation, formally written

$$\left(\alpha\frac{\partial}{\partial x}+\beta\frac{\partial}{\partial y}+\gamma\right)\xi(x,y)=\epsilon(x,y). \tag{60}$$

Rotating the plane to new axes $X=0$, $Y=0$, inclined at an angle $\tan^{-1}(\beta/\alpha)$ to the old ones, we find the relation may be written

$$\left(\sqrt{(\alpha^2+\beta^2)}\frac{\partial}{\partial X}+\gamma\right)\xi'(X,Y)=\epsilon'(X,Y), \tag{61}$$

where $\xi'(X,Y)=\xi(x,y)$, $\epsilon'(X,Y)=\epsilon(x,y)$. That is, scheme (60) may be regarded as a series of Markoff processes running side by side in the direction $\tan^{-1}(\beta/\alpha)$ and independent one of the other. This indicates the degenerate nature of the first-order scheme.

The simplest second-order scheme (and thus the simplest non-degenerate scheme) is the symmetric autoregression

$$\xi_{st}=\alpha(\xi_{s+1,t}+\xi_{s-1,t}+\xi_{s,t+1}+\xi_{s,t-1})+\epsilon_{st}. \tag{62}$$

Written in the form

$$\left[\Delta_s^2+\Delta_t^2+\left(4-\frac{1}{\alpha}\right)\right]\xi_{st}=\epsilon_{st} \tag{63}$$

(where Δ is the central difference operator) the analogy with the continuous relation, a stochastic Laplace equation, becomes apparent:

$$\left[\left(\frac{\partial}{\partial x}\right)^2+\left(\frac{\partial}{\partial y}\right)^2-\kappa^2\right]\xi(x,y)=\epsilon(x,y). \tag{64}$$

The continuous relation is the easier to handle (as was also the case for relations (56), (60)) due to its essentially less artificial nature. Explicit results have been obtained for scheme (62) (Van der Pol & Bremmer, 1950; Stöhr, 1950), but these are not simple. From the formula (Titchmarsh, 1948, p. 201)

$$\frac{1}{4\pi^2}\iint\frac{e^{i(X\omega_1+Y\omega_2)}\,d\omega_1\,d\omega_2}{(\omega_1^2+\omega_2^2+\kappa^2)^{\mu+1}}=\left(\frac{r}{2\kappa}\right)^{\mu}\frac{K_\mu(\kappa r)}{\Gamma(\mu+1)}\quad(r=\sqrt{(X^2+Y^2)}), \tag{65}$$

we deduce that for scheme (64)

$$\xi(x,y)=\iint_{-\infty}^{+\infty}\epsilon(x+X,y+Y)\,K_0(\kappa r)\,dX\,dY \tag{66}$$

and

$$\phi(X,Y)=\frac{r}{2\kappa}K_1(\kappa r), \tag{67}$$

where K_ν is the modified Bessel function of the second kind, order ν. Since $\lim rK_1(r)=1$ the correlation coefficient corresponding to (67) is

$$\rho(r)=\kappa r K_1(\kappa r). \tag{68}$$

Scheme (64) is a special case of the general second-order stochastic difference equation. According to the usual classification into elliptic, parabolic, and hyperbolic forms, (64) would be regarded as a circular form with centre at the origin. An investigation of the remaining second-order schemes would be of great interest, both theoretically and practically.

The correlation function (68) is of interest in that it may be regarded as the 'elementary' correlation in two dimensions, similar to the exponential $e^{-\alpha|x|}$ in one dimension. Both correlation curves are monotone decreasing, but (68) differs in that it is flat at the origin, and that its rate of decay is slower than exponential.

Attempts have been made to represent two-dimensional correlograms as sums of exponentials, largely because the exponential has proved itself the natural choice in one dimension, and the observed curves display the same monotonic decay. However, it is apparent from the previous paragraphs that the exponential function has no divine right in two dimensions, while the example of the last section indicated that a K_1 function fitted the observations better than did an exponential.

Two-dimensional processes can be constructed which have exponential correlation functions, but these are very artificial. For example, Matérn (1947) has shown that the correlation function $\exp-\alpha\sqrt{(X^2+Y^2)}$ corresponds to a spectral function $(\omega_1^2+\omega_2^2-\alpha^2)^{-\frac{1}{2}}$. The simplest process with such a spectral function may be formally written

$$\left[\left(\frac{\partial}{\partial x}\right)^2+\left(\frac{\partial}{\partial y}\right)^2-\alpha^2\right]^{\frac{1}{2}}\xi(x,y)=\epsilon(x,y) \tag{69}$$

and it is difficult to visualize a physical mechanism which would lead to such a relation.

