Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource

John Haslett; Adrian E. Raftery

Space–time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource

By JOHN HASLETT†

Trinity College, Dublin, Ireland

and ADRIAN E. RAFTERY

University of Washington, Seattle, USA

[Read before The Royal Statistical Society at a meeting organized by the Research Section on Wednesday, May 25th, 1988, Professor A. P. Dawid in the Chair]

SUMMARY
We consider estimation of the long term average power output from a wind turbine generator at a site for which few data on wind speeds are available. Long term records of wind speeds at the 12 synoptic meteorological stations are also used. Inference is based on a simple and parsimonious approximating model which accounts for the main features of wind speeds in Ireland, namely seasonal effects, spatial correlation, short-memory temporal autocorrelation and long-memory temporal dependence. It synthesizes deseasonalization, kriging, ARMA modelling and fractional differencing in a natural way. A simple kriging estimator performs well as a point estimator, and good interval estimators result from the model. The resulting procedure is easy to apply in practice.

Keywords: Deseasonalization; Fractional differencing; Kriging; Optimal interpolation; Persistence

1. Introduction

The Irish government has, in recent years, been considering the possibility of using wind energy to meet a significant portion of Ireland's energy needs. This paper describes a project aimed at developing methods for the evaluation of Ireland's wind power resource.

This resource may be exploited at various scales. Large scale wind farms, involving some hundreds of wind turbines, could supply the electricity grid with a significant proportion of its energy needs; one study envisaged up to 25 % (Gibbons et al., 1979). Isolated communities, such as on the islands, where electricity costs are particularly high, could be supplied by medium scale turbines. One interesting small scale project is the heating of greenhouses, where the demand for energy correlates seasonally with its availability. There are many difficult and different problems in this broad task; here we concentrate on the quantification of the resource at a specific site.

Data on the availability of wind energy in Ireland are sparse, because few data exist at locations of interest for wind energy purposes. Only at 12 synoptic meteorological stations have detailed and reliable records of wind speeds and directions been compiled over a long period; see Fig. 1. These data are described in Section 2. Data at a potential site will typically be few or non-existent, although there
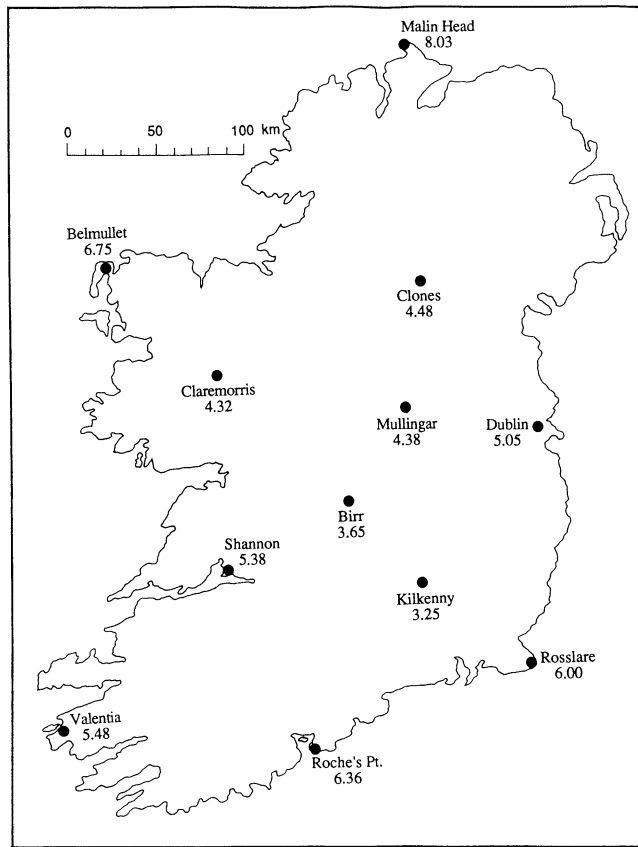
Fig. 1. Irish Meteorological Service synoptic meteorological stations, with mean wind speeds (in metres per second) for 1961–78

may be some broad guidance available from an experienced evaluation of the local topography. The kinetic energy in the wind, and the power available to any specific turbine, is a non-linear function of wind speed, and it is therefore necessary to estimate the full distribution of wind speeds, and not merely an average.

Before starting to operate a wind turbine or wind farm at a potential site, it would be usual to record wind data there for a short period of several weeks or months. As we shall see in Section 3, a simple estimator based solely on such a short run of data performs poorly, and a better estimator results by 'adjusting' such an estimator, using the long term records at the synoptic stations. This estimator is calculated by first deseasonalizing the data, and then applying a general least squares approach. Given our spatial context, this has close links with 'kriging'. However, the standard errors yielded by standard assumptions of temporal independence, or even short-memory temporal dependence, are much too small, due mainly to the presence of temporal persistence, namely non-negligible dependence between observations a long time span apart.

We therefore develop alternative interval estimators based on a simple and parsimonious approximating model which accounts for the main features of the data, namely seasonality, spatial correlation, short-memory temporal dependence and

temporal persistence; see Section 4. The resulting standard errors are quite accurate. The model also yields estimates of the full distribution of wind speeds, and hence of the available kinetic energy in the wind; see Section 5.

2. The Data

The data are hourly wind speeds and directions at each of 12 synoptic meteorological stations during the period 1961–78; see Fig. 1. The wind speeds were recorded in knots (1 knot = 0.5148 m/s), but converted to metres per second for analysis. Here we present the main features of the data that turned out to be important for our purpose; further description can be found in Haslett and Kelledy (1979) and Raftery et al. (1982).

At all levels of temporal aggregation, standard deviations are correlated with means. Also, marginal distributions are noticeably asymmetric. Standard exploratory techniques suggested taking a square root transformation, and this does, indeed, stabilize the variance over both stations and time periods, and make the marginal distributions approximately normal. This same transformation has been found to be helpful in similar contexts by other workers; see Brown et al. (1984) and Carlin and Haslett (1982).

Another problem is the choice of level of temporal aggregation. For example, Brown et al. (1984) work with hourly data, Barros and Estevans (1983) use weekly averages, and Balling and Cervany (1984) rely on monthly aggregates. Our procedure for estimating power at the new site is in two stages; first, estimating the distribution of wind speeds at the new site, which is easier for more aggregated data, and, secondly, translating this into available kinetic energy, which is easier for less aggregated data; see Section 5. The extent to which speeds at one station can be predicted from contemporaneous speeds at others increases quite rapidly up to a level of aggregation of about one day, and thereafter more slowly. Guided by these observations, we have based our analyses on the square roots of daily mean wind speeds.

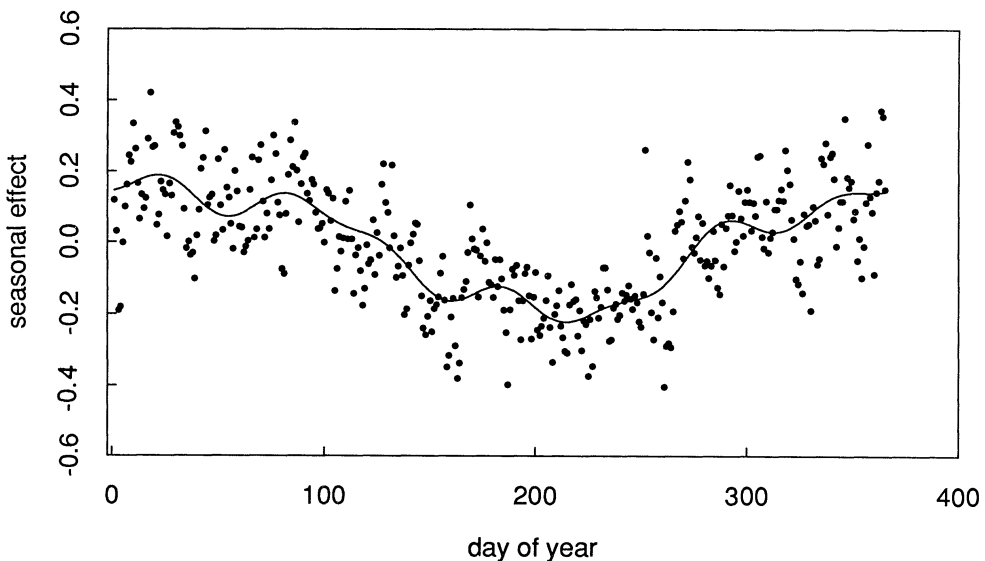Wind speeds vary with time of year, although the seasonal effect is not very strong,

Fig. 2.   Seasonal effects. The dots show the average of the square root of the daily means over all stations and years, for each day of the year. The full line is the estimated seasonal effect.

accounting for about one-quarter of the total variance. We estimated the seasonal effect by calculating the average of the square roots of the daily means over all years and stations for each day of the year, and then regressing the results on a set of annual harmonics; see Fig. 2. Subtraction of the estimated seasonal effect from the square roots of the daily means then yields deseasonalized data, hereafter referred to as velocity measures.

Contemporaneous velocity measures at different stations are highly correlated, and the correlations are clearly related to the distances between stations; see Fig. 3. The correlations involving one station—Rosslare—are much lower than the others. This may be because its position in the extreme south-east of the country makes it subject to meteorological influences which do not affect the other stations. We have omitted Rosslare from our calculations. The spatial correlation pattern revealed by Fig. 3 changes little with time of year.

The data exhibit some short-memory temporal autocorrelation; see Fig. 4. There are striking similarities between its pattern and extent at the different stations.

3. A Kriging Estimator

Suppose we have data on wind velocity measures at m places, labelled $1, \ldots, m$. One of these, the new site, is labelled k and there we have a short run of data collected on n consecutive days, $t = t_0, \ldots, t_0 + n - 1$. At the other places, the synoptic meteorological stations, we have long runs of data collected on N consecutive days, $t = 1, \ldots, N$. Here $n \ll N$, and $1 \leqslant t_0 < t_0 + n - 1 \leqslant N$. Let $\mathbf{X}_t = (X_{1t}, \ldots, X_{mt})^T$, where X_{it} is the velocity measure at place i on day t. Then, as we shall see in Section 5, knowledge of the mean and variance of X_{kt} can be translated into quite precise knowledge of the average available kinetic energy in the wind at the new site. This section and the next one are mainly concerned with estimating the mean and variance of X_{kt}.
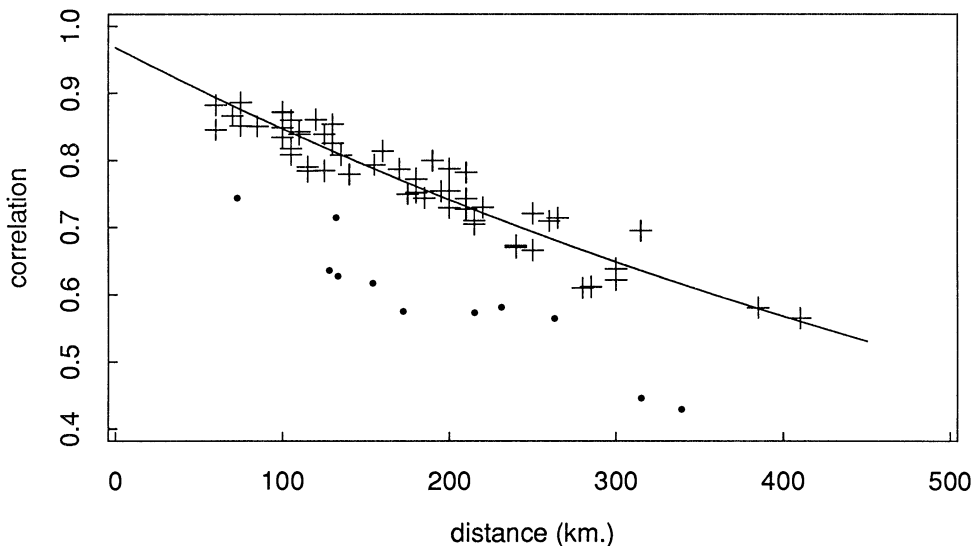

Fig. 3.   Distance–correlation plot. Each cross corresponds to a pair of synoptic stations; the dots correspond to pairs which include Rosslare. The full line is the fitted relationship (3.3) with $\alpha = 0.968$ and $\beta = 0.00134$.

Let $\mu_i = E[X_{it}]$ ($i = 1, \ldots, m$), and $\bar{X}_{i,t,n} = n^{-1} \Sigma_{s=1}^n X_{i,t+s-1}$. A simple estimator of μ_k is $\bar{\mu}_k = \bar{X}_{k,t_0,n}$, the average of the observed velocity measures at place k. A simple estimator of the variance of $\bar{\mu}_k$ is

$$\bar{\sigma}_{k,t_0,n}^2 = \left\{ n^{-1} \sum_{s=1}^n (X_{k,t_0+s-1} - \bar{\mu}_k)^2 \right\}/(n-1) \tag{3.1}$$

The estimator $\bar{\mu}_k$ takes no account of the long runs of data at other places, and we now develop an estimator which does exploit these. The analyses in Section 2 indicate that the correlation between wind speeds at different places is strongly related to the distance between them, and suggest that the covariance structure can be reasonably well approximated by the relations

$$\text{cov}(X_{it}, X_{jt}) = \sigma_X^2 r_{ij} \tag{3.2}$$

where

$$r_{ij} = \begin{cases} 1 & \text{if } i = j \\ \alpha \exp(-\beta d_{ij}) & \text{if } i \neq j. \end{cases} \tag{3.3}$$

In (3.3), $0 \leqslant \alpha \leqslant 1$, $\beta \geqslant 0$, and d_{ij} is the distance (in kilometres) between places i and j. If $\alpha < 1$ there is a 'nugget effect', namely spatial correlation less than one at arbitrarily small distances. This is due to measurement error and very small scale effects (Journel and Huijbregts, 1978).

If one assumes that (3.2) and (3.3) hold, and ignores temporal dependence, the general least squares estimator of μ_k is

$$\tilde{\mu}_k = \mathbf{a}_k^{\text{T}} (\bar{\mathbf{X}}_{t_0,n} - \ddot{\boldsymbol{\mu}})/a_{kk} \tag{3.4}$$

In (3.4) $\mathbf{A} = \mathbf{R}^{-1}$, where $\mathbf{A}$ is the $m \times m$ matrix with elements (a_{ij}), and $\mathbf{R}$ is the $m \times m$ matrix with elements (r_{ij}) defined in (3.3); $\mathbf{a}_k = (a_{1k}, \ldots, a_{mk})^{\text{T}}$; $\bar{\mathbf{X}}_{t_0,n} = (\bar{X}_{1,t_0,n}, \ldots, \bar{X}_{m,t_0,n})^{\text{T}}$; and $\ddot{\boldsymbol{\mu}} = (\ddot{\mu}_1, \ldots, \ddot{\mu}_m)^{\text{T}}$, where

$$\ddot{\mu}_i = \begin{cases} \bar{X}_{i,1,N} & (i \neq k) \\ 0 & (i = k). \end{cases}$$

(3.4) was obtained as the general least squares estimator, given the spatial covariance structure consisting of a smooth isotropic function of distance. This approach underlies many 'kriging' procedures (Journel and Huijbregts, 1978; Ripley, 1981), and we therefore call $\tilde{\mu}_k$ a kriging estimator. It has also been used in meteorological applications, where it is known as 'optimal' or 'objective interpolation' (Gandin, 1965; Creutin and Obled, 1982; Tabios and Salas, 1985; Thiebaux and Peddar, 1986). The results are fairly robust to some misspecification of the spatial covariance structure (Cressie, 1985; Brooker, 1986).

It is clear that $\tilde{\mu}_k$ is unbiased. If (3.2) holds and the observations are temporally uncorrelated, then

$$\text{var}(\tilde{\mu}_k) = \sigma_X^2 \mathbf{a}_k^{\text{T}} \mathbf{R} \mathbf{a}_k/a_{kk}^2 n \tag{3.5}$$

If one assumes, in addition, that $\mathbf{a}_k^{\text{T}} \mathbf{X}_t$ is normally distributed, which does seem to be approximately the case for our data, then $\tilde{\mu}_k$ is also normally distributed, and interval estimators result from (3.5). Even if $\mathbf{a}_k^{\text{T}} \mathbf{X}_t$ is not normally distributed, $\tilde{\mu}_k$ will still be approximately normally distributed in large samples, under regularity conditions.
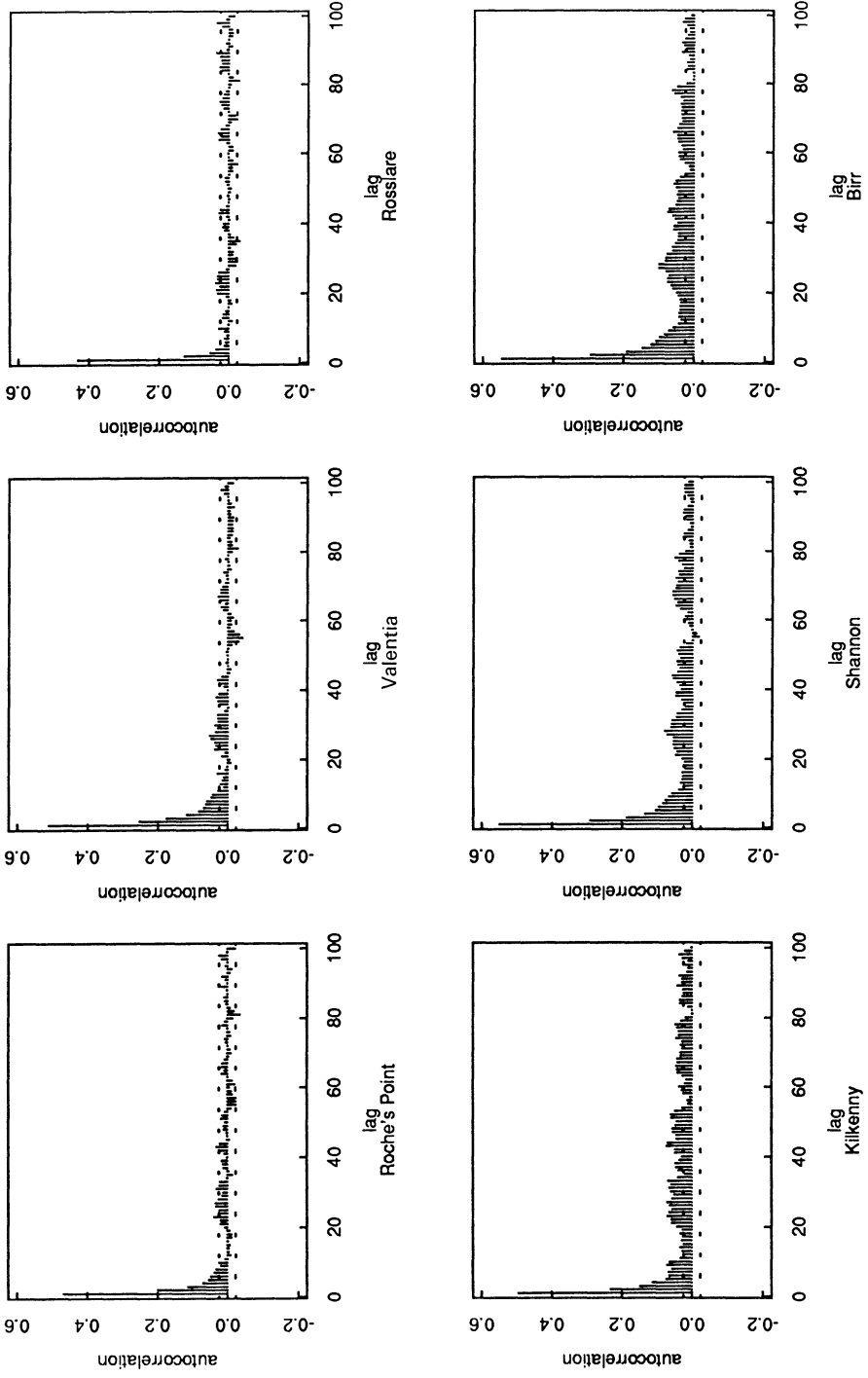
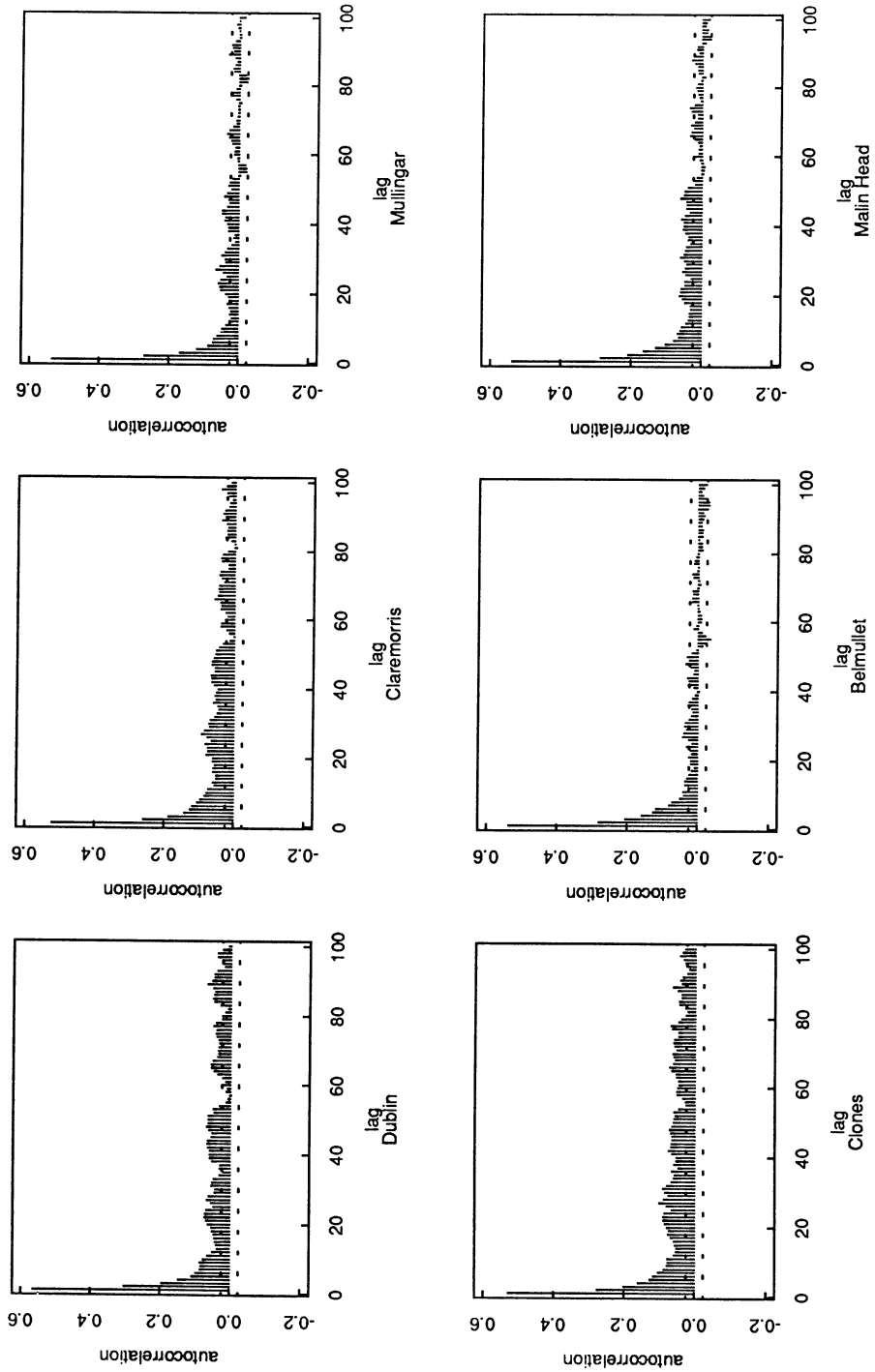Fig. 4. Autocorrelation functions of the velocity measures for the 12 synoptic stations.

TABLE 1

Empirical and theoretical mean squared errors (multiplied by 10 000) of $\bar{\mu}_k$ and $\tilde{\mu}_k$. The mean squared errors based on (3.1) and (3.5) do not take account of temporal autocorrelation. The mean squared errors based on (4.10) take account of both short- and long-memory temporal dependence.

	MSE ($\bar{\mu}_k$)		MSE ($\tilde{\mu}_k$)		
n	Empirical	From (3.1)	Empirical	From (3.5)	From (4.10)
20	595	197	190	66	204
40	366	83	159	33	160
80	254	43	131	16	126
160	155	21	107	8	99
320	101	10	82	4	78

In order to assess the performance of $\tilde{\mu}_k$ as compared with the simpler estimator $\bar{\mu}_k$, and also to assess the variance estimators given by (3.1) and (3.5), we carried out an approximate cross-validation exercise, the results of which are shown in Table 1. For each of several values of n we calculated $\tilde{\mu}_k$ and $\bar{\mu}_k$ for each disjoint data run of length n at each station in turn, and compared these with $\ddot{\mu}_k$, considered to be the 'true' value for this purpose. α, β and σ_X^2 were estimated once from the entire data set, as described in Section 4; this provides a good approximation to a complete recomputation on the deletion of each station in turn.

As an estimator of μ_k, $\tilde{\mu}_k$ performs better than $\bar{\mu}_k$, particularly for short data runs. For example, for runs of length $n = 20$ days, using $\tilde{\mu}_k$ rather than $\bar{\mu}_k$ reduces the empirical mean squared error by about 68 %. It appears, empirically, that to achieve the same precision, the simpler estimator would require about six times as much data. The gain in precision decreases as n increases. The empirical distributions of $\bar{\mu}_k$ and $\tilde{\mu}_k$ did not deviate appreciably from normality.

The estimated variances of $\bar{\mu}_k$ obtained from (3.1), and those of $\tilde{\mu}_k$ obtained from (3.5) are, however, clearly quite inaccurate. Not only are they much too small for all values of n, but the extent to which they fall below the empirical variances increases with n. When $n = 320$, they are too small by factors of about 10 and 20, respectively. This indicates that simply taking into account the short-memory temporal autocorrelation suggested by Fig. 4 would not, by itself, be sufficient to make the standard errors accurate, as it would, effectively, just multiply all the standard errors by a constant.

This suggests that the data exhibit persistence, or long-memory temporal dependence, one of whose manifestations is that the sampling variance of the sample mean decreases more slowly, asymptotically, than the usual rate for short-memory processes, $O(n^{-1})$ (Hosking, 1982, 1984a). Further evidence for persistence is provided by Fig. 5, which shows, for each station, the periodogram of the residuals from a fitted AR(9) model. In Fig. 5, the short-memory temporal dependence has been largely removed, and yet there is a concentration of power at low frequencies, which is characteristic of long-memory temporal dependence (Graf *et al.*, 1984). The patterns at the different stations are quite similar to each other.

Most models for long-memory temporal dependence imply an approximately linear relationship between power and log-frequency at low frequencies. Inspection of Fig. 5 reveals that for our data, at the very lowest frequencies, power is slightly less than such models would lead us to expect. This may be due to the fact that Fig. 5 shows the periodogram from a finite sample and not the true spectrum, rather than to inappropriateness of such models. The implied truncation of the autocovariance function may lead to some negative bias at the lowest frequencies.

4. Modelling and Interval Estimation

4.1. *A Model*

We base inference for μ_k on a single model for the entire space–time process $\{\mathbf{X}_t : t \in \mathbb{Z}\}$, as follows:

$$X_{it} = \mu_i + \nabla^{-d}\phi(B)^{-1}\theta(B)\varepsilon_{it} \qquad (i = 1, \ldots, m) \tag{4.1}$$

where

$$\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{mt})^{\mathrm{T}} \overset{\text{iid}}{\sim} \mathrm{MVN}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{R}) \tag{4.2}$$

In (4.1), B is the backward shift operator such that $B\varepsilon_{it} = \varepsilon_{i,t-1}$, $\nabla = (1 - B)$, $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$. It is assumed that $0 \leqslant d \leqslant \frac{1}{2}$ and that the zeros of $\phi(B)$ and $\theta(B)$ lie outside the unit circle. ∇^{-d} is defined by the binomial series expansion of $(1 - B)^{-d}$.

In (4.1), temporal persistence is modelled by the use of fractional differencing (Granger and Joyeux, 1980; Hosking, 1981). (4.1) and (4.2) imply that the second-order moment structure of the space–time process is isotropic and stationary in both space and time. This is true approximately, but not exactly; assuming it enables the resulting model to be applied easily at a new site for which little other information about spatial and temporal covariances is available.

4.2. *Identification and Preliminary Estimation*

We used the following procedure to identify the orders of the polynomials $\phi(B)$ and $\theta(B)$, and to obtain initial estimates of the parameters. This is important, because the more exact estimation method described in Section 4.3 is feasible only if reasonable starting values are available. The full data set, consisting of $N = 6574$ observations at each of $m = 11$ stations, was used.

1. Form preliminary estimates of α and β by regressing $\log\{\mathrm{corr}(X_{it}, X_{jt})\}$ on d_{ij}, using the fact that $\mathrm{corr}(X_{it}, X_{jt}) = \mathrm{corr}(\varepsilon_{it}, \varepsilon_{jt})$ for (4.1).
2. Form m approximately spatially independent series $\{Y_{it} : t = 1, \ldots, N\}$ ($i = 1, \ldots, m$), where $\mathbf{Y}_t = \mathbf{C}\mathbf{X}_t$, and $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{mt})^{\mathrm{T}}$. Here $\mathbf{C}$ is a lower triangular $m \times m$ matrix such that $\mathbf{C}\mathbf{R}\mathbf{C}^{\mathrm{T}} = \mathbf{I}$; $\mathbf{C}$ is constructed by the Gram–Schmidt orthogonalization procedure.
3. Find an autoregressive filter which accounts for most of the short-memory dependence in the $\{\mathbf{Y}_t\}$ series, and filter each of the m series $\{Y_{it} : t = 1, \ldots, N\}$ with it. This yields a data set which is free, to a good approximation, of both spatial and short-memory temporal dependence, and whose main feature is persistence. The filter we used was of order nine.
4. Form means of n consecutive values from the output of step 3, for several subsets of the data, and several values of n. For each value of n, calculate the sampling
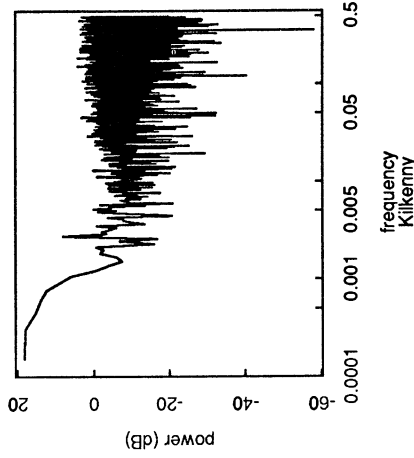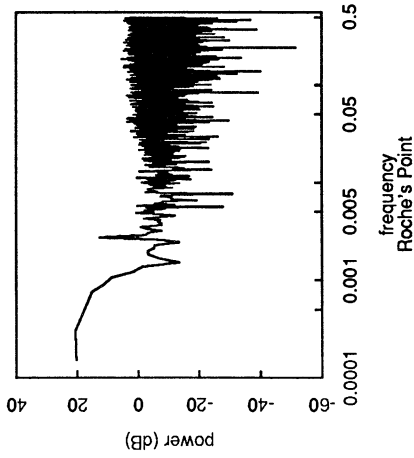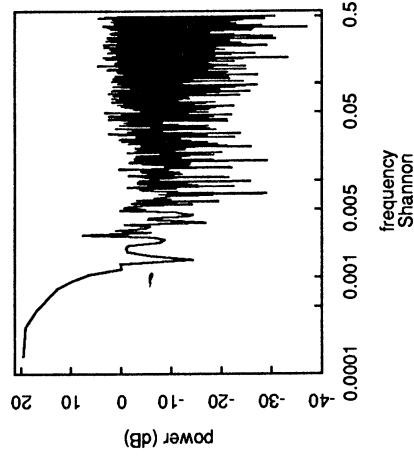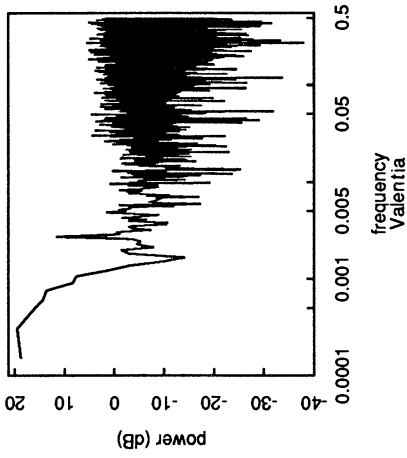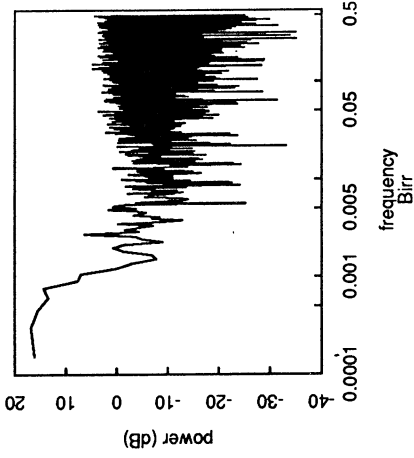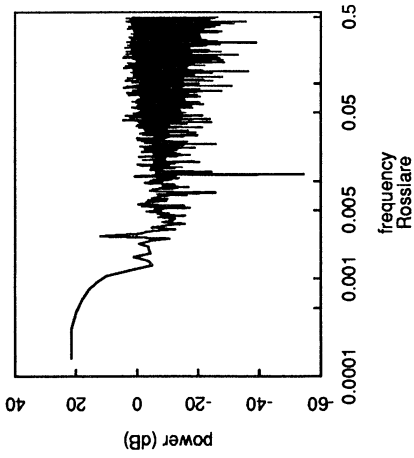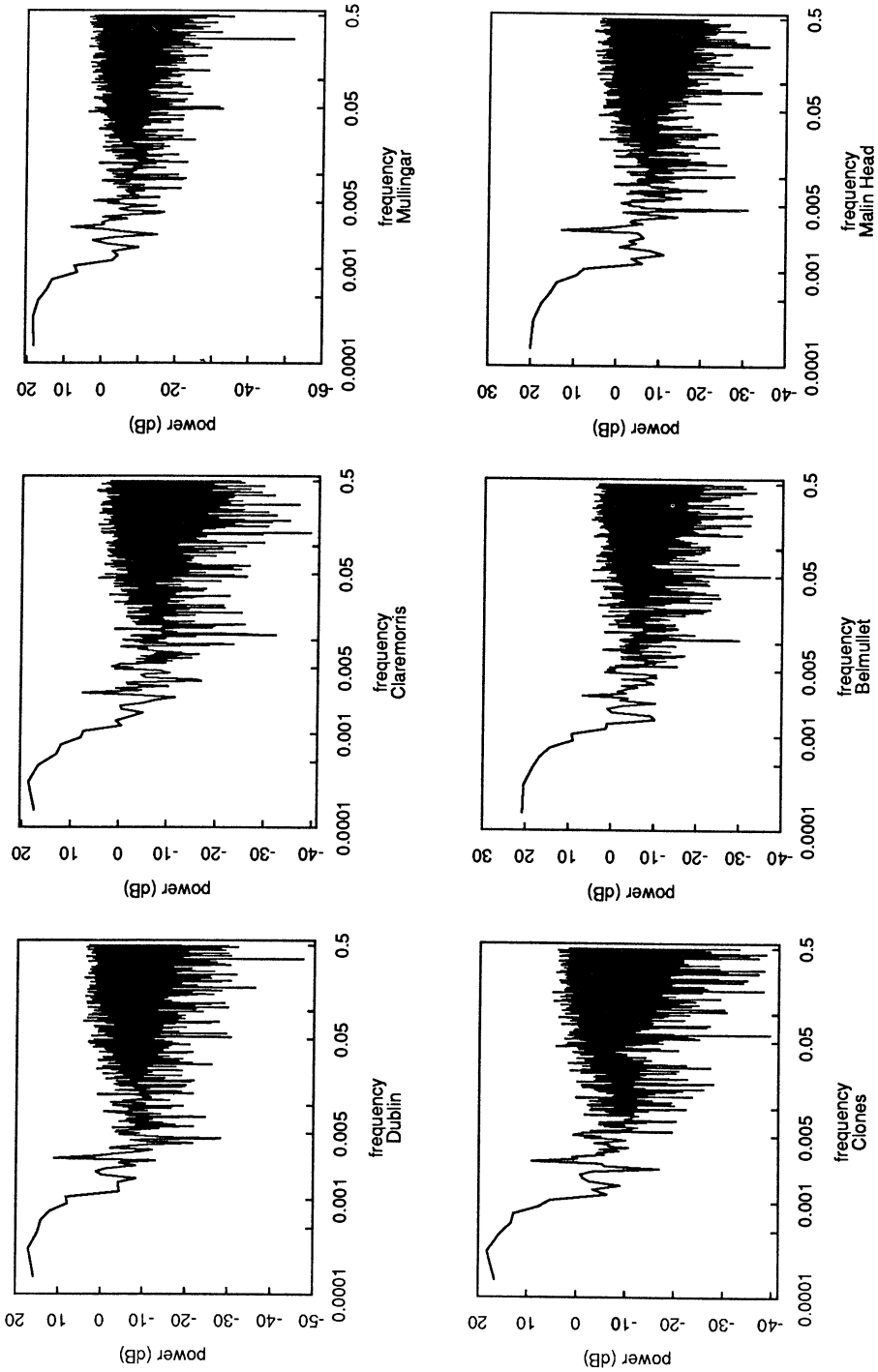
Fig. 5. Log-periodograms of the velocity measures for the 12 synoptic stations, after being passed through a short-memory AR(9) filter.

variance of these sample means. Then, regress the logarithm of the sampling variance on log n, and take one-half of one minus the estimated slope as an initial estimate of d, $\tilde{d}$, say. This is motivated by the asymptotic results of Hosking (1982, 1984a).

5. Form the m series $\{\nabla^{\tilde{d}} Y_{it}: t = 1, \ldots, N\}$. A fast and accurate way of implementing this filter is described in Section 4.3.

6. Identify a common ARMA(p, q) model for the m series $\{\nabla^{\tilde{d}} Y_{it}: t = 1, \ldots, N\}$, and estimate it, yielding initial estimates of $\phi(B)$ and $\theta(B)$. We identified an AR(2) model for our data.

4.3. *Maximum Likelihood Estimation*

Let $\mathbf{X}^t = \{\mathbf{X}_1, \ldots, \mathbf{X}_t\}$ and $\mathbf{X}_i^t = \{X_{i1}, \ldots, X_{it}\}$. Then the likelihood can be calculated exactly by noting that, conditionally on $\mathbf{X}^{t-1}$, $\mathbf{X}_t$ has a multivariate normal distribution, where $E[X_{it} | \mathbf{X}^{t-1}] = E[X_{it} | \mathbf{X}_i^{t-1}]$, $\text{var}[X_{it} | \mathbf{X}^{t-1}] = \text{var}[X_{it} | \mathbf{X}_i^{t-1}]$, and $\text{corr}(X_{it}, X_{jt} | \mathbf{X}^{t-1}) = \alpha \exp(-\beta d_{ij})$. The one-dimensional conditional means, $E[X_{it} | \mathbf{X}_i^{t-1}]$, and variances, $\text{var}[X_{it} | \mathbf{X}_i^{t-1}]$, may be calculated by inserting the autocorrelations for the fractionally differenced ARIMA(p, d, q) process (Hosking, 1981) into the Durbin–Levinson recursion (Ramsey, 1974).

Maximum likelihood estimators can then be found by numerically maximizing the likelihood. This is, however, a demanding task, computationally. For example, a single evaluation of the likelihood takes about three hours of CPU time on a VAX 11/780, and finding the maximum likelihood estimator would take at least 45 hours, even with good starting values. Carlin *et al.* (1985) and Carlin (1987) did obtain estimates for long-memory time series models by numerically maximizing the exact likelihood, which seems to have been practicable because they were working with short, one-dimensional, series of lengths less than 220.

However, a fast and accurate approximation can be found, as follows. First, we note that, to an excellent approximation, the conditional means and variances may be found using only the partial autocorrelations for the fractionally differenced ARIMA(0, d, 0) process, and not those for the full ARIMA(p, d, q) process, which are much more complicated (Hosking, 1981). Then we can find approximate maximum likelihood estimators of μ and σ_ε^2 analytically, and find a concentrated likelihood which is a function only of α, β, d, $\phi(B)$ and $\theta(B)$. Finally, we approximate the partial linear regression coefficients of the ARIMA(0, d, 0) process, the calculation of which dominates the CPU requirements. The resulting approximation reduces the required CPU time by a factor of about 70 for our data, and appears to be quite accurate.

We now describe the approximation used in more detail. We first note that, approximately,

$$E[X_{it} | \mathbf{X}_i^{t-1}] = u_{it} + w_t \mu_i \tag{4.3}$$

and

$$\text{var}[X_{it} | \mathbf{X}_i^{t-1}] = \sigma_X^2 \kappa \prod_{j=1}^{t-1} (1 - \phi_{jj}^2) = v_t \tag{4.4}$$

In (4.3) and (4.4)

$$u_{it} = \phi(B)\theta(B)^{-1} \sum_{j=1}^{t-1} \phi_{tj} X_{i,t-j} \tag{4.5}$$

$$w_t = 1 - \phi(1)\theta(1)^{-1} \sum_{j=1}^{t-1} \phi_{tj} \tag{4.6}$$

where the ϕ_{tj} are the partial linear regression coefficients for the ARIMA(0, d, 0) process, given explicitly by Hosking (1981), and κ is the ratio of the innovations variance to the process variance for the ARMA(p, q) process with parameters $\phi(B)$ and $\theta(B)$, as defined by equation (3.4.4) of Box and Jenkins (1976).

Given values of α, β, d, $\phi(B)$ and $\theta(B)$, approximate maximum likelihood estimators of μ and σ_ε^2 are then available analytically as

$$\hat{\mu}_i = \sum_{t=1}^{N} w_t(X_{it} - u_{it})v_t^{-1/2} \Big/ \sum_{t=1}^{N} w_t^2$$

$$\hat{\sigma}_\varepsilon^2 = (Nm)^{-1} \sum_{t=1}^{N} (\mathbf{X}_t - \mathbf{u}_t - w_t\hat{\mu})^{\mathrm{T}} \mathbf{A}(\mathbf{X}_t - \mathbf{u}_t - w_t\hat{\mu})v_t^{-1}$$

where $\mathbf{u}_t = (u_{1t}, \ldots, u_{mt})^{\mathrm{T}}$ and $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_m)^{\mathrm{T}}$. The concentrated log-likelihood, namely the log-likelihood with μ and σ_ε^2 replaced by $\hat{\mu}$ and $\hat{\sigma}_\varepsilon^2$, is then approximately

$$l(\alpha, \beta, d, \phi(B), \theta(B)) = \text{constant} - \tfrac{1}{2}Nm \log \hat{\sigma}_\varepsilon^2 - \tfrac{1}{2}N|\mathbf{R}| \qquad (4.7)$$

and approximate maximum likelihood estimators can be found by maximizing this numerically. It is computationally efficient to maximize l as a function of the other variables conditionally on d, for each trial value of d.

The recursive calculation of the ϕ_{tj} and the evaluation of (4.3) dominate the CPU time requirements. We approximate these, noting that by Hosking (1981), $\phi_{tj} \sim -\pi_j$ and $\pi_j \sim O(j^{-d-1})$ as $t \to \infty$, where the π_j are the π weights of the ARIMA(0, d, 0) process, as defined by Box and Jenkins (1976). Our approximation consists of taking these asymptotic relationships to hold exactly for $j > M$, where M is some integer, and then taking the π_j to be constant, at their approximate average value, for $M < j \leqslant t - 1$. This yields

$$\sum_{j=1}^{t-1} \phi_{tj}X_{i,t-j} \approx \sum_{j=1}^{M} \phi_{tj}X_{i,t-j} - \sum_{j=M+1}^{t-1} \pi_j X_{t-j}$$

$$\approx \sum_{j=1}^{M} \phi_{tj}X_{i,t-j} - M\pi_M d^{-1}\{1 - (M/j)^d\}\bar{X}_{i,M+1,t-1-M}. \qquad (4.8)$$

(4.8) is then substituted into (4.5), and a similar approximation is used in (4.6). This provides a good approximation to the likelihood function as a whole, and so opens the possibility of Bayesian, as well as likelihood, inference for fractionally differenced models. Numerical experimentation indicated that choosing $M = 100$ gives good results over a wide range of values of d and N. For our application, the approximation reduced CPU time by a factor of about 70, so that the CPU time required for a single evaluation of the likelihood went down from three hours to about $2\frac{1}{2}$ minutes.

Hosking (1984b) proposed an approximation which is similar in spirit to (4.8). However, (4.8) seems to be somewhat faster for the very long series we are dealing with here ($N = 6574$), and seems also to avoid the starting value problem for the long-memory filter. Spectral approximations to the maximum likelihood estimator in the one-dimensional case have been proposed by Fox and Taqqu (1986), Beran (1986) and Geweke and Porter-Hudak (1983); these have not, however, been generalized to the multivariate, or space–time, context. Short-memory autoregressive approximations have been suggested by Granger and Joyeux (1980) and Li and

McLeod (1986), but we felt it important to retain the long-memory property in the approximation used.

Neither the finite sample, nor the asymptotic distribution of the maximum likelihood estimator for models such as (4.1) appears to be known. However, Yajima (1985) has shown that the maximum likelihood estimator for a one-dimensional specialization of (4.1), namely the fractionally differenced ARIMA(0, d, 0) process with known mean, has the usual asymptotic normal distribution, while a similar result for the purely spatial specialization of (4.1) follows from Mardia and Marshall (1984). We conjecture that the usual result does hold for (4.1), and could be proved by combining the arguments used in these two papers. We give approximate standard errors based on this conjecture, equal to the square roots of the diagonal elements of a numerical approximation to minus the inverse Hessian of the approximate concentrated log-likelihood function, evaluated at its maximum.

The approximate maximum likelihood estimators, with their approximate standard errors in parentheses, are $\hat{\alpha} = 0.968\,(0.0013)$, $\hat{\beta} = 0.00134\,(0.000025)$, $\hat{d} = 0.328\,(0.0029)$, $\hat{\phi}_1 = 0.010\,(0.0123)$, $\hat{\phi}_2 = -0.063\,(0.0123)$ and $\hat{\sigma}_\varepsilon^2 = 0.246$. Thus, the long-memory effect is found to be large, and the existence of a fairly small nugget effect is suggested by $\hat{\alpha}$ being slightly, but apparently significantly, less than one.

4.4. Model Checking

Suppose that $\mathbf{e}_t = \mathbf{C}(\mathbf{X}_t - \mathbf{u}_t - w_t\hat{\boldsymbol{\mu}})v_t^{-1/2} = (e_{1t}, \ldots, e_{mt})^T$. Then, conditionally on the estimated model, the e_{it} should be independent and identically distributed normal random variables with mean zero and variance σ_ε^2. The quantile–quantile plots, autocorrelations and cumulative periodograms of the residuals are in good agreement with this. There is, however, a small number of clearly non-zero cross-correlations at lag zero. These are all less than 0.2 in absolute value, and reflect the fact that, as can be seen from Fig. 3, the assumed spatial correlation function (3.3) is not exact; this seems unavoidable. The most important check on the model for our purposes is the quality of the interval estimates it yields; this is investigated in Section 4.5.

In analysing another, one-dimensional, set of wind speed data, Lawrance and Lewis (1985) concluded, using techniques based on third-order moments, that their data exhibited non-linearity. Similar analyses, and also calculations based on fourth-order moments (Granger and Andersen, 1978), failed to reveal any such non-linearities in our data.

4.5. Interval Estimation at a New Site

Let us now consider the situation of Section 3, where we have a short run of data at a new site k, on the basis of which we wish to estimate μ_k. As we saw, the estimator $\tilde{\mu}_k$ defined by (3.4) performs well as a point estimator. An interval estimator now follows by noting that, conditional on the estimated parameters of the model (4.1), $\tilde{\mu}_k$ is unbiased and normally distributed with

$$\text{var}(\tilde{\mu}_k) = \sigma_X^2 \mathbf{a}_k^T \mathbf{R} \mathbf{a}_k \{ n + \sum_{j=1}^{n-1} (n-j)\rho_j \}/n^2 \qquad (4.9)$$

where ρ_j are the autocorrelations for the fractionally differenced ARIMA(p, d, q) process. However, (4.9) is inconvenient because the ρ_j are rather complex. A convenient

approximation to (4.9) is

$$\text{var}(\tilde{\mu}_k) \approx 2\pi f(0)\mathbf{a}_k^T\mathbf{R}\mathbf{a}_k\{n + \sum_{j=1}^{n-1}(n-j)\rho_j^{(d)}\}/n^2 \tag{4.10}$$

where the $\rho_j^{(d)}$ are the autocorrelations of an ARIMA(0, d, 0) process given explicitly by Hosking (1981), and $f(0)$ is the spectrum at zero of the estimated ARMA(p, q) model with parameters $\hat{\phi}(B)$ and $\hat{\theta}(B)$, and innovations variance σ_ε^2 (Fuller, 1976, Theorem 6.1.2).

To assess the accuracy of standard errors based on (4.10), we extended the cross-validation exercise reported in Section 3 to include them; see Table 1. The theoretical standard errors are quite accurate, and, in particular, they capture the non-standard rate of decline of the mean squared error of the empirical errors quite well.

It would, of course, be possible to use as a point estimator the exact maximum likelihood estimator of μ_k based on the model (4.1), rather than the kriging estimator $\tilde{\mu}_k$. However, this is a much more complicated solution, and some numerical work indicated that there is little to be gained from adopting it. This is not surprising, given the result of Adenstedt (1974) and Beran and Künsch (1985), that the loss of efficiency incurred by using the sample mean is small for one-dimensional long-memory processes.

5. Estimating Wind Power

The power P (in watts per square metre) in a wind with instantaneous speed V (in metres per second) is due to its kinetic energy, and is

$$P = \tfrac{1}{2}\lambda V^3 \tag{5.1}$$

where λ is the density of air, namely 1.227 kg/m^2 at 1013 mbar and 15 °C (Golding, 1976; Kaye and Laby, 1946, p. 34). Not all of this energy is available to a given turbine. Indeed, there is a theoretical upper bound of 0.593 (the Betz coefficient) to the proportion of this energy which may be extracted. The amount which is available to a particular turbine is a complicated function of V, the parameters of which are specific to that turbine. A realistic figure for a modern efficient turbine is of the order of 0.35. For maximum generality, therefore, we concentrate here on estimation of the long-run average of P at a new site. Brown *et al.* (1984) give an example of the computation of wind power from a specific turbine.

By (5.1), the average power in the wind at place i on day t is $\tfrac{1}{2}\lambda\overline{V_{it}^3}$, where $\overline{V_{it}^3}$ is the average cubed wind speed at place i on day t. However, we have formulated our model in terms of the velocity measures X_{it}. In order to deduce an empirical relation between $\overline{V_{it}^3}$ and X_{it} we constructed a log–log plot of $\overline{V_{it}^3}$ against $Z_{it} = X_{it} + s_t$, where s_t is the seasonal effect for day t. This is shown in Fig. 6, and is based on 180 days sampled randomly from each of the synoptic meteorological stations.

Fig. 6 suggests the approximate relation

$$E[\overline{V_{it}^3} \mid Z_{it}] = \gamma Z_{it}^\delta \tag{5.2}$$

with parameters which are approximately constant across place and time of year. We have taken $\delta = 5$, which accounts for almost as much of the variation (93.4%) as the
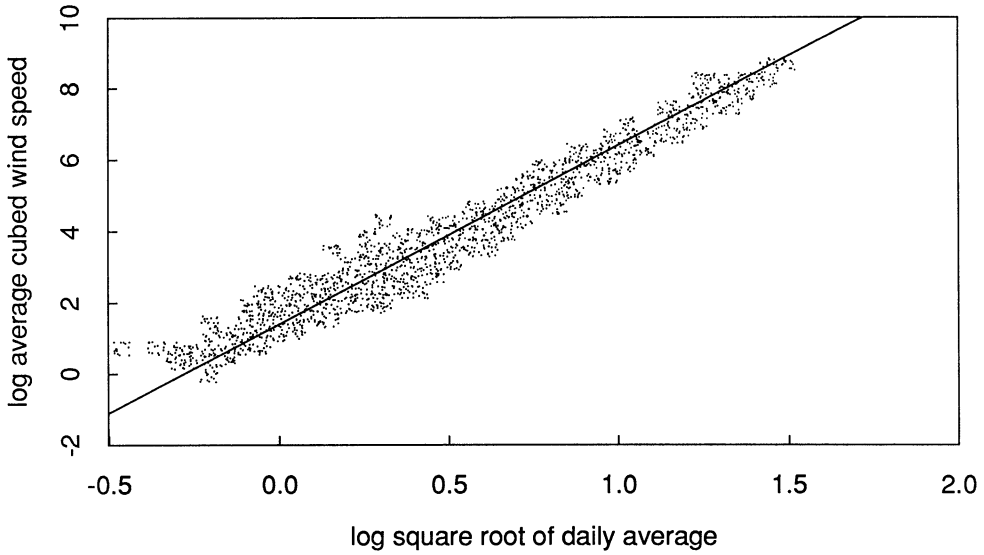
Fig. 6.   Plot of log (average cubed wind speed) (log $\overline{V_{it}^3}$) against log (square root of daily average) (log Z_{it}) for 180 days sampled randomly from each of the synoptic stations. The full line is the fitted line with slope $\delta = 5$.

best fitting line, which accounts for 94.3 % of the variation and for which $\delta = 4.6$. This yields the very simple result

$$E[\overline{V_{it}^3}] = \gamma E[Z_{it}^5]$$
$$= \gamma\{(\mu_i + s_t)^5 + 10(\mu_i + s_t)^3\sigma_X^2 + 15(\mu_i + s_t)\sigma_X^4\}. \tag{5.3}$$

The estimated value of γ is 3.63.

To obtain an estimate, $\hat{P}_k$, for the average power at place k, we replace μ_k by $\tilde{\mu}_k$ in (5.3), and average over all values of t for a year. This yields

$$\hat{P}_k = \tfrac{1}{2}\lambda\gamma \frac{1}{365.25} \sum_{t=1}^{366} \omega_t\{(\tilde{\mu}_k + s_t)^5 + 10(\tilde{\mu}_k + s_t)^3\sigma_X^2 + 15(\tilde{\mu}_k + s_t)\sigma_X^4\} \tag{5.4}$$

where ω_t is 0.25 for February 29th and 1.0 for all other days. Upper and lower confidence bounds for P_k may be obtained by replacing $\tilde{\mu}_k$ in (5.4) by the upper and lower confidence bounds for μ_k derived in Section 4.

In Table 2, we show the results of implementing this procedure for several examples involving each of the synoptic stations, a collection of starting values ranging over years and time of year, and several values of n. The 'true' values are obtained from (5.4) with $\tilde{\mu}_k$ replaced by $\ddot{\mu}_k$.

These results reflect the satisfactory performance of the interval estimator of μ_k. They also show how much harder it is to estimate mean power than mean speed. For example, in the example of the first line of Table 2, it is possible to estimate μ_k to within about $\pm 13\%$ at a 95% confidence level after 20 days, but mean kinetic energy can be estimated only to within a factor of about 2. The last line of Table 2 illustrates the fact that even when n is greatly increased, the uncertainty remains considerable.

To bring these results into more concrete form, we note that an average power of

TABLE 2
Estimated power (in kW/m²) based on short data runs at the synoptic stations. The point and interval estimates are obtained from the corresponding estimates of μ_k, using (5.4). The kriging point estimator of μ_k, given by (3.4), and the model-based interval estimator of μ_k, based on (4.10), are used.

Site	Starting date	n	Point estimate	'True' value	95% confidence limits	
					lower	upper
Malin Head	1 January, 1961	20	.37	.57	.19	.65
Roche's Point	5 February, 1962	20	.38	.35	.23	.62
Valentia	12 March, 1963	20	.21	.25	.11	.38
Kilkenny	15 April, 1964	40	.10	.09	.06	.16
Shannon	20 May, 1965	40	.24	.24	.16	.35
Birr	24 June, 1966	40	.10	.11	.06	.14
Dublin	29 July, 1967	80	.21	.21	.14	.31
Claremorris	1 September, 1968	80	.18	.15	.12	.26
Mullingar	6 October, 1969	160	.17	.16	.13	.23
Clones	29 January, 1971	160	.14	.16	.10	.20
Belmullet	8 April, 1973	320	.37	.39	.27	.51
Malin Head	22 February, 1974	320	.70	.57	.49	.96

0.37 kW/m^2 (as at Belmullet) corresponds to 3240 kW h/m^2/annum of energy. Thus, for a horizontal axis turbine with a 5 m blade, and a cross-section therefore of 79 m², and an average efficiency of 0.35, this corresponds to an energy production of about 90 MW h/annum on average. For comparison Irish electricity production in 1985–86 was about 10^7 MW h/annum (Electricity Supply Board, 1986).

6. Discussion

We have proposed a procedure for estimating wind power at a new site which gives reasonable answers and is easy to apply in practice. Inference is based on a simple and parsimonious approximating model which synthesizes deseasonalization, kriging, ARMA modelling and fractional differencing in a natural way.

We have focused here on the evaluation of the average power output to be expected in the long term from a wind turbine at a given site. We have ignored many questions however. For example, the basic data pertain to a standard height of 10 m, with

the exception of Malin Head at 22 m; the question of height extrapolation is discussed, with references, in Brown *et al.* (1984). Furthermore, other factors, such as variability, influence the value of the wind power resource to, for example, an electricity utility. Short term predictions of power available and required are of considerable importance for the control of an electricity grid. Such questions need model predictions at short timescales, of the order of one hour or less, and are not provided by this model; see Brown *et al.* (1984) and Lou and Corotis (1985). The long term variability of the resource is another key issue in discussions of questions such as the 'capacity credit' of a proposed wind farm; see Haslett and Diesendorf (1981) and Carlin and Haslett (1982). The model developed in Section 4 provides some basis for answering this and other questions about the resource. For example, it could be used for medium term prediction and control.

Our model is based on the assumption that the second-order moment structure of the space–time process is constant over Ireland (except the south-east corner). This assumption holds approximately, but not exactly, and it plays the important role of providing an estimate of the second-order properties at a new site for which there are not enough data to estimate them independently. Thus, the error bounds above must be somewhat optimistic, being conditional on the estimated values of all the parameters of the model, other than the mean, and on their being site independent. A natural way to relax it would be to cast the problem in a parametric empirical Bayes framework (Morris, 1983). This would be much more complicated, and some numerical work indicated that, in our case, neither exact knowledge of the second-order moment structure nor the imposition of a joint probability distribution on the μ_i would greatly improve precision over our method. However, it may well be worth investigating for other problems of this type.

We have not made use of the available wind direction information. Wind turbines are able to turn quite rapidly so as to be optimally placed for electricity generation with respect to the current wind direction. Thus, information on wind directions would be of use only indirectly in assessing the resource, if it enabled us to estimate the distribution of wind speeds with more precision; our calculations suggest that it would not. For example, we decomposed each wind speed into components parallel and perpendicular to the prevailing wind direction, an approach advocated by McWilliams and Sprevak (1985). The strong relationship between spatial correlation and distance shown in Fig. 3, which is crucial to our method, disappeared. Even assuming exact knowledge of all spatial correlations did not lead to appreciable improvement in the estimator of μ_k when wind speeds were decomposed into components.

One refinement which may well lead to increased precision is the incorporation of prior expert opinion about μ_k, usually that of a meteorologist. This can be done using Bayes' theorem. A simple, approximate, Bayesian solution follows by noting that posterior uncertainty about μ_k is much greater than that about any of the other parameters, and approximating the likelihood for μ_k by a normal density, obtained by assuming the other parameters to be known exactly. It can also be done, approximately, in a non-Bayesian way by regarding the prior mean as another estimator, and using the prior variance to combine the two estimators in the usual way. If the prior distribution is normal, these two approaches should give very nearly the same answer.

A variety of alternative approaches could be taken. For example, Deutsch and Pfeiffer (1981) outline a different approach to space–time modelling which introduces

spatial structure by ordering the neighbours of each site, rather than by modelling the spatial covariance structure. Their method seems less applicable to the present problem. Another approach is suggested by the meteorological plausibility of considering wind speeds to be governed by regimes which succeed one another according to a renewal process.

The evaluation of wind power has been considered for other countries, including Argentina (Barros and Estevans, 1983), New Zealand (Cherry, 1980), Denmark, the UK (Musgrove, 1987; Halliday, 1984), the USA (Pennell et al., 1980; Justus et al., 1976) and Spain (Adell et al., 1987). Many of these make use of meteorological models of the air-flow away from the Earth's surface. However, such studies have, typically, given little explicit indication of the precision of their estimates. Exceptions include Corotis (1977) and Barros and Estevans (1983).

There have been few studies of the spatial co-variability of wind speeds in this context. Exceptions include Balling (1984), Barros and Estevans (1983), Corotis et al. (1977) and Carlin and Haslett (1982); these have, however, adopted very simple descriptions only. The only detailed approach known to us of adjusting short series of wind speeds, by reference to longer series, is that of Barros and Estevans (1983), but this ignored the temporal autocorrelations. Such adjustments are well known in the hydrology literature as 'augmentation' procedures; see Vogel and Stedinger (1985) for a recent review. These do not, however, typically model the correlations spatially, and usually ignore even the short term temporal auto-correlation. Mollison (1980, 1986) has also used a similar idea to augment wave data by wind data.

One consequence of the long-memory property is that the gain in precision from extending records in time at one site rapidly becomes small. However, the availability of even small amounts of data at nearby places increases precision considerably, due to the strong spatial correlation. For the purpose of estimating the wind resource, it therefore seems worthwhile to collect wind speed data at a much denser grid of locations, perhaps using simple anemometers. This seems likely to be a more efficient use of resources than recording wind speeds for an extended period at a small number of additional locations. The results of Percival (1983, 1985) suggest also that there is not much to be gained for the present purposes by increasing the frequency at which observations are made. Of course, there are parameters of interest other than long run average power, such as extrema, concerning which we can make no such recommendations.

Acknowledgements

References

Adell, L., Zubiar, R., Martin, F., Ferrando, F., Moreno, P., Varona, L. and Pantoja, A. (1987) Development of a methodology for the estimation of wind energy resources in relatively large areas: application to the eastern and central parts of Spain. *Sol. En.,* **38**, 281–295.

Adenstedt, R. K. (1974) On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.,* **2**, 1095–1107.

Balling, R. C. and Cervany, R. S. (1984) Analysis of time and space variations in long term monthly averaged wind speeds in the United States. *Wind Engng,* **8**, 1–8.

Barros, V. R. and Estevans, E. A. (1983) On the evaluation of wind power from short wind records. *J. Climate Appl. Meteorol.,* **22**, 1116–1123.

Beran, J. (1986) Estimation, testing, and prediction for self-similar and related processes. Unpublished doctoral dissertation, ETH, Zurich.

Beran, J. and Künsch, H. (1985) Location estimators for processes with long range dependence. Unpublished Research Report No. 40, Sem. für Statist., ETH, Zurich.

Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis, Forecasting and Control,* revised edn. San Francisco: Holden-Day.

Brooker, P. I. (1986) A parametric study of robustness of kriging variance as a function of range and relative nugget effect for a spherical variogram. *Math. Geol.,* **18**, 477–488.

Brown, B. G., Katz, R. W. and Murphy, A. H. (1984) Time series models to simulate and forecast wind speed and wind power. *J. Climate Appl. Meteorol.,* **23**, 1184–1195.

Carlin, J. B. (1987) Seasonal analysis of economic time series. Unpublished PhD thesis, Department of Statistics, Harvard University.

Carlin, J. B., Dempster, A. P. and Jonas, A. B. (1985) On models and methods for Bayesian time series analysis. *J. Economet.,* **30**, 67–90.

Carlin, J. B. and Haslett, J. (1982) The probability distribution of wind power from a dispersed array of wind turbine generators. *J. Appl. Meteorol.,* **21**, 303–313.

Cherry, N. J. (1980) Wind energy resource survey methodology. *J. Indust. Aerodynam.,* **5**, 247–280.

Corotis, R. B., Sigl, A. B. and Cohen, M. (1977) Variance analysis of wind characteristics for energy conversion. *J. Appl. Meteorol.,* **16**, 1149–1157.

Cressie, N. (1985) Fitting variogram models by least squares. *Math. Geol.,* **17**, 563–586.

Creutin, J. D. and Obled, C. (1982) Objective analysis and mapping techniques for rainfall fields: an objective comparison. *Wat. Resour. Res.,* **18**, 413–415.

Deutsch, S. J. and Pfeiffer, P. E. (1981) Space-time ARMA modeling with contemporaneously correlated innovations. *Technometrics,* **23**, 401–409.

Electricity Supply Board (1986) *59th Annual Report.* Dublin: Electricity Supply Board.

Fox, R. and Taqqu, M. S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.,* **14**, 517–532.

Fuller, W. A. (1976) *Introduction to Statistical Time Series.* New York: Wiley.

Gandin, L. S. (1965) *Objective Analysis by Least Squares.* Israel Program for Scientific Translations.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.,* **4**, 221–238.

Gibbons, T. G., Haslett, J., Kelledy, E. and O'Rathaille, M. (1979) *The Potential Contribution of Wind Power to the Irish Electricity Grid.* SORL Report 7902, Statistics and Operations Research Laboratory, Trinity College, Dublin.

Golding, E. W. (1976) *The Generation of Electricity by Wind Power.* London: Spon.

Graf, H., Hampel, F. R. and Tacier, J.-D. (1984) The problem of unsuspected serial correlations. In *Robust and Nonlinear Time Series Analysis* (eds J. Franke, W. Haerdle and R. D. Martin), pp. 127–145. New York: Springer.

Granger, C. W. J. and Andersen, A. P. (1978) *An Introduction to Bilinear Time Series Models.* Goettingen: Vandenhoeck and Ruprecht.

Granger, C. W. J. and Joyeux, R. (1980) An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.,* **1**, 15–29.

Halliday, J. A. (1984) Analysis of wind speed data recorded at 14 widely separated UK meteorological stations. *Wind Engng,* **8**, 50–73.

Haslett, J. and Diesendorf, M. (1981) The capacity credit of windpower—a theoretical analysis. *Sol. En.,* **26**, 391–401.

Haslett, J. and Kelledy, E. (1979) The assessment of actual wind power availability in Ireland. *En. Res.,* **3**, 333–348.

Hosking, J. R. M. (1981) Fractional differencing. *Biometrika,* **68**, 165–176.

——(1982) Some models of persistence in time series. In *Time Series Analysis: Theory and Practice 1* (ed. O. D. Anderson), pp. 641–653. Amsterdam: North-Holland.

——(1984a) Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. Technical Report 2752, Mathematics Research Center, University of Wisconsin.

——(1984b) Modeling persistence in hydrological time series using fractional differencing. *Wat. Resour. Res.*, **20**, 1898–1908.

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

Justus, C. G., Hargraves, W. R. and Yalchin, A. (1976) Nationwide assessment of potential output from wind powered generators. *J. Appl. Meteorol.*, **15**, 673–678.

Kaye, G. W. C. and Laby, T. H. (1948) *Tables of Physical and Chemical Constants*, 9th edn. London: Longman.

Lawrance, A. J. and Lewis, P. A. W. (1985) Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. *J. R. Statist. Soc.* B, **47**, 165–202.

Li, W. K. and McLeod, A. I. (1986) Fractional time series modelling. *Biometrika*, **73**, 217–221.

Lou, J.-J. and Corotis, R. B. (1985) Prediction of generated wind turbine power with high frequency wind speed series. In *Time Series Analysis: Theory and Practice 7* (ed. O. D. Anderson), pp. 205–210. Amsterdam: North-Holland.

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

McWilliams, B. and Sprevak, D. (1985) Stochastic modelling of wind speed and direction. In *Time Series Analysis: Theory and Practice 7* (ed. O. D. Anderson), pp. 195–203. Amsterdam: North-Holland.

Mollison, D. (1980) The prediction of device performance. In *Power from Sea Waves* (ed. B. Count), pp. 135–172. London: Academic Press.

——(1986) Wave climate and the wave power resource. In *Proc. IUTAM Symp. Hydrodynamics of Ocean Wave-energy Utilisation* (eds D. V. Evans and A. F. de O. Falcao), pp. 135–156. Heidelberg: Springer.

Morris, C. N. (1983) Parametric empirical Bayes inference: theory and applications *J. Amer. Statist. Ass.*, **78**, 47–65.

Musgrove, P. J. (1987) Wind energy conversion—recent progress and future prospects. *Sol. Wind Technol.*, **4**, 37–50.

Penell, W. T., Barchet, D. L., Wendell, L. L. and Heister, T. R. (1980) Meteorological aspects of energy resources and selecting the sites. *J. Indust. Aerodynam.*, **5**, 223–246.

Percival, D. B. (1983) The statistics of long memory processes. Unpublished PhD thesis, Department of Statistics, University of Washington, Seattle.

——(1985) On the sample mean and variance of a long memory process. Technical Report No. 69, Department of Statistics, University of Washington, Seattle.

Raftery, A. E., Haslett, J. and McColl, E. (1982) Wind power: a space–time process? In *Time Series Analysis: Theory and Practice 2* (ed. O. D. Anderson), pp. 191–202. Amsterdam: North-Holland.

Ramsey, F. L. (1974) Characterization of the partial autocorrelation function. *Ann. Statist.*, **2**, 1296–1303.

Ripley, B. D. (1981) *Spatial Statistics.* New York: Wiley.

Tabios, G. Q. and Salas, J. D. (1985) A comparative analysis of techniques for spatial interpolation of precipitation. *Wat. Resour. Bull.*, **21**, 365–380.

Thiebaux, H. J. and Pedder, M. A. (1986) *Spatial Objective Analysis with Applications in Atmospheric Science.* London: Academic Press.

Vogel, R. M. and Stedinger, J. R. (1985) Minimum variance streamflow augmentation procedures. *Wat. Resour. Res*, **21**, 715–723.

Yajima, Y. (1985) On estimation of long-memory time series models. *Austral. J. Statist.*, **27**, 303–320.

Discussion of the Paper by Haslett and Raftery

Professor R. L. Smith (University of Surrey): This paper is an excellent example of the development of statistical methodology to solve a substantial applied problem.

The problem is typical of those which arise in what may loosely be termed the environmental sciences—by these I include such fields as hydrology, meteorology, air pollution and numerous problems with a biological flavour. As such, the methods used will be of interest to workers in all these fields.

The authors' approach incorporates many techniques. After initial exploratory analysis they propose a 'kriging' estimator for interpolation at a new site, exploiting spatial correlations. Further analysis leads them to identify a model incorporating long-range and short-range temporal correlations. The method of fitting, based on an approximate likelihood function, makes an original contribution to the computational aspect of time series models, and finally the model is applied, not without further difficulties, to the prediction of wind power.

In seeking some aspect on which to comment in more detail, my attention naturally fell on the long-memory aspects, which of all the authors' techniques are the least well understood at the moment. I therefore went back to the last time that a paper before this Society was substantially concerned with this theme, Lawrance and Kottegoda (1977), and found the following quotation:

'Long-term dependence has in the past been analysed using the rescaled adjusted range . . . ; the method has been propounded by Mandelbrot and Wallis . . . and so far it has no competitors.'

The rescaled adjusted range has not been nearly so prominent in the recent literature on this subject. Why did the method become fashionable, and why did it become unfashionable again?

Part of the reason, no doubt, lies in the introduction of the fractional differencing concept. Although there have been many theoretical papers on this subject, there are few containing really substantial applications, and this paper is to be welcomed if only for that reason.

Nevertheless, this approach is very much model dependent. The analyst who is uncertain whether to use a long-memory model at all may prefer a nonparametric robust approach to the estimation of d. One such has been proposed by Geweke and Porter-Hudak (1983). Assuming a spectral density of the form

$$f(\lambda) = O(\lambda^{-2d}), \qquad \lambda \to 0,$$

their method is based on the approximate linearity of $\log I_N(\lambda)$, the logarithm of the periodogram based on N observations, in $\log \lambda$. Roughly, they fit a least squares linear regression to $\log I_N(\lambda_{j,N})$ against $\log \lambda_{j,N}$ for $j = 1, 2, \ldots, n \, (\ll N)$, where $\lambda_{j,N} = 2\pi j/N$ is the jth Fourier frequency, and estimate $-2d$ as the slope of that regression.

This approach has some analogies with estimating the tail of a probability distribution. For example, under an assumption of the form

$$f(\lambda) = a\lambda^{-2d}[1 + b\lambda^c + o(\lambda^c)], \qquad c > 0,$$

it can be shown that the optimal n is of the order $N^{2c/(2c+1)}$, with corresponding mean squared error of the order $N^{-2c/(2c+1)}$. The calculation mimics Hall (1982) in the tail estimation context; Hall and Welsh (1984, 1985) have considered some other aspects of this.

There is a technical difficulty with this calculation; namely that the standard sampling properties of the periodogram (approximately independent and exponentially distributed ordinates at the Fourier frequencies) fail in the extreme lower tail under a long-memory model. This is also a technical gap in the paper of Geweke and Porter-Hudak, and may have something to do with the levelling-off of the periodogram in the extreme lower tails of the authors' Fig. 5.

Returning to the methodological aspects of the paper, in view of the wide range of potential applications it is worth examining some of the assumptions from a broader viewpoint than just whether they were justified for this particular data set. I had some doubts about both equation (3.3), where there is no allowance for any kind of directional dependence, and the constancy of autoregressive moving average coefficients across all sites in Section 4.1. Do the authors have any comments on whether such assumptions are likely to prove restrictive in trying to apply the model in other contexts? What alternatives are available?

Overall, this paper must be praised as a major piece of applied work, for the development of new methodology, for its contribution to the computational aspect of long-memory model fitting and not least for the theoretical developments that it will stimulate. It is an ideal contribution to the proceedings of this Society.

I have great pleasure in proposing a vote of thanks.

Professor Denis Mollison (Heriot-Watt University, Edinburgh): Where Richard Smith has discussed the theoretical content of the paper, I shall concentrate on the applied side. The problem addressed by the authors is indeed of practical importance, and their conclusion is somewhat depressing: even with nearly a year's data from a new site ($n = 320$), confidence intervals for the mean resource have a $\pm 30\%$ spread (Table 2), where we might have assumed an accuracy 4–5 times as great before they pointed out the importance of long term memory dependence (see Table 1 *et seq.*) Errors of this magnitude ($\pm 30\%$) would affect the unit cost of wind power by about $\pm 20\%$ (British Wind Energy Association, 1987), which could be crucial for a resource which is on the verge of economic viability.

The authors have mentioned possible improvements in accuracy based on the use of the same data set, such as the use of Bayesian priors. An alternative, exploiting our understanding of atmospheric dynamics, would be to use a hindcasting model such as that of the UK Meteorological Office (Golding, 1980), which has produced estimates for an approximately 50 km grid covering North-west Europe including Ireland since about 1978. Short period measurements for a specific site could be used to calibrate estimates from such a model, which might first be modified to take account of local topography.

In the other direction, an alarming possibility is that the wind climate may be appreciably non-stationary on the timescale considered (say 10–50 years). Carter and Draper (1988) have recently pointed out strong evidence for a significant increase in wave power for sites south and west of Ireland, possibly as large as a doubling of the mean resource over the period 1960–90. Admittedly they did not detect a significant change in wind climate at the sites that they considered, but since waves are generated by winds (mainly
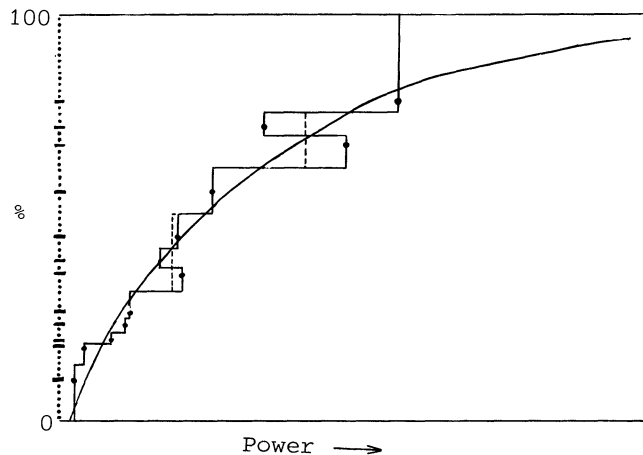
Fig. 7. Estimates of the distribution function of monthly mean power obtained (a) from the empirical distribution function of the set of predictors $\{W_i\}$ (smooth curve) and (b) by plotting data values P_i (●) against the same vertical scale, i.e. the position of the corresponding month i (——) among the order statistics of $\{W_i\}$ (schematic, based on data from Mollison (1980))

non-local; see, for example, Mollison (1986)) their work certainly implies that similarly significant changes could also occur in the wind power resource.

A small point, but of some importance, is that the seasonal variation has been assumed to be the same at all sites. It would be interesting to know whether the authors investigated this, and whether their conclusions might be sensitive to this assumption.

The authors' main model, with long term memory, is in the end only used for confidence intervals. The estimator itself turns out to be in reasonable agreement with their earlier estimator, which they therefore fall back on. The latter is essentially an average of the short term data weighted according to their simpler 'inverse covariance' model (equation (3.4)).

This encourages me to describe a model of my own (Mollison, 1980) for a similar problem, the augmentation of short term data on wave power by longer term wind information. The approach was rather different, but there are sufficient similarities that each may illuminate the other. My approach was initially based on a model for a wave power measure P_i, the average power observed in month i, in terms of a predictor based on the average value of the fifth power of wind speed W_i,

$$\ln P_i = k + \ln W_i + c_i.$$

Like the authors' equation (3.4) this is a linear relation between transformed values of short term and long term variables.

This parametric model yields estimates P_i for the longer period, and in particular an estimate and confidence interval for the mean wave power resource. For instance, with wave data for two years ($n = 24$) and wind data for 13 years ($N = 156$), the confidence interval was estimated at $\pm 13\%$. However, results were sensitive to the details of the model; the estimates P_i ranged up to more than twice the highest observed value, and thus the estimate of the mean resource was sensitive to the power of wind speed used in defining W_i.

A nonparametric alternative is to assume only that P depends monotonically on W. If this is the case, we can estimate the distribution function of P using all the values of W to determine the vertical scale, i.e. we plot P_i against the position of i among the order statistics of $\{W_i\}$ (Fig. 7). A non-decreasing estimate of the distribution function can be ensured by a monotone least squares regression (dotted line in Fig. 7).

This method has several advantages, apart from its minimum of assumptions. There is no need to estimate the relational parameter k, which is the main contributor to the uncertainty in our estimate of the mean resource, so it is not surprising that there is little if any loss of accuracy in the estimate of the mean resource. Indeed, simulations for my particular data set, admittedly with a slightly different

treatment of the highest end of the power range, gave a narrower confidence interval, $\pm 10\%$, than for the parametric model.

Perhaps the greatest advantage of the nonparametric method, however, is that it can be interpreted as giving weights to the short term data; namely, data month i is given weight proportional to the number of months in the ordered sequence $(W_{(j)})$ for which it is the closest data month. (A slight refinement is to share out weights equally where data months are in the wrong order, i.e. among months for which the monotone least squares regression takes the same value. Simulations suggest that this also slightly increases the accuracy of the estimate of the mean resource.)

The complete set of short term data can then be used, with these weights, as a representative resource sample; for instance, in the wind and wave power contexts such a set can be used to optimize device design (see, for example, Mollison (1980)). There should be no difficulty in extending this representation to the authors' case of a number of synoptic stations; their equation (3.4) essentially gives weights to the various synoptic stations and thus could be used to combine sets of weights derived as above for the individual stations.

The nonparametric method may fail to represent extreme conditions, especially in a sample where there are few observations in what, on the evidence of the background data W_i, were the most extreme months. I would argue that this is an advantage, in that it makes it clear that we lack this information; it is precisely in these circumstances that we would be unwise to rely on the parametric model. In particular, it indicates that where extremes are of interest, as in design survival tests, further data or different estimation techniques are required. However, knowledge of extremes is unnecessary for power output estimates, since almost by definition they will be beyond the output limit of economic devices.

There remains the problem of long term memory. Even taking monthly averages, the sequence (W_i) showed a (seasonally detrended) serial correlation of 0.2. In the light of the authors' analysis, it would clearly be desirable to reassess my estimates of confidence intervals.

The methodology of the paper has much wider generality than applications to renewable energy, but it is applications such as this which motivate developments in the methodology, and John Haslett and Adrian Raftery's exposition balances the interest of the two in a way that is most welcome. I have much pleasure in seconding the vote of thanks.

The vote of thanks was passed by acclamation.

Dr C. A. Glasbey (Scottish Agricultural Statistics Service, Edinburgh): I enjoyed this paper, which is an attractive blend of theory and practice, and a good example of the usefulness of statisticians.

A principal components analysis of the spatial covariance matrix gives an alternative perspective on its structure. Based on **R**, 80% of the spatial variability is accounted for by a daily average, and half of what remains by a linear gradient across Ireland. By way of comparison, I am involved with the Scottish Centre of Agricultural Engineering in studying local variability in solar radiation in the Pentland Hills, to the south of Edinburgh. We have also found a square root transformation to be appropriate for stabilizing variances. In our case, three-quarters of the spatial variability about a daily mean is explained by a linear gradient. Most of this variability is concentrated in a few days when either a north–south, or an across-the-ridge, effect occurs.

Meteorologists have a rule-of-thumb that about 30 years of weather data are optimal to represent current climatic variability, because longer periods are affected by drifts in climate. Arising out of this, how does long-term memory relate to climatic drift? And, would the authors have used 100 years of data if they had had them available?

Have the authors considered the possibilities which exist, for larger values of n, of increasing the robustness of inference. For example, elements in row k of **R** could be estimated, to guard against the 1 in 12 chance of being at another 'Rosslare'! Equation (5.3) looks highly sensitive to the normality assumption. An estimator constructed by resampling the data may perform better.

Two points of detail: I could not understand how it is possible that some of the data points in Fig. 6 correspond to $\overline{V^3} < Z^6$, and the results in Table 2 look unexpectedly good. If log-normal approximations are used and small correlations ignored, then the squared distance between the vectors of point and 'true' estimates is about 6. This lies in the lower 10% tail of a χ^2_{12} distribution!

Dr Eric Renshaw (University of Edinburgh): The authors are to be congratulated for presenting such a stimulating array of theoretical and practical aspects of space–time modelling. Long-range memory processes are of particular importance, and insight into their behaviour can be obtained by considering

spatial persistence through the interaction model (Bartlett, 1971, 1975)

$$X_{ij}(t + dt) = X_{ij}(t)(1 + \lambda dt) + \sum_{r,s=-\infty}^{\infty} a_{rs}\{X_{ij}(t) - X_{i+r,j+s}\} dt + dZ_{ij}(t) + o(dt)$$

where $\{X_{ij}(t)\}$ is a lattice process on $i, j = \ldots, -1, 0, 1, \ldots$ and $\{dZ_{ij}(t)\}$ denotes white noise. The associated spectrum (Renshaw, 1984) for equal $\{X_{ij}(0)\}$ is

$$f(\omega_1, \omega_2; t) = (\sigma^2/\psi)[\exp(\psi t) - 1]$$

where

$$\psi(\omega_1, \omega_2) = 2\left\{\lambda + \sum_{r,s=-\infty}^{\infty} a_{rs}[1 - \cos(r\omega_1 + s\omega_2)]\right\},$$

and this is especially useful for seeing how the form of the interaction weights $\{a_{is}\}$ determines overall spatial structure.

For example, the one-dimensional case with nearest neighbour weights $a_{-1} = a_1 = \alpha$, $a_r = 0$ (otherwise), leads to $\psi = 2\lambda + 8\alpha \sin^2(\frac{1}{2}\omega)$. If $\alpha < -\lambda/4$ then $\psi < 0$ for $\omega_0 < \omega \leqslant \pi$ where $\omega_0 = \cos^{-1}(1 + \lambda/2\alpha)$, whence $f(\omega; t) \sim -\sigma^2/\psi$ as $t \to \infty$. Thus ω_0 defines an 'outer scale of pattern' in the sense that if $\omega < \omega_0$ then $f(\omega; t)$ does not approach a stationary limit as $t \to \infty$. So changing λ enables us to alter the range of the scales of pattern present in the stationary part of the process. If $\lambda = 0$ then for small ω we have the inverse square law

$$f(\omega; t) \sim -(\sigma^2/2\alpha)\omega^{-2} \qquad (\alpha < 0),$$

while the Cauchy-type weights $a_r = k/[|r|(|r| + 1)]$ $(r \neq 0)$ yield pure '$1/\omega$ noise', i.e.

$$f(\omega; t) \sim (\sigma^2/2k\pi)\omega^{-1}.$$

Interest in spatial persistence requires us to extend this by constructing a process which possesses a genuine power law spectrum $f(\omega; t) \sim$ constant $\times \omega^{-d}$ for non-integer $d > 0$. This may be achieved by using a similar fractional differencing approach to the authors'. For the ARIMA(0, d, 0) process $x_t = (1 - B)^{-d}\varepsilon_t$ yields negative binomial weights which suggests putting

$$a_r = c \binom{|r| + d - 1}{|r|} \qquad (r \neq 0).$$

These give rise to

$$\psi(\omega) = 2\lambda - 4c[2 \sin(\frac{1}{2}\omega)]^d \cos[\frac{1}{2}(\pi - \omega)d],$$

and so

$$f(\omega; t) \sim [\sigma^2/4c \cos(\frac{1}{2}\pi d)]\omega^{-d} \qquad (\text{if } \lambda = 0)$$

as required.

Professor Knut Conradsen (Technical University of Denmark, Lyngby): I congratulate the authors on a stimulating paper that in an impressive way applies recent developments in time series and spatial statistics in the analysis of large data sets. The paper clearly demonstrates the importance of involving statisticians in work that is otherwise often done exclusively by physicists and engineers.

My comments relate to the problems around the spatial interpolation. In geostatistics we apply different types of minimum mean squared error estimates based on different models for the spatial auto-covariance. It is common folklore that the results of such an interpolation (a so-called kriging) are fairly insensitive to misspecifications of the spatial covariance structure; cf. the remarks following equation (3.4).

In Figs 8 and 9 are shown the kriging variance and the kriging weights in a simple kriging problem with three observations. The semivariogram is spherical with nugget effect c_0 and sill $c_0 + c_1$. We see that the kriging weights are fairly sensitive to changes in the relative nugget effect $c_0/(c_0 + c_1)$. Our experience with geochemical samples (stream sediments) is that this may have very serious effects whenever the data structure deviates from the model. In this sense, we should not consider kriging to be a fairly robust technique.

My second remark is related to the first, namely the question of a proper modelling of the spatial autocovariance. The authors have chosen the exponential given in equation (3.3). In the interpolations the behaviour of the autocorrelation close to zero is very important. In the region say between 0 and

Kriging   variance

Fig. 8

weight
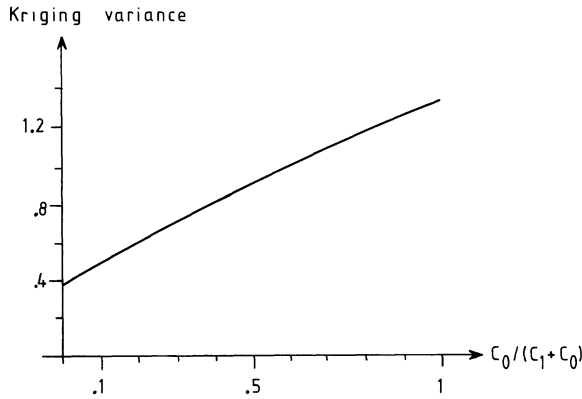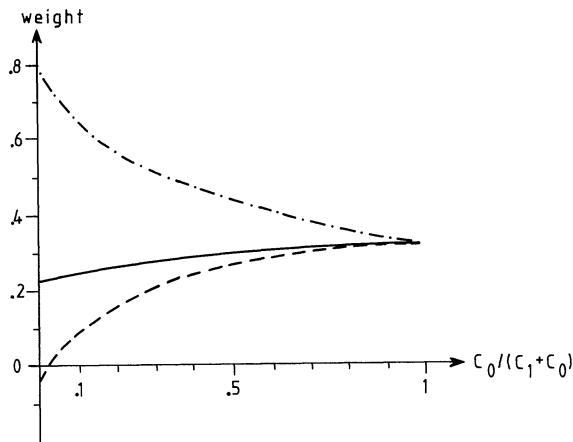
Fig. 9

50 km, however, the fit offered by the authors is not very adequate. A closer scrutiny of Fig. 3 shows that the correlations between 60 km and 100 km vary around 0.87, with no systematic decrease in that region. From two Danish meteorological stations separated only by 6 km a correlation of 0.87 was found (based on 7500 observations). If we add this observation and re-estimate the correlation structure, the outcome could be as in Fig. 10.

In actual interpolations this could be of importance. The model checking in the paper is based on a cross-validation technique, and therefore only correlations between sites with larger differences are used. It will be trivial to modify the correlation structure, and my remarks only serve the purpose of pointing out some possible pitfalls in modelling spatial data.

Dr I. T. Jolliffe (University of Kent): I would like to thank the authors for a stimulating paper, which uses, in an interesting way, some relatively recent ideas from time series analysis and spatial modelling on a real data problem. I have three comments, two of which relate to the somewhat strange behaviour of the data from the station at Rosslare. Without knowing anything about the siting of the station, it seems more likely to me that the difference between it and the other stations is due to local topography rather than to a regional effect. The main part of the discrepancy noted in the paper between Rosslare and the other stations is in the interstation correlations (Fig. 3), but it may be that it is the different autocorrelation structure at Rosslare (Fig. 4) which is the more fundamental difference. Consider the following (oversimplified) model involving two stations only.

Fig. 10

Let ε_{1t} and ε_{2t} be the noise terms for the two stations, each with variance σ_ε^2 and with

$$\text{corr}[\varepsilon_{1t}, \varepsilon_{2t}] = \varrho_\varepsilon.$$

Suppose that the velocity measures X_{1t} and X_{2t} follow AR(1) models

$$X_{1t} = \phi_1 X_{1t-1} + \varepsilon_{1t}$$
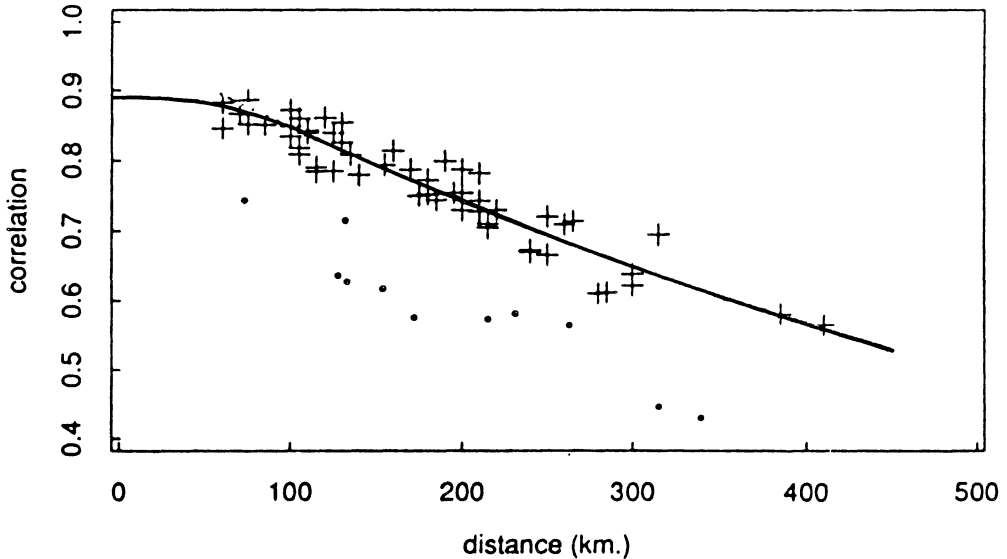$$X_{2t} = \phi_2 X_{2t-1} + \varepsilon_{2t}.$$

Then

$$\text{var}[X_{it}] = \sigma_\varepsilon^2(1 - \phi_i^2)^{-1}, \qquad i = 1, 2,$$

and

$$\text{cov}[X_{1t}, X_{2t}] = \sigma_\varepsilon^2 \varrho_\varepsilon (1 - \phi_1\phi_2)^{-1},$$

so the correlation between X_{1t} and X_{2t} is given by

$$\varrho_X^2 = \varrho_\varepsilon^2 \frac{(1 - \phi_1^2)(1 - \phi_2^2)}{(1 - \phi_1\phi_2)^2} = K\varrho_\varepsilon^2,$$

say. Now $K \leq 1$, and the amount by which ϱ_X^2 is shrunk relative to ϱ_ε^2 depends on the difference between the denominator and numerator of K, namely $(\phi_1 - \phi_2)^2$. There is no shrinkage when $\phi_1 = \phi_2$, but as ϕ_1 and ϕ_2 diverge so shrinkage increases. Thus, the smaller *cross*-correlation for Rosslare may be an indirect effect of smaller *auto*correlation. I would welcome the authors comments on this.

The second question regarding Rosslare is to ask whether the cross-validation exercise has been extended to predict the values for Rosslare. If the results are reasonable for this atypical site, it would increase confidence that worthwhile predictions can be made at new sites.

My final point is a brief question concerning model (4.1). The authors allow any past dependence of one X_{it} series on another to be explained entirely in terms of correlation between noise terms. To what extent is this less flexible than allowing direct dependence of X_{it} on X_{jt-1}, say, for $i \neq j$?

Dr C. Chatfield and Dr M. Yar (University of Bath): Given the mammoth nature of the project, the authors have done well to restrict the length of the paper but they have inevitably had to leave out some details, and our comments are mostly questions to clarify a few obscurities.

First a definition of 'synoptic' would avoid the need to look it up in the dictionary. Secondly, a brief description of 'kriging' would be enlightening. As we understand it, kriging is a two-dimensional interpolation and smoothing method, used in the mining industry, which is related to spline smoothing

(for example, see Wegman and Wright (1983)). Our third minor query is to ask why Fig. 5 presents periodograms rather than smoothed spectra which might be easier to interpret. A common vertical scale might also assist comparisons.

Our main query concerns equation (4.1) which assumes that the *same* univariate model is appropriate at each site, with the *same* ϕ and θ. We would like further justification of this assumption. We are also puzzled because in Section 4.2 the model appears to be fitted, not to the Xs (as implied by equation (4.1)), but to the fractionally differenced filtered Ys. As we understand it the same autoregressive (AR) filter of order 9 and the same d value is used for each series. How was the AR filter selected and what form does it take? This is one of the first reported cases of fractional differencing that we have seen, and we would also like further justification of this aspect. It is not obvious why the more usual differencing with an integer d value is not used. We suspect that fractional differencing arises from the shape of the (filtered?) spectrum near zero frequency, and that d is constrained to lie within the interval $[0, \frac{1}{2}]$ to obtain a finite variance.

Looking at Fig. 4, our first reaction was that there are substantial differences in the behaviour of the autocorrelation function (ACF) at different sites and that it is difficult to see 'striking similarities between its pattern and extent at the different stations' as suggested by the authors. At Rosslare, for example, the autocorrelations are 'small' at lags 5 or more and we see no need for any kind of differencing. However, at Clones, the ACF does not damp down to zero even at lag 100 and our first reaction is to take first differences, rather than fractional differences. A model for simple differences is easier to fit and to understand. If the same seasonal filter was used on each of the raw data series, could some of the long term persistence be induced by the imperfect nature of the seasonal filter? Returning now to the periodograms in Fig. 5, it is difficult to say whether they have similar properties or not (see our earlier comment on presentation). As the short-memory variation has been removed from each series, the periodograms are bound to look fairly similar in that variation is concentrated at low frequencies.

The final step in Section 4.2 says that a common ARMA model is identified for all the $\{\nabla^d Y_{it}\}$ but gives no indication how this is done. Was an AR(2) model identified for every single site, and, if not, how were the disparities between the selected models resolved?

Dr J. T. Kent (University of Leeds): I would like to congratulate the authors on a masterly application of ideas from spatial analysis, time series and long-range correlation to an important practical problem. My comments are directed to the initial data processing, which consists of three steps.

(a)   Start with the hourly average wind speed, $U(t)$ say.
(b)   Calculate daily averages, $\bar{U}(t)$, say.
(c)   Make a power transformation $\bar{U}(t)^{\alpha}$, with $\alpha = \frac{1}{2}$, to produce an approximate Gaussian time series.

Does the choice of power $\alpha = \frac{1}{2}$ depend on the scale of temporal aggregation, i.e. would $\alpha = \frac{1}{2}$ still be appropriate if weekly or monthly averages were used instead of daily averages? Related considerations arise in mining where log-normal spatial processes (corresponding to $\alpha = 0$) are observed. It is found that, to a good approximation, log-normality often persists over several scales of spatial aggregation; see, for example, Dowd (1982).

If we also take account of the average hourly wind direction then $U(t)$ can be regarded as the radial component of a two-dimensional wind velocity vector $V(t) = (V_1(t), V_2(t))$. The simplest model for the marginal distribution of $V(t)$ is bivariate normal with mean zero and isotropic covariance matrix, so that $U^2(t)$ is proportional to a χ_2^2 variate. The Wilson–Hilferty transformation of $U(t)$ to achieve approximate normality corresponds to $\alpha = \frac{2}{3}$. Further if the mean of $V(t)$ is non-zero we would expect a choice of α nearer to unity. Thus the fact that the preferred choice $\alpha = \frac{1}{2}$ is smaller than $\alpha = \frac{2}{3}$ suggests, perhaps not surprisingly, that the distribution of $V(t)$ is more heavily tailed than the normal distribution.

Steps (b) and (c) can be carried out in either order, i.e. we might transform before taking averages. Indeed we might have defined the initial data $U(t)$ to be the hourly average of wind speed to some power rather than of wind speed itself, especially as it is the cubed wind speed which is proportional to energy. Can the authors given some insight into their preferred ordering of steps?

Mr P. B. Brontë-Hearne (Crawley): I would like to draw attention to the power law equation. In finding a site for a wind turbine generator (WTG) there has to be a relationship between the type of WTG and the power law equation. No mention has been made of the height at which the wind speeds were measured. There has to be a relationship between the hub height of the WTG and the mean wind speeds

measured at that height. A simple formula used is (Halliday, 1983)

$$V_Z = V_H \left(\frac{Z}{H}\right)^n$$

where V_Z is the mean wind speed at height Z and V_H is the mean wind speed measured at height H, which is usually accepted as 30 m. As wind speed varies considerably according to the type of ground over which the wind passes n is a variable which depends on the nature of the terrain. This will have an effect on the suitability of the WTG selected. n can range from 0.1 (sand and ice) to 0.25–0.4 (for urban terrains) but is generally accepted to be $\frac{1}{7}$.

If isovents had been plotted for the various recording stations showing the availability of winds over a (perhaps yearly) period the Irish Government might have been more amenable to the provision of wind turbine generation.

Dr R. J. Bhansali (University of Liverpool): I would also like to congratulate the authors on an interesting and substantial empirical study. I have two brief questions. First, what checks did the authors make, apart from plotting the log-periodogram against the logarithm of frequency, before deciding that they are dealing with a long-memory model? Parzen (1983) has proposed an index for diagnostic checking of long-memory models. Are the authors aware of Parzen's work and have they experience of using this index?

Secondly, the authors tried to subject their data to the standard multivariate autoregressive moving average model fitting exercise and, if so, what sort of results did they find? Were they totally discouraging?

Professor Toby Lewis (University of East Anglia): I add my congratulations to the authors on a highly effective use of statistical methodology in the service of an important social need. I have a couple of tangential comments on aspects of the model.

First, regarding wind direction, there was the surprising observation in Section 6 that, when wind speed at each station was decomposed into components parallel and perpendicular to the prevailing wind direction, the relation between interstation correlation and distance d_{ij} disappeared. I do not know whether the correlations were calculated from signed components $v_i \cos(\sin)\theta_i$, absolute components $|v_i \cos(\sin)\theta_i|$ or square roots; in any case the non-dependence on d_{ij} seems counter-intuitive. Would the authors tell us a little more?

Secondly, a comment on Fig. 3 (which I offer in the spirit of 'lateral thinking'): the model (3.3) for r_{ij} in terms of d_{ij} fits well, but there is an outlier, Rosslare, already discussed by Dr Jolliffe and others; the correlations involving Rosslare are too low. However, we might equally say that the distances to Rosslare are too short! Take for instance point P, i.e. (Dublin, Rosslare), in Fig. 11. The distance from Dublin to Rosslare is only OP, but we would like it to be OQ, right up to the fitted curve. Then why not move Rosslare? If we draw circles on the map with centres such as Dublin and radii such as OQ, the desired new location for Rosslare emerges. In the spirit of Anglo-Irish entente, the new location proves to be in England—just. It is at Hartland Point on the north Devon coast (Fig. 12). Replotting the 11 Rosslare correlation points in Fig. 3 with distances adjusted to Hartland Point we obtain the points ⊙ in Fig. 11, now lying comfortably on or near the fitted relationship. Incidentally, the points R and S for Belmullet and Malin Head, lying a little off the fitted curve, could be brought nicely on to it if we shifted Rosslare, not to Hartland Point, but to the location marked * on Fig. 12. This is the Devon village of Sheepwash. But I feel that I should stay with Hartland Point, as more fitting to the *gravitas* of Dr Haslett and Dr Raftery's admirable paper.

The following contributions were received in writing after the meeting.

Dr Jan Beran (American Statistical Association, Bernoulli Society): This paper demonstrates once more the importance of long-range dependence for statistical analysis, in particular for the construction of confidence intervals. So far theory and applications have mainly focused on time series. Here we have spatial data, though the long-range dependence only occurs in the time dimension. The paper might stimulate research on long-memory processes with a more general index variable.

The computation of the confidence intervals does not take into account that d (and also the autoregressive moving average parameters) has to be estimated. Is the effect of estimation negligible? For instance for the location parameter of a process with a one-dimensional index variable such confidence
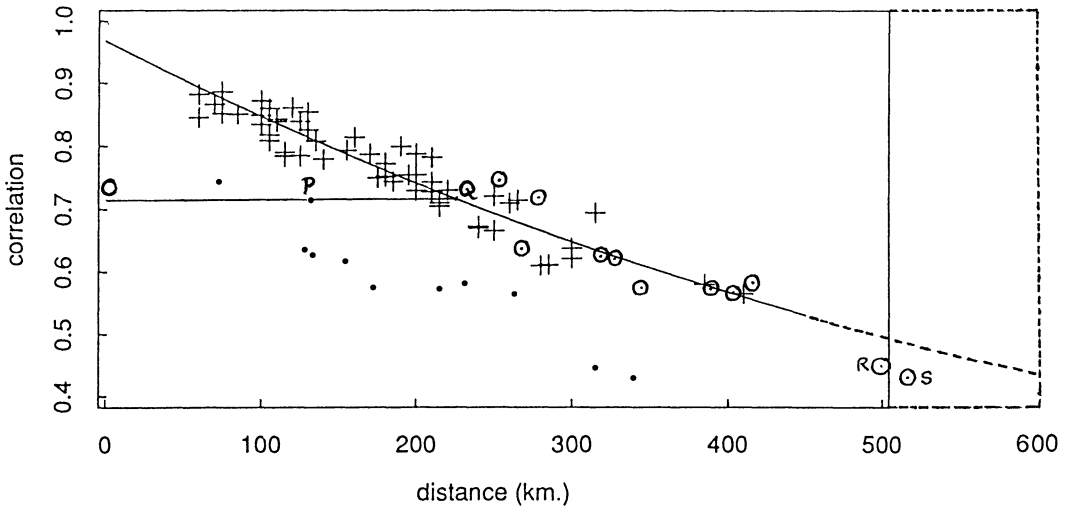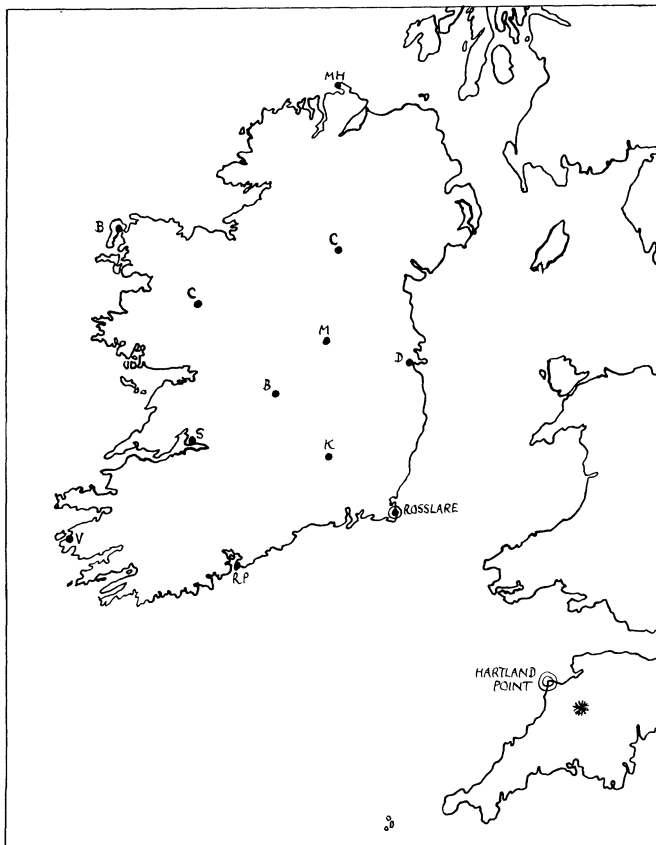
Fig. 11.   Extension of Fig. 3

Fig. 12.   Extension of Fig. 1

intervals are clearly too narrow so that the variability of d has to be built into the procedure. It might be possible to use similar techniques for the model considered in this paper. ·

Dr J. B. Carlin (La Trobe University, Bundoora, Victoria): The authors have synthesized several ideas from time series and spatial statistical modelling, in a novel and imaginative way, to address a practical problem of considerable difficulty.

A feature of the paper is the use of long-memory time series models. It is salutary to see such clear evidence in these data of the need for models that go beyond the finite spectra of the autoregressive moving average class. The analysis shows that inferences based on inappropriate short-memory models may be quite misleading for assessing the variability or uncertainty of estimates of long term levels. Unfortunately, with shorter time series, it may be much more difficult to assess the nature of low frequency variation by examining the data (i.e. d may be difficult to estimate). Nevertheless, we should be aware of the potential sensitivity in conclusions to such features of fitted models (Carlin, 1987; Carlin and Dempster, 1988).

On a more technical level, the authors have developed a new and apparently very successful method of approximating the likelihood of the fractionally differenced ARIMA(p, d, q) process. Further details justifying the method, as well as some systematic evaluations of its performance, would be welcome, as this could be a major contribution towards overcoming the computational difficulties that are a major constraint in the wider application of long-memory models. The authors' computational times are consistent with my experience. Even using the authors' approximation, maximum likelihood estimation seems bound to be computationally costly: it would be interesting to learn about the numerical maximization algorithm that they used.

Finally, a few comments about the applied problem: the authors' modelling success, as reflected by the almost uncanny agreement of the theoretical and empirical (cross-validatory) mean squared errors shown in Table 1, relies on some remarkable empirical regularities observed in their data. For instance, they argue that it is reasonable to assume a common seasonal pattern, and the same univariate time series structure, for each of their sites, as well as assuming the simple isotropic spatial dependence model (excluding the unfortunate Rosslare). These assumptions could be violated in countries other than Ireland, with its maritime climate and relatively low relief, so that caution must be exercised in the extension of these methods to other locations. Also, from a limited amount of data at a new, candidate wind power site, it might be difficult to assess whether or not the site has peculiarities like those of Rosslare. Here the input of expert meteorological knowledge would be important. Another feature that weighs heavily in the real-world conclusions of the study is the use of the simple model for expected power output, given by equation (5.2) and supported by the data of Fig. 6. This enables the authors to predict power output simply from an estimate of the long term mean of the square root of daily wind speed. Is there any physical rationale for equation (5.2), or perhaps empirical evidence to support it from other sources? Finally, in Section 5 one might assume that the quantity of ultimate interest, V_{it}^3, should be approximately a *continuous time* average: what is the justification for using the average of *hourly* wind speeds instead?

Professor N. A. C. Cressie (Iowa State University, Ames) **and Professor F. Pesarin** (Universita degli studi di Padova): We have several comments and questions.

(a) We cannot obtain the authors' data set from the published literature and encourage them to make it available for others to perform alternative analyses.

(b) Is there any advantage to analysing power directly, rather than building a model for wind speed and then converting to power?

(c) Why did the authors ignore two stations, Cork and Casement, from the 14 reported by Haslett and Kelledy (1979)? They are spatially close to Roche's Point and Dublin respectively and would allow verification of the small lag correlation behaviour assumed in equation (3.3).

(d) Choice of exponential covariance in equation (3.3) implies sample paths that are continuous (when there is no nugget effect) but not differentiable. At the scale of spacing of the synoptic stations this does not matter, but if wind turbines were to be clustered around centres of population small-scale sample path behaviour is important. If the fitted space–time model were used to simulate the wind speed at all scales, the answers may be inappropriate for certain questions at small scale. The rate of approach of the spatial correlation function to the abscissa could be checked by using data from Cork and Casement, two synoptic stations omitted by the authors.

(e)   We see spatial inhomogeneity in the time series of Fig. 4. Stations Valentia, Roche's Point and Rosslare do not have the same long-range dependence as the other stations. Was this seen in the diagnostics used on the residuals from the authors' model (4.1) (which assumes a temporal operator on spatially stationary errors that is homogeneous across space)?

(f)   Residuals are different from errors; residuals contain spurious correlations that bias estimation of the error correlation structure. The authors' 'original' data are residuals, having first been deseasonalized.

(g)   The seasonal component was assumed deterministic for all the calculations but is clearly estimated.

(h)   The authors make the point that, under long-range temporal dependence, there is little loss of asymptotic efficiency in using unweighted means. A similar phenomenon occurs in space; Kramer and Donninger (1987) give a result of this type for a simultaneous spatial autoregressive Gaussian process.

(i)   The wind speed data exhibit high spatial correlation, severely reducing the effective number of 'spatial observations'. Without the spatial homogeneity assumption referred to (and questioned) in (e), estimators would be highly variable.

(j)   We think that the term 'kriging estimator' is inappropriate. Kriging refers to prediction, which should be distinguished from estimation. Kriging is what is needed here, but estimation ignores the question of variability in the potential observations. Data are recorded using instruments that will be different from the turbines that will actually generate the power. Thus it is the *variability with regard to the turbines* that should be considered. This is known as the 'change of support problem' in the geostatistics literature and is ignored by considering inference on means.

A. P. Dempster (Harvard University, Cambridge): My comments are limited to the statistical modelling and are based on experience with similar time series models also estimated by maximum likelihood (ML) methods albeit only univariate and much shorter series. Readers may find a forthcoming paper by Carlin and Dempster (1989) more accessible than the paper by Carlin *et al.* (1985) and Carlin (1987).

A basic difficulty in dealing with 11 simultaneous and long ($n = 6574$) time series is the possible wild proliferation of parameters. The authors deal with this by ruthlessly enforcing parsimony, e.g. using common fixed seasonal patterns and common simple whitening filters, for all the series. The simple linear model for space correlation implicitly assumes that the pairwise cross-spectra are constant across frequency and all have zero phase shifts. While the extreme parsimony renders the ML method feasible, I wonder whether it is overdone, especially with such long series. In particular, could data analysis show dependence of correlation on frequency and perhaps location-related phase shifts at different frequencies?

My main comment is to question the authors' approach to long-memory dependence. It seems to me that the Fig. 5 periodograms of AR(9)-whitened series removes not only 'short-memory' dependence but also makes the spectra flat across 99% of the frequency range, i.e. from 0.005 to 0.5, and shows only a hint of increase across a further 0.8%, i.e. from 0.001 to 0.005. Thus only about 1/500th of the periodogram ordinates suggest further long-memory dependence, and sampling theory for these few points is not yet well understood, so they are difficult to interpret, leading me to question whether d can be safely estimated from the AR(9) residuals. In addition, the AR(9) series itself is quite capable of representing something indistinguishable from long-memory dependence via roots near unity.

A different criticism applies to the ML procedure and applies also to my own work with Carlin. The high apparent accuracy with which d is estimated results from, in effect, using the fractional differencing term in the model to shape spectra across the full frequency range 0–0.5. A very different value of d could be operating near zero frequency, yet the procedure could completely miss this fact. Indeed, the low frequency power need not be a power law at all. For example, it might have peaks near the 11- or 22-year sunspot cycles, yet the data would have no sensitivity. It is sobering that with so much data we really cannot identify important low frequency phenomena without strong assumptions. What are the practical implications for forecasting energy yields?

Peter Guttorp and Paul D. Sampson (University of Washington, Seattle): We raise two questions and propose an alternative, nonparametric approach to the authors' spatial covariance model which does not require a stationary or isotropic covariance structure and so obviates the *ad hoc* approach of eliminating Rosslare from the analysis.

The authors do not suggest an explanation for the long term memory evidence. Can it be related to climatological principles? Similarly, can meteorological theory be used to model the seasonal variation? This would seem more appropriate than fitting harmonics. One may want to use a local smoother with a higher degree of flexibility to estimate the seasonal term. The effect on the spectral estimates of a local smoother is less clear than that of harmonics. Perhaps some insight can be had using Mallow's (1980) concept of linear parts of non-linear smoothers.

We are developing a method for estimating non-stationary anisotropic spatial covariances from repeated observations at a set of stations (Sampson, 1986). Many applications require estimation everywhere. Since the estimator (3.4) does not apply for extrapolation to a location without pilot data, this requires a spatial analysis more closely related to standard kriging methods. We model *spatial dispersions* $v_{ij} = \text{var}(X_{it} - X_{jt})$ as a general function of the geographic locations of stations i and j by applying multidimensional scaling (MDS) to the matrix (v_{ij}) to obtain a new two-dimensional representation of the sampling stations in which the spatial dispersion function (or variogram) satisfies the common assumption of stationarity and isotropy (i.e. determined only by metric distances between station locations). Station pairs that are weakly correlated will be located relatively further apart in the MDS representation than they are geographically. We estimate the spatial dispersion v_{ij} between any two locations in the geographic plane using the composite of

(a)   the monotone relationship between spatial dispersion and the interstation distances in the MDS representation and
(b)   a smooth mapping between the geographic and MDS representations.

This mapping embodies the nature of the manifest anisotropy and non-stationarity; it can be depicted graphically using biorthogonal grids (Bookstein, 1978).

Applying MDS to the sample covariance matrix for the Irish wind power data, we obtained Fig. 13 (cf. Fig. 1). The stations around the coast are relatively further from the stations in the middle of the island, indicating that covariance between coastal stations and inland stations is weaker than that among inland stations. Rosslare is furthest displaced in accordance with its relatively weak covariance with all other stations. Fig. 14 displays the success of MDS in representing the dispersions v_{ij} as a function of distance in Fig. 13. The authors refer to some studies of robustness to misspecification of the spatial covariance structure. However, these are limited to misspecification of stationary structures. Part of the non-stationarity in these data is due to a gradient in the station variances: decreasing variance from the north-west to the south-east. Fig. 3 does not show this.

Our approach to spatial covariance cannot be directly integrated into the likelihood estimation framework of Section 4. However, the authors' maximum likelihood estimates of parameters of the isotropic spatial correlation function in equations (4.1) and (4.2) are little changed from the preliminary estimates obtained by regressing $\log[\text{corr}(X_{it}, X_{jt})]$ on d_{ij}. This suggests that we may simplify their estimation procedure by holding the parameters of the spatial covariance process in the likelihood fixed. Then the likelihood is expressed in terms of a fixed estimate of the spatial correlation matrix R for which we would propose substituting our nonparametric estimate. This estimate could be refined as necessary.

Dr John Henstridge (Perth): In a large applied project such as this there are always alternative possible approaches. Two occur to me.

First, the modelling of the series does not discuss the lagged cross-correlations. Given that the stations are several hundred kilometres apart and that weather patterns tend to move from west to east, I would have expected a delay of up to 12 hours between the west coast and east coast stations. This could be readily modelled using for example the spectral methods of Hannan and Thompson (1974).

Secondly, it is clear from the periodograms in Fig. 4 that the temporal persistence referred to is on a timescale of several years. (It could not be much less since the seasonal component has been removed and the AR(9) model would remove most of the variance over shorter periods.) In my experience with long term meteorological data such temporal persistence is likely to be due in part to changes in the measuring equipment and in the environment around the measuring station rather than in the weather. It is not unusual for stations themselves to be moved. However, the methods of this paper could be used to predict the daily velocity measures at each station from the measures at the other stations and the discrepancy between the actual and predicted records could be expected to highlight sudden changes in the mean. This can then be corrected if felt justified. It is likely that there would remain a long-memory dependence component but on a reduced scale.
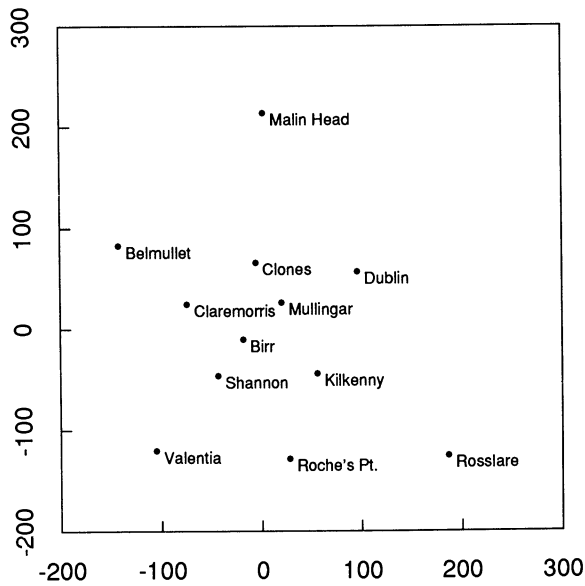
Fig. 13.   MDS representation of the IMS monitoring stations based on estimated spatial dispersions v_{ij}
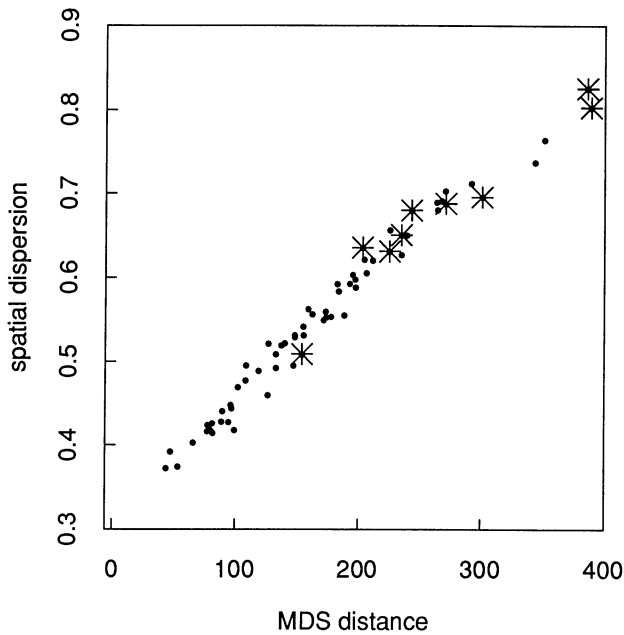
Fig. 14.   Plot of spatial dispersion v_{ij} versus interstation distance in the MDS representation: *, station pairs involving Rosslare

Dr D. A. Jones (Institute of Hydrology, Wallingford): Given the contrast in performance between short- and long-memory models, it would be interesting to include medium-memory models for consideration. Such models might reasonably be defined in terms of their partial autocorrelations. For example, for a model with three parameters a, b and c, let

$$\phi_{11} = a, \quad \phi_{22} = b, \quad \phi_{jj} = c \ (3 \leqslant j \leqslant M) \text{ and } \phi_{jj} = 0 \ (M < j),$$

where $M = 50$ or $M = 100$. This corresponds to an AR(M) process. An alternative model might allow ϕ_{jj} to taper linearly to zero, but sample estimates might suggest more appropriate behaviour.

Some of the difficulties reported with ARIMA(p, d, q) processes arise from the calculation of their partial autocorrelation functions: one possibility is to move to models that are parameterized directly via these functions, much as above, with a suitable behaviour for ϕ_{jj} as j increases. Modelling directly in terms of the partial autocorrelations would fit in with the authors' existing estimation scheme, while avoiding the need for approximations. The only disadvantage seems to be that the rather mesmeric statements of model structure, such as equation (4.1), are lost.

Dr Richard W. Katz (National Center for Atmospheric Research, Boulder): This paper provides a useful method for synthesizing several statistical characteristics that are typical of climatic variables such as wind speed. These characteristics include non-normal distribution, seasonal cycles and temporal and spatial correlation. The most novel aspect of this work concerns the issue of long-memory dependence. Models that possess long-memory dependence are sometimes considered in the water resources literature, especially as one possible chance mechanism to explain the origin of the so-called 'Hurst phenomenon' (Hosking, 1984). However, such models are not routinely considered by climatologists in fitting variables such as wind speed.

Convincing evidence is provided in this paper that taking into account temporal correlation (both short memory and long memory) is necessary for providing reliable standard errors in the estimation of mean wind speed. Climatologists are well aware of the need to correct for the effect of short-memory correlation on the standard error of time averages. In particular, a formula that is essentially a special case of equation (4.10), but ignores long-memory correlation, has been frequently employed in the meteorological literature (e.g. Jones (1975)).

Finally, stationarity on an interannual timescale has been assumed in all the analyses contained in the paper, but one of the issues in climatology over which the most controversy currently exists concerns whether or not the climate is undergoing permanent change (e.g. Wigley and Jones (1981)). Moreover, non-stationarity is an alternative chance mechanism to long-memory dependence for explaining the Hurst phenomenon (Bhattacharya *et al.*, 1983). Consequently, the conclusions of this paper relating to the efficient allocation of resources for measuring wind speed need to be qualified.

Professor Hans R. Künsch (Eidgenössische Technische Hochschule, Zurich): I was very pleased to see here another example of data which clearly exhibit long-range dependence. It is the first multivariate example that I know of. The model considered by the authors is a simple and useful subclass among the large number of possible multivariate models. It implies that not only all autocovariances and auto-spectra but also all cross-covariances and cross-spectra are proportional. I infer that the authors have checked this assumption at an exploratory stage.

The approximation to log-likelihood studied by Fox and Taqqu (1986) and Beran (1986) is Whittle's approximation. It is available also in the multivariate case, see Whittle (1953), theorem 6. For model (4.1) it equals

$$\log \sigma_\varepsilon^2 + \log(\det R) + \sigma_\varepsilon^{-2} \int |1 - e^{i\lambda}|^{2d} |\phi(e^{i\lambda})|^2 |\theta(e^{i\lambda})|^{-2} \sum_{j,k} (R^{-1})_{jk} I_{N,jk}(\lambda) \, d\lambda$$

where $I_{N,jk}$ is the cross-periodogram. Approximating the integral by a sum an evaluation of this expression should not take much central processor unit time.

Finally I would like to propose a slight variant of the estimator (3.4) and its approximate variance (4.10). For simplicity we take in the estimation problem of Section 3 $N = Mn$ and $t_0 = N - n + 1$. Other values of t_0 can be handled similarly. We consider the following estimator depending on coefficients α_j

$$\hat{\mu}_k = n^{-1} \sum_{t=t_0}^{N} X_{kt} + \sum_{j \neq k} \alpha_j \left(n^{-1} \sum_{t=t_0}^{N} X_{jt} - N^{-1} \sum_{t=1}^{N} X_{jt} \right).$$

Under model (4.1) the covariance between block sums $\Sigma_{t=1}^{n} X_{it}$ and $\Sigma_{t=sn+1}^{(s+1)n} X_{jt}$ for large n is approximately

$$\sigma_\varepsilon^2 r_{ij} c(\phi, \theta, d) n^{1+2d} \tfrac{1}{2}(|s + 1|^{1+2d} - 2|s|^{1+2d} + |s - 1|^{1+2d});$$

see Cox (1984). If these covariances hold exactly, the optimal coefficients α_j can be obtained easily. The variance of $\hat{\mu}_k$ is then equal to

$$\sigma_\varepsilon^2 c(\phi, \theta, d) n^{2d-1} [1 - u_M^2 / v_M (1 - a_{kk}^{-1})]$$

where $u_M = 1 - \frac{1}{2}M^{-1}[1 - (M-1)^{2d+1} + M^{2d+1}]$, $v_M = 2u_M - 1 + M^{2d-1}$ and $a_{kk} = (R^{-1})_{kk}$. The factor $1 - u_M^2/v_M(1 - a_{kk}^{-1})$ gives the decrease in the variance due to the information at other sites. Because u_M and v_M converge to unity rather slowly, it can be close to unity even if a_{kk}^{-1} is small, i.e. the spatial dependence is strong. This shows that the information from other sites is useful only if the records there are much longer than at the site of interest. The statement of the last paragraph of the paper thus seems to be too optimistic.

Dr W. K. Li (University of Hong Kong): I would like to concentrate my comments on the modelling aspect. In practice, it is rather unlikely that all m stations would exhibit the same long term and short term autocorrelation structure. Therefore model (4.1) appears to be a simplification and a more general model with d, $\phi(B)$ and $\theta(B)$ depending on i could be entertained. Of course, the modelling would become more difficult. In a recent report, Hui and Li (1988) consider fractionally differenced periodic processes where d or $\phi(B)$ are allowed to vary over different seasonal periods. The results may be applicable to the present problem. Since model (4.1) only makes use of the information provided by the distances between stations it is more akin to the so-called contemporaneous autoregressive moving average models studied by Camacho et al. (1987) than to a spatial time series over a rectangular lattice. Thus the approach of Mardia and Marshall (1984) may not be needed here. It also seems that some sort of approximations to ∇^d or the exact likelihood are unavoidable in practice and in my experience such approximations are quite satisfactory with sufficiently long records of data. Finally, the maximum likelihood estimate $\hat{\alpha}$ is rather close to unity although its approximate standard error is only 0.0013. Have the authors considered a model with α set equal to unity?

Professor Alan Lippman (Brown University, Providence): Raftery and Haslett have proposed a reasonable model for daily average wind speed in Ireland. The clean spatial correlation structure implied by Fig. 3 enables the authors to make effective use of a kriging-type estimator for the expected daily mean wind speed, which yields, at any location, good estimates based on little data. The model that they propose for the daily mean provides remarkably reliable estimates of the variance of the kriging estimate of the expected daily mean.

It is unfortunate, though understandable, that the authors could see no way to estimate the distribution of the wind speed (not the daily mean). If we had the true distribution of wind speeds it would be trivial to calculate the expected power production, as power production is a known, turbine-dependent, non-linear function of wind speed.

The authors instead use a clever two-part approach to achieve their goal, first modelling the daily mean and then using the model to estimate expected power production. It is on the second part, involving the use of the kriging estimate and its error bounds, that I would like to comment.

Although I am not well versed in the mechanics of turbines, the authors' assumption that power production is proportional to the power in the wind appears hazardous to me, as this ignores the effects of extrema. This is a point that the authors mention briefly but could prove important. Turbines shut down at high wind speeds. Ignoring this could lead to overestimating power production. I assume that the authors have already considered this, but I would be interested to see a modified Fig. 6, plotting log(power produced) (for a specific type of turbine) versus log(daily mean).

Granting that power production is proportional to the power in the wind, could an improvement be made in its estimation by using more than just the kriging estimate of the daily mean and its error bounds? It should be possible at a new site to estimate some statistics of the wind speed, for example the variance of the square root of wind speed. I pick this quantity since the authors observed that the square root of wind speed was approximately normal. An estimate of this variance, when used in conjunction with an estimate of the expected daily mean, might yield a better estimate of the expected cubed wind speed. A 20-day sample period yields 480 hourly samples, sufficient perhaps for a reasonable estimate of this variance, and while there would be seasonal effects to consider, I would not anticipate anything like long-memory dependence. So, another modification of Fig. 6, this time by adding a third dimension, variance of the square root of wind speed, might be revealing.

Professor K. V. Mardia (University of Leeds): The terminology of 'kriging estimation' in the paper could be somewhat misleading. Usually kriging is used for prediction whereas in the paper the term is used for parameter estimation. In fact, let $\mathbf{X} = (\mathbf{X}_1, X_2)'$ be $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the usual partitioning for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ where X_2 is the scalar variable at the new site. Then, from conditional expectation we have

$$\mu_2 = E(X_2|\mathbf{X}_1) + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1).$$

The authors' estimator $\hat{\mu}_2$ of μ_2 at the new site, given by equation (3.4), is obtained on replacing $\boldsymbol{\mu}_1$ in

the right-hand side of this equation by the sample mean of all the N observations, and $E(X_2|\mathbf{X}_1)$ and $\mathbf{X}_1$ by the sample means of X_2 and $\mathbf{X}_1$ based on the n observations respectively, $n < N$. Of course, the tools in both cases are similar as we are using

    (a)   the conditional expectation and

    (b)   a covariance scheme.

I do not believe that the robustness of $\hat{\mu}_2$ for values of α and β follows from the previous studies related to prediction. However, we might expect it to be true but, as has been pointed out by the authors, the variance of $\hat{\mu}_2$ will be definitely influenced by the estimated values of σ_X^2, α and β. Therefore an efficient method of estimation is desirable. It is common in geostatistics to plot semivariograms rather than correlation functions, particularly for processes which have stationary increments but are not stationary. Might the use of semivariograms also be fruitful for long-range correlations?

The authors indicate that by combining known results on asymptotic normality of Mardia and Marshall (1984) with others they could obtain similar results for their model. However, the nugget parameter causes some theoretical difficulty as it lies on the boundary of the parameter space. For a further discussion of this topic see Watkins (1988).

The authors removed the data at Rosslare in estimating α and β. This might indicate that there is some effect of the wind direction in general. The behaviour of 'co-kriging estimation' through wind velocity rather than just wind speed will depend heavily on the underlying cross-covariance structure. Which cross-covariance scheme was used by the authors?

Dr A. I. McLeod (University of Western Ontario, London): Contrary to a statement made at the beginning of the second last paragraph of Section 4.3, the asymptotic distribution of the parameter estimates in a univariate ARIMA(p, d, q) process with $|d| < 0.5$ has been derived by Li and McLeod (1986).

The model used by Haslett and Raftery can be viewed as a long-memory extension of the contemporaneous autoregressive moving average model of Camacho et al. (1987a, b).

Yosihiko Ogata (Institute of Statistical Mathematics, Tokyo): I am concerned about an implication of Fig. 5, i.e. that all periodograms in the figure have common peaks at the one-year period, in spite of deseasonalization of the data using the estimate in Fig. 2. This indicates that the seasonal effect at each station may not be quite the same as those at the other stations. I would like to describe a possible analysis for such a case in relation to the interpolation problem.

Consider the original data X_{it} as the spatiotemporal data $X(t_i, x_i, y_i)$ on $[0, T] \times A$, where A is the rectangular region of Fig. 1 including Ireland. Then consider a three-dimensional spline function $h(t, x, y|\mathbf{c})$ parameterized by $\mathbf{c}$. Since quite a few parameters will be required to obtain sensible estimates of the trend, I consider the penalized log-likelihood, where, besides the standard roughness penalties for the spline function

$$\Phi_1(h) = \int_A \int_0^T \left(\frac{\partial^2 h}{\partial t^2}\right)^2 dt\, dx\, dy$$

and

$$\Phi_2(h) = \int_A \int_0^T \left[\left(\frac{\partial^2 h}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 h}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 h}{\partial y^2}\right)^2\right] dt\, dx\, dy,$$

the seasonality constraint is given by

$$\Phi_3(h) = \int_A \int_{T_0}^T [h(t - T_0, x, y) - h(t, x, y)]^2 dt\, dx\, dy,$$

where $T_0 = 365.24$ days. Alternatively, we may regard the original data as the superposed spatiotemporal data $X(s_i, x_i, y_i)$ on $S \times A$, where S is the one-dimensional torus identical with $[0, T_0]$, and a very heavy weight is imposed on the penalty for the periodicity,

$$\Phi_3(h) = \int_A [h(0, x, y) - h(T_0, x, y)]^2 + \left[\frac{\partial h}{\partial t}(0, x, y) - \frac{\partial h}{\partial t}(T_0, x, y)\right]^2$$

$$+ \left[\frac{\partial^2 h}{\partial t^2}(0, x, y) - \frac{\partial^2 h}{\partial t^2}(T_0, x, y)\right]^2 dx\, dy.$$

To obtain suitable weights, I employ the Bayesian interpretation of the penalized likelihood (Akaike, 1979): the sum of the weighted penalties is considered to be proportional to the logarithm of prior probability density $\pi(\mathbf{c}|w_1, w_2, w_3)$ of the parameters $\mathbf{c}$, and the penalized log-likelihood is considered to be the log(posterior distribution). Then the marginal of the posterior (the Bayesian likelihood),

$$\Lambda(\sigma, w_1, w_2, w_3) \;=\; \int L(\mathbf{c}|\sigma)\pi(\mathbf{c}|w_1, w_2, w_3)\,d\mathbf{c},$$

is maximized to obtain the optimal weights.

The estimated spline function can be used for interpolating the seasonal effect at any locations. Further the so-called universal kriging procedure, subtracting the trend of the estimated spline, can then be carried out. However, assuming that the sample space of the spatiotemporal random field is restricted to a class of smooth spline functions, we have an alternative kriging method using the Gaussian posterior distribution of the parameter $\mathbf{c}$. See Ogata (1988) for a longer version of these comments and Ogata and Katsura (1988) for some details and numerical performance of the related spatial problems.

Dr W. D. Ray (Birkbeck College, London): Current statistical literature frequently deals with a far too idealistic model which is deemed to be sacrosanct; a theory is then developed to the finest detail with the pious hope that sometime, somewhere, data will be found to fit. It is nice to see a paper which is more data orientated and which checks out early features through exploratory analysis, to evaluate, for example, likely transformations and levels of aggregation. Another time series paper by Harvey and Durbin two years ago on seat belt legislation was also in this vein, but such contributions are not as common as they ought to be.

The paper is fairly self-contained and complete, but I have a few peripheral comments. I was surprised that the estimated seasonal effect in Fig. 2 required several harmonics; the scatter seems to indicate that fewer would have sufficed. The striking homogeneous short-memory autocorrelations of Fig. 4 are remarkable, particularly the positive aspects. So too is the common pattern of low frequency–long-memory persistence in Fig. 5, hence the need for fractional differencing, and these data provide a good example of its necessity.

The commonality feature of the wind data at the synoptic sites in Ireland is fortunate to allow the relative simplicity of models (4.1) and (4.2), but this feature may not be present in other applications when some clustering may be necessary.

It was not too surprising that the numerical aspects of maximum liklihood estimation are a problem here, a factor which also becomes acute when handling non-linear times series with large data sets. Thus the approaches to obtain approximations are to be commended.

The comment in Section 4.4 that non-linearities were not present in these wind data could have been amplified by providing a few statistics, which could then have been useful for future researches. The agreement of the mean squared errors from equation (4.10) with the empirical results seems rather flattering to the approximation.

This work is a very good example of time series modelling carried out in the true spirit of data leading the way. The classes of models (4.1) and (4.2) are sufficiently wide to be of use in a greater variety of applications, and probably will.

Dr E. M. Scott (University of Glasgow): I would like to make some comments on related problems.

My first comment concerns the non-Bayesian nature of the analysis. Given the nature of the problem (and others in environmental sciences), it is likely that prior information on a specific site would be available and that potential covariates might exist, which could and should be incorporated in the analysis.

Secondly, an important problem, not tackled in the paper, involves the question of the siting of the synoptic stations and whether there might be any possibility of developing the modelling approach to identify 'optimal' sites for wind farms, which could then be investigated in more detail.

Finally, the removal of the 12th station from the analysis raises interesting questions concerning the coarseness of the synoptic site grid relative to the degree of spatial variability in wind over a large geographical area.

There must be many sites where the global wind model is difficult to apply because of local conditions.

How should we balance siting and number of synoptic stations with the spatial variability of the response?

Professor Michael Stein (University of Chicago): Standard asymptotic results often do not apply in a spatial context. For example, in Section 3, the authors state that the estimate $\tilde{\mu}_k$ will be 'approximately normally distributed in large samples' even if the observations are not jointly normally distributed. However, the phrase 'large samples' is quite vague, and could refer to either N, the number of days, or m, the number of sites, or both, being large. If m is large but N is not, then there is no reason to think that $\tilde{\mu}_k$ will be approximately normally distributed, despite the fact that the 'sample size' mN is large. A second example is in Section 4.3, where a reference to Mardia and Marshall (1984) is made to support a conjecture that the maximum likelihood estimates of the parameters in equation (4.1) will have the usual asymptotic normal distribution. The result of Mardia and Marshall (1984) requires that the size of the observation region grows as the number of observation sites grows. In the present problem, the observation region, Ireland, is unlikely to grow to satisfy someone's theorem. Stein (1987, 1988) considers inferences for spatial processes based on an increasing number of observations in a fixed region. In any case, the model given by equation (4.1) can be thought of as a multiple time series model, and I would guess that the parameter estimates are asymptotically normal as N increases.

Another problem that I would like to raise is making inferences about a spatial correlation function over distances less than the shortest distance between any two observation sites. Beyond the restriction that correlation functions be positive definite, there is no logical constraint on the form of the correlation function over these distances. In particular, Fig. 3 shows some evidence that the correlation function flattens out over shorter distances, in which case the authors' estimate of the nugget effect would tend to be too small. Although misspecification of the form of the correlation function over these distances would not affect the results of the authors' cross-validation studies, it would affect inferences at a new site which was very close to one of the existing sites.

Professor Paul Switzer (Stanford University): When estimating the value of spatial processes at unobserved sites from data at observed sites the specification of the spatial correlation structure can be of major importance. The approach used by Haslett and Raftery is to approximate the contemporaneous correlation r_{ij} between any two sites i and j by a fitted exponential function of the corresponding intersite distance. Such a smoothing and parameterization of spatial correlation has two immediate advantages —it allows reasonable estimation of the spatial correlation structure when there is little or no time replication and it gives the needed estimates of correlations between observed and unobserved sites.

However, when there is substantial time replication, as there appears to be with these Irish wind data, then the r_{ij} will be well determined for every pair of existing sites. These well-determined intersite correlations will typically not all agree with any simple parametric function of intersite distance. Indeed, it is noted in Fig. 3 that correlations involving the Rosslare site fit poorly to the assumed exponential correlation model, and this station is removed from subsequent analyses. If there are potential sites of interest nearby, the removal of Rosslare from the analyses could constitute an important waste of available data.

Considering the substantial amount of time replication that is available from these data, it would be preferable to avoid parameterizing the correlations between the existing 12 sites. In the absence of a purely distance-dependent correlation model we need an alternative method to estimate correlations between the data sites and potential unobserved sites. A suggestion for such a program has been made by Switzer (1988). The suggestion uses both the fitted parametric correlation model and the directly estimated correlations between data sites for this.

Specifically, let R and $\tilde{R}$ respectively be 12×12 correlation matrices between pairs of sites, the first estimated directly from each pair of observed time series and the second obtained from the fitted exponential correlation model, say. Further, let R_k^* and $\tilde{R}_k$ respectively denote 12×1 correlation vectors between the putative site k and each of the 12 data sites, the first given by the expression below and the second obtained from the exponential correlation model. As the putative site k approaches an observed site i, then the proposed R_k^* vector coincides with the ith column of the directly estimated correlation matrix R. Other properties of the proposal are described in Loader and Switzer (1988). The proposal is

$$R_k^* = [R\tilde{R}^{-1}]\tilde{R}_k.$$

Winson Taam and Brian S. Yandell (University of Wisconsin–Madison): The authors have chosen to use an exponential structure to model the spatial dependence among these unequally spaced weather stations. They also indicated that another approach would be to collect data on a denser grid of locations. Given an equally spaced rectangular lattice, the space–time model will be essentially the same as that discussed by the authors except that the spatial structure is being modelled by a specific class of spatial

models in place of the exponential correlation structure. In particular, the spatial correlation can have a spatial autoregressive moving average structure defined in Besag (1972) or Tjøstheim (1978). We need to estimate the covariance matrix for the likelihood estimation. Because of the regular grid structure, we can use a torus to approximate the covariance $\mathbf{R}$. Taam (1988) has indicated the approximation rate for that spectral approximation. The advantages of this approach include modelling the local spatial dependency, simplifying the computation of likelihood estimates for the spatial portion and representing the spatial structure in spectral terms. This last feature can answer the question that Haslett and Raftery asked at the end of Section 4.3. This approach is one way to handle the boundary problem when a likelihood estimation is used. The fractional differencing may still be used in the temporal part of the model because we have proposed an alternative way to model the spatial part of the model if the data were collected from a rectangular lattice.

We could relax the parametric nature of the Haslett–Raftery model by setting the problem in a Bayesian context of multivariate smoothing splines (Wahba, 1983, 1985). Consider the model

$$X_{it} = f_i(t) + \varepsilon_{it}$$

with ε_{it} independent identically distributed normally with variance σ_ε^2 and $f_i(t)$ having a multivariate normal distribution in time and space. The covariance for $f_i(t)$ could be

(a) completely general (symmetric non-negative definite, but no further structure),
(b) a Kronecker product of a spatial and a temporal covariance or
(c) a Kronecker sum of a spatial and a temporal covariance.

Case (b) includes the model considered by Haslett and Raftery as a special case. Model (c) is much simpler, with correlated means but no cross-correlation over time. This hierarchy of models provides a framework for testing model adequacy and avoids the parametric assumptions made in the paper. This nonparametric approach may be viewed as an exploratory method to identify a model, or as a means to confirm the adequacy of a parametric model (Cox et al., 1988). The computational cost will be considerable. Bates et al. (1987) provided a general algorithm for multivariate smoothing splines and indicated that, without special attention to the design, computation becomes prohibitive on a VAX computer with over 400 data points. We can use the ideas in Yandell (1988) on block diagonalization to modify one-dimensional spline code (Hutchinson, 1984; Reinsch, 1967) to compute estimates for (c) quickly. The same idea may also help to reduce computation for case (b).

Professor D. M. Titterington (University of Glasgow) **and Mr P. Jamieson** (James Howden & Co. Ltd): We should like to comment briefly on the body of the paper and to make further remarks about an aspect of wind power referred to right at the end of Section 6.

The first comment is to continue the Rosslare saga. No matter where the port is relocated as a result of the paper and discussion, the Rosslare data should surely be incorporated at some stage. Fig. 3 suggests that this should be feasible, using a different value of β.

The second remark is to wonder whether or not the methods of the paper can be developed to create contour maps of wind speed and/or direction. With the incorporation of the time variable, these could lead to fascinating animated films of the wind behaviour over Ireland.

Of more interest to us, however, is the problem of high winds and the associated loadings imposed on wind turbines. In view of the high cost of these machines and the length of time (about 25 years) envisaged for their period of service, it is very important to be able to predict long term extremes of wind and to translate these into extremes of stress on the turbines. While there are adequate models for the stress from the literature on structures, the complicated statistical description of wind speeds at even a single location precludes the availability of analytical solutions, so far as extreme wind speeds are concerned. Our investigations so far have accordingly taken the form of simulation exercises.

Professor H. Tong (University of Kent at Canterbury): Spatial time series models are as important as partial differential equation models in the physical sciences.

In the authors' approach, spatial dependence is modelled in equation (4.1) via the ε_{it}s. This is similar in spirit to the 'diagonal' approach of Chan and Wallis (1978) in multiple time series. In the present context, $E[X_{it}|\mathbf{X}^{t-1}]$ does not depend on $\mathbf{X}_j^{t-1}$, $j \neq i$. Am I correct in suspecting that this could be a serious constraint? Without nonparametric regression estimates of these, I could not tell whether substantial information might be lost due to the assumption. I suspect that it would if the new station is close to one of the synoptic stations and if the timescale is short. $E[X_{it}|X_{js}]$ could also be non-linear.

It is rather artificial and time consuming to model long-range memory by fractional differencing. I feel that a Markovian model such as a non-linear autoregression (NLAR) would be a much more natural way to go about it. The difficulty is that it is not so easy to identify a suitable NLAR model. H. Künsch, D. Tjøstheim and I have been experimenting with NLAR models of the following form with this objective in mind:

$$X_t = X_{t-1} + \alpha I(X_{t-1} \leq 0) - \beta I(X_{t-1} > 0) + \varepsilon_t$$

($\alpha > 0$, $\beta > 0$), where I is an indicator function. (The model is a random walk model if $\alpha = \beta = 0$). It is ergodic. The hope is that it is neither geometric ergodic nor mixing. Unfortunately we were unable to reach any conclusion.

In addition to Fig. 5, it would be informative to have periodograms before the AR(9) filter.

Dr Andrew Walden (BP Exploration Co. Ltd, London): The long-memory temporal dependence raises some interesting questions. The authors acknowledge the main problem in recognizing long-memory dependence, i.e. it is difficult or impossible in practice to distinguish between spectral shape caused by truncating the autocovariance function of a long-memory process (through the use of a finite sample) from spectral shape arising from a process which does not satisfy the long-memory model. Several of the spectra of Fig. 5 show decay rates of 12 dB/octave (i.e. f^{-4}) at a frequency as low as 0.0005. By restricting d to $0 \leq d \leq 0.5$, the authors implicitly restrict frequency decay rates to be no greater than f^{-1} at such low frequencies. Do the authors feel that this problem is sufficient explanation of this discrepancy? Did they consider spectral approaches to the estimation of d such as that of Janacek (1982)?

It is interesting to consider physical mechanisms for red noise spectra similar to those seen in Fig. 5. An ensemble of purely random processes, each with an autocovariance of the form $\exp(-|\tau|/\tau_0)$ and its own correlation time τ_0 can generate red noise spectra with differing decay rates in different frequency ranges depending on the distribution of τ_0. This has been used to model the river level at the mouth of the Nile (Montroll and Shlesinger, 1982) for which the predominant decay is f^{-1}. Mechanisms for higher decay rates are discussed in Halford (1968).

The **authors** replied later, in writing, as follows.

Our project had a specific goal, namely the estimation of the mean kinetic energy in the wind at a site for which only a short run of data is available. To do this, we produced a model which was easy to apply at a new site, exploiting the remarkable empirical regularities highlighted by Dr Carlin and Dr Ray. We could have developed a more complicated model which might have better described some fairly minor features of the synoptic data, but this would have made the method harder to apply at a new site, and numerical work referred to in Section 6 indicates that it would not have improved the results. Modelling the existing data was not an end in itself.

Nevertheless, Professor Smith rightly says that the wide range of potential applications justifies looking for models more general than equation (4.1). Indeed, more than half the discussants suggested ways of elaborating the model. Equation (4.1) is a special case of the general model

$$\Phi(B)(\mathbf{Z}_t - \mu - \mathbf{s}_t) = \nabla^{-d}\Theta(B)\varepsilon_t. \tag{1}$$

In equation (1) $\mathbf{Z}_t$ is the vector of undeseasonalized velocity measures on day t, $\Phi(B) = I - \Theta_1 B - \cdots - \Phi_p B^p$ and $\Theta(B) = I - \Theta_1 B - \cdots - \Theta_q B^q$, where $\Phi_1, \ldots, \Phi_p$ and $\Theta_1, \ldots, \Theta_q$ are $m \times m$ matrices such that the zeros of the determinantal polynomials $|\Phi(B)|$ and $|\Theta(B)|$ are outside the unit circle, $\mu = (\mu_1, \ldots, \mu_m)^T$, $\mathbf{s}_t = (s_{1t}, \ldots, s_{mt})^T$ is a vector of seasonal effects, $\nabla^{-\mathbf{d}} = (\nabla^{-d_1}, \ldots, \nabla^{-d_m})^T$, and

$$\varepsilon_t \overset{\text{IID}}{\sim} \text{MVN}(0, V).$$

As Dr McLeod points out, equation (1) is an extension and synthesis of many proposals in the literature, most of which are cited in Camacho *et al.* (1987a).

If equation (1) is unconstrained, parameters proliferate wildly, as Professor Dempster has noted. Each parameter in equation (1) is associated with either a single site or a pair of sites and so may be constrained to be a function of position and/or (directed) separation which is either

(a) constant,
(b) deterministic and parametric,
(c) deterministic and nonparametric or
(d) stochastic.

Our model (4.1) is based on constraints of types (a) and (b), while several discussants suggest constraints of type (c). Stochastic constraints lead to parametric empirical Bayes models (Deely and Lindley, 1981; Morris, 1983). This is intellectually the most satisfying approach, but it is also the most difficult, and only Professor Ogata has had the courage to tackle it.

Model (1) encompasses virtually all the suggestions for model elaboration made by discussants. With suitable adaptation, the methods of statistical analysis developed in Section 4 may be applied to it.

Data analysis

Dr Kent's comparison of the square root transformation at different levels of aggregation with the log-normal transformation elsewhere is perceptive; Carlin and Haslett (1982) found this effective for hourly data. He is correct in his surmise that transforming and aggregating could have been performed in reverse order. This might indeed have led to a simpler approach than in Section 5, as implicitly sought by several contributors concerned with power considerations. However, one of the practical criticisms levelled at our solution by our meteorological colleagues is that our method, developed for data disaggregated to the level of days, is applicable with difficulty to several valuable short runs of data that are already available, but published solely as means, and to data that might be collected by particularly cheap 'run-of-the-wind' anemometers which simply return a mean wind speed for the observation period. Such data cannot be disaggregated to days, never mind hours, before transformation. Our general approach can be used for such data, but the details of the method require modification.

Dr Kent's components of wind speed model has been used in the literature (McWilliams *et al.*, 1979) for hourly data. Almost uniformly preferred is the Weibull model, and Carlin and Haslett's (1982) square root transformation is related to a classical transformation of Weibull data to normality (Dubey, 1967; Johnson and Kotz, 1970).

Professor Guttorp and Professor Sampson ask whether seasonal variation could be modelled using meteorological theory. We know of no way of doing this. Wind arises because of temperature differences, so the (relatively weak) seasonal pattern in wind speeds is related to a superposition of (usually much stronger) temperature patterns at different places. This, together with atypical wind patterns around the equinoxes, may suggest a meteorological explanation for the need to use several harmonics which troubled Dr Ray.

For simplicity and ease of application at a new site, we assumed the seasonal effect to be constant throughout Ireland, although, as Professor Ogata points out, there are slight differences between stations. His proposals for modelling these differences are interesting, and we hope that he will try them out on our data.

Rosslare

Rosslare is an outlier because the correlations with the other stations are too low. We simply removed it from the analysis. Professor Switzer points out that, if there are potential sites of interest nearby, this could be an important waste of data. In Ireland, the main sites of interest for wind energy are in the west and the north-west, so that the removal of Rosslare in the south-west is not a problem.

Of course, if the outlying station had been in a location of interest for wind energy, we could not have dealt with it so simply. Professor Lewis proposes an excellent practical way of overcoming the difficulty which, combined with Professor Titterington and Mr Jamieson's suggestion of a change in β, suggests a whole battery of *ad hoc* ways of dealing with isolated particularities in spatial covariance structures. Professor Guttorp and Professor Sampson and Professor Switzer outline more general methodologies for dealing with non-stationarities in the spatial covariance structure, on which we comment later.

Dr Jolliffe speculates that the unusual behaviour of Rosslare may be due to local topography rather than to a regional effect. The meteorologists, frankly, are puzzled. The station is sited somewhat unfortunately in that the winds from the prevailing direction tend, rather more than should be the case in ideal circumstances, to pass over the village. But departures from the ideal siting can apparently be found at all stations.

Dr Jolliffe also says that the lower cross-correlations between Rosslare and other stations could be a by-product of lower autocorrelations at Rosslare. We find it difficult to see dramatic differences between the autocorrelations at Rosslare and other stations from Fig. 4 and Fig. 5; in particular, the pattern at Rosslare is similar to those at Roche's Point and Valentia. Roche's Point provides an informal test of Dr Jolliffe's hypothesis. The cross-correlation between Rosslare and Roche's Point is about one-quarter less than would be predicted from equation (3.3). Inspection of Fig. 4 indicates that the short term

autocorrelation structures at both stations are well approximated by AR(1) models, while Fig. 5 shows that the long-range dependence patterns are also similar. Dr Jolliffe's own calculation yields $K = 0.9994$, so that differences in autocorrelations are unlikely to explain the difference in cross-correlation.

Dr Jolliffe also asks whether we extended the cross-validation exercise to predict the values for Rosslare. Some cross-validation on Rosslare was performed. Using 52 weeks of data at Rosslare, a 5-year mean wind was predicted by $\tilde{\mu}_k$ with an error of 1.0%; this error ranked fifth smallest of the 12. For a longer 18-year mean the error was 1.5% which ranked second out of 12. This is perhaps another example of the remarkably ($p < 0.10$) good fortune pointed out by Dr Glaseby!

Spatial covariance structure

To respond briefly to Dr Chatfield and Dr Yar, kriging can be viewed as a minimum mean square error interpolator or predictor in a stochastic process context. Cressie (1985) reminds us that it has been reinvented many times and is similar, for example, to the well-known Wiener filter. Professor Mardia's contribution shows yet another familiar face of the technique. It is more frequently applied in spatial problems with no time replication. A key step in kriging is the estimation of the relationship between (spatial) correlation and distance. In our case, as Professor Tong points out, we model the (temporal) cross-correlation of the ε_{it}s as a function of distance. In this sense the method can be thought of as a multiple time series model, as Professor Stein remarks.

We must disappoint Professor Titterington and Mr Jamieson: we have declined to interpolate the mean wind speed from other means, for with only 12 data points and the expectation on physical grounds of spatial non-stationarity of the mean, this would be foolhardy. Nevertheless the similarities between the difficulties arising in the two problems are important, and many contributors have drawn attention to the fact that we have available here (as typically in geostatistics) very little evidence to guide us at short-range separation. As Professor Cressie and Professor Pesarin point out, we did have some additional data. These could have been used to adjudicate between the suggestions by Professor Conradsen, Professor Stein and others that the nugget effect is greater than we estimated, and that of Dr Li that it be ignored altogether. In retrospect, these data, which were omitted on meteorological advice owing to the length of record in one case and anomalous data in the other, could have proved useful here. As Professor Conradsen points out, a change in the variogram structure can have dramatic effects on the kriging weights. What is at issue here, however, is the variance of the difference between an optimal and a suboptimal estimator, based on a correct and an incorrect variogram respectively. This is not as dramatic; see comment (h) by Professors Cressie and Pesarin. The correct estimation of the 'kriging variance' does depend critically on the variogram, as remarked by Professor Mardia.

Professor Smith, Professor Lewis, Professor Guttorp and Professor Sampson and Professor Switzer are all concerned that equation (3.3) is not sufficiently general. Our numerical work indicated that, for our purpose, precision is not greatly improved even by assuming knowledge of the exact spatial covariance structure at the new site, which is presumably the best that we can do. Thus equation (3.3) appears to be sufficiently general for our application. However, in view of other potential applications, it is worth considering generalizations, especially as doing so is unlikely to complicate the statistical analysis greatly.

One such is suggested by Professor Smith, who asks whether equation (3.3) could be generalized to allow for directional dependence. This is not too difficult. If ϕ_{ij} is the angle of the line joining stations i and j, restricted to the range $[\phi_0, \phi_0 + \pi]$ for some ϕ_0, then we can replace the lower equation in (3.3) by

$$r_{ij} = \exp[-g(d_{ij}, \phi_{ij})].\tag{2}$$

A simplification of equation (2) which may often be reasonable is to set

$$g(d, \phi) = g_1(d) + g_2(\phi).\tag{3}$$

In equation (3) we may specify functional forms such as $g_1(d) = \alpha + \beta d$ and $g_2(\phi) = \exp\{\kappa[\cos(\phi - \bar{\phi}) - 1]\}$, suggested by the von Mises distribution. This could represent a situation in which correlation is strongest along a direction $\bar{\phi}$, and declines as we deviate from that direction. For the wind data, however, generalizations such as equation (2) do not seem necessary.

Professor Lewis finds the lack of directional information counter-intuitive. We were disappointed also. To give flavour to this, some observed correlations are shown in Table 3. For simplicity we confine attention to Belmullet, and its correlation with other stations, in 1970.

TABLE 3

Correlations between wind speeds at Belmullet, and other stations in 1970, based on (transformed) daily averages, and daily averages of (signed) east–west, north–south components

Station	Square root	East–west	North–south
Roche's Point	0.57	0.02	0.21
Valentia	0.70	0.33	0.29
Rosslare	0.35	0.16	0.18
Kilkenny	0.65	0.31	0.09
Shannon	0.75	0.15	0.12
Birr	0.78	0.00	− 0.04
Dublin	0.70	0.13	− 0.06
Claremorris	0.87	0.29	0.30
Mullingar	0.77	0.08	0.06
Clones	0.80	0.34	0.25
Malin Head	0.89	0.39	0.23

We speculate that aggregation is the source of the difficulty, and that more detailed modelling at the level of hours would be needed to exploit this directional information properly. This would probably need greater attention to lagged correlations reflecting the weather systems, as suggested by Dr Henstridge, and, if we understand Professor Mardia's final point correctly, to cross-covariances between components. We feel that this would contribute little extra to the final result.

Professor Guttorp and Professor Sampson outline a nonparametric method for estimating non-stationary and anisotropic spatial covariances. This looks promising and reveals subtle but potentially important features of the wind data which could not easily be detected otherwise. It also accommodates Rosslare in a smooth way and provides estimates of the spatial covariance at all locations. A remaining question is whether the estimated covariance structure is guaranteed to be positive definite.

Professor Switzer's alternative proposal is interesting because it provides a way of modifying the assumed global spatial covariance to take account of local structure. However, it is designed for the situation where no data are available at the new site, which was not the case for us. Also, it is not guaranteed to yield a positive definite spatial covariance matrix. Ideally, such a proposal should give weight to the data at a new site that increases with its amount. Devising a scheme which weights data at a new site appropriately while preserving positive definiteness is a difficult challenge.

Dr Taam and Professor Yandell suggest setting the problem in a Bayesian context of multivariate smoothing splines. This is an interesting idea, although the problems of implementation are formidable, and we look forward to more research on this topic. Their more specific proposals for the situation where the data are on a lattice are also interesting, although they do not seem directly relevant to the present problem.

Why long-memory?

Meteorologists have long been aware that the sample mean may exhibit behaviour that is inconsistent with short-memory dependence, which they often call 'potential predictability' (Madden, 1976; Shukla and Gutzler, 1983; Trenberth, 1985). However, as Dr Glasbey and Dr Katz point out, they have tended to attribute such behaviour to the rather vaguely defined concept of 'climatic drift', which they clearly think of as a form of non-stationarity. By contrast, in the closely related area of hydrology, similar phenomena are often observed, and long-memory dependence is widely accepted as an explanation for them.

We continue to believe that wind speeds in Ireland probably do exhibit long-memory dependence. The decrease in the empirical mean squared errors in Table 1 is too rapid to be compatible with most reasonable models for non-stationarity in the mean. Further, certain kinds of behaviour often described as climatic drift can be represented by long-memory processes. Dr Glasbey reports the meteorologists' rule-of-thumb that climatic drift manifests itself in periods greater than 30 years. For a fractionally differenced model with our estimated $d = 0.328$, the variance of a 30-year mean is about the same as

that of the mean of 25 independent daily observations! Thus our model implies that disjoint 30-year periods may have quite different means, giving the appearance of climatic drift.

Professor Dempster points out that Fig. 5 does not conclusively establish that the data have a long-memory component, rather than, say, cycles of lengths close to the 11- and 22-year sunspot cycles. In support of the long-memory hypothesis, we can only point to the empirical behaviour of the sample means in Table 1, the lack of apparent cycles or monotonic trends in plots of long series of annual means (up to 40 years) such as those in Raftery *et al.* (1982) and the analogy with hydrology. Professor Dempster also says that the AR(9) filter is capable of representing something indistinguishable from long-memory dependence via roots near unity. However, an autoregressive root near unity cannot account for behaviour of the kind that we observed, such as the behaviour of the sample means, which is characteristic of long-memory dependence, but quite different from non-stationarity.

Dr Chatfield and Dr Yar ask whether some of the long-memory dependence could be explained by the imperfect nature of the seasonal filter, also pointed out by Professor Ogata. Fig. 5 shows that this cannot be so. At each station there is a local peak in the periodogram around the annual frequency resulting from the failure to remove all the seasonal variation, but this is well separated from the low frequency ordinates which reveal the long-memory dependence.

Dr Henstridge suggests that some of the long-memory effect may be due to changes in measuring equipment and in the environment around the stations, and perhaps even to displacements of the stations themselves. Apparently the measuring equipment has not been changed, except at Malin Head, where the anemometer was raised in about 1965; an empirical adjustment (similar to that suggested by Mr Brontë-Hearne) was made here to preserve continuity. Urban spread has latterly reached some of the stations, originally placed 2–3 miles from the towns, but it seems that during the period 1961–78 this was not regarded as a problem.

Why fractional differencing?

Several discussants suggested ways of modelling the observed long term dependence other than fractional differencing. Dr Chatfield and Dr Yar wonder why we did not use first differencing. The reason is that this yields a non-stationary model of the random walk type, which would conflict with the behaviour of the sample means in Table 1.

Dr Jones suggested a medium-memory model. This is interesting, although the three-parameter model written down is formally a short-memory model, and the behaviour of the sample mean would reflect this. Thus it seems unlikely that such a model could adequately account for the empirical mean squared errors in Table 1. However, the idea of defining the model in terms of the partial autocorrelations is valuable; in this connection we draw attention to the pioneering paper of Ramsey (1974), which is often overlooked.

Professor Tong suggests another model, but he is not sure whether or not it has the long-memory property. It is appealingly simple, and so we hope that he, Professor Künsch and Dr Tjøstheim continue this research.

Dr Beran points out that long-range dependence may exist in space as well as in time, and Dr Renshaw has made a start on modelling it.

Model elaboration

Professor Smith, Dr Chatfield and Dr Yar, Professor Cressie and Professor Pesarin and Dr Li are concerned that forcing the autoregressive moving average (ARMA) coefficients to be constant across sites in model (4.1) may be unduly restrictive, while Dr Jolliffe, Professor Tong and (implicitly) Dr Henstridge suggest allowing direct dependence of X_{it} on X_j^{t-1} for $j \neq i$. Professor Guttorp and Professor Sampson suggest allowing a gradient in variance across Ireland. All these suggestions lead to special cases of model (1). We chose model (4.1) after experimenting with other special cases of model (1) because it was the simplest model which enabled us to achieve our objective, not because it captures every feature of the synoptic data.

On the basis of Fig. 4, Professor Cressie and Professor Pesarin comment that Valentia, Roche's Point and Rosslare do not have the same long-range dependence as the other stations. Detailed features of empirical autocorrelation functions such as those in Fig. 4 are notoriously difficult to interpret, and we preferred to rely on Fig. 5 which indicates that the low frequency characteristics at these three stations are similar to those at the others. Dr Henstridge expects some time delay of up to 12 hours between the west coast and east coast stations; our exploratory analyses, some of which are described in Raftery *et al.* (1982), showed this not to be important at the daily level of aggregation. Professor Guttorp and Professor

Sampson detect a gradient in variance over Ireland; we agree that this is present, but it is slight and has little effect on the performance of the estimators (3.4) and (4.10).

In answer to Dr Bhansali, model (4.1) is not a special case of the standard multivariate ARMA (MARMA) model as defined, for example, by Tiao and Box (1981), because the MARMA model does not allow for long-memory dependence. The standard MARMA model is a special case of model (1), however. We did some MARMA modelling at an early stage of our project (Raftery et al., 1982), but this was not very satisfactory in terms of our main goal