The examination of the two sets of uniformity data was begun by Mr I. D. Dick of the New Zealand Department of Scientific and Industrial Research, who also in large degree stimulated the author to the present investigation. Thanks are also due to the supervisor of the punched card calculations, Mr E. W. Jones.

REFERENCES

BARTLETT, M. S. (1946). Stochastic processes. Mimeographed N. Carolina lecture notes.
BATCHELOR, L. D. & REED, H. S. (1924). Relation of the variability of yields of fruit trees to the accuracy of field trials. *J. Agric. Res.* **12**, 245–83.
DANIELL, P. J. (1946). Contribution to discussion on stochastic processes. *J. R. Statist. Soc.* B, **8**, 88–90.
DOOB, J. L. (1944). The elementary Gaussian processes. *Ann. Math. Statist.* **15**, 229–82.
MATÉRN, B. (1947). Metoder att uppskatta noggrannheten vid linje- och provytetaxering. *Medd. SkogsforsknInst.* **36**, no. 1.
MERCER, W. B. & HALL, A. D. (1911). The experimental error of field trials. *J. Agric. Sci.* **4**, 107–32.
MORAN, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.
NEYMAN, J. (1952). *Lectures and Conferences on Mathematical Statistics and Probability*. Graduate School, U.S. Dep. of Agriculture.
QUENOUILLE, M. H. (1949). Problems in plane sampling. *Ann. Math. Statist.* **20**, 355–75.
STÖHR, A. (1950). Über einige lineare partielle Differenzgleichungen mit konstanten Koeffizienten. *Math. Nachr.* **3**, 295–315.
TITCHMARSH, E. C. (1948). *Introduction to the Theory of Fourier Integrals*. Oxford University Press.
VAN DER POL, B. & BREMMER, H. (1950). *Operational Calculus*. Cambridge University Press.
WHITTLE, P. (1952). Tests of fit in time series. *Biometrika*, **39**, 309–18.
WHITTLE, P. (1953). The analysis of multiple stationary time series. *J. R. Statist. Soc.* B, **15**, 125–39.
WIENER, N. (1949). *The Extrapolation, Interpolation and Smoothing of Stationary Time-Series*. New York, Wiley.
WILLIAMS, R. M. (1954). The choice of sampling intervals for systematic samples from populations with stationary correlation. (Unpublished paper.)
WOLD, H. (1938). *A Study in the Analysis of Stationary Time Series*. Upsala.

[EDITORIAL NOTE. The preceding paper by Whittle (on p. 443) and the following paper by Patankar (on pp. 459–60) both use as an illustration of theory Mercer & Hall's (1911) wheat-plot data. On a first inspection it might be thought that the serial correlations in (i) the column $s = 0$ and (ii) in the row $t = 0$ of Whittle's Table 1, p. 443, should correspond with Patankar's correlations described as (i) 'along the rows, results over the whole data, original' (last column of Table 5·2, p. 460) and (ii) 'along the columns, results over the whole data' (last column of Table 5·1). Reference to the authors has, however, confirmed that the quantities computed are differently defined. Thus if

$i = 1, 2, \ldots, 25$ denotes columns running from west to east,

$j = 1, 2, \ldots, 20$ denotes rows running from north to south,

the co-variance in Whittle's expression for r_{s0} is

$$\sum_{i=1}^{25}\sum_{j=1}^{20-s} x_{i,j}x_{i,j+s} - \sum_{i=1}^{25}\sum_{j=1}^{20-s} x_{i,j} \times \sum_{i=1}^{25}\sum_{j=1}^{20-s} x_{i,j+s}/\{25(20-s)\},$$

with similar expressions for the variances. The corresponding covariance used by Patankar is, however,

$$\sum_{i=1}^{25}\left\{\sum_{j=1}^{20-s} x_{i,j}x_{i,j+s} - \sum_{j=1}^{20-s} x_{i,j} \times \sum_{j=1}^{20-s} x_{i,j+s}/(20-s)\right\},$$

with similar expressions for the variances. Thus Whittle bases his serial correlations on the total variation and covariation while Patankar's correlations are based on within-column (or within-row) variation and covariation. As might be expected, Whittle's correlations are larger than Patankar's because of the variation between rows and columns.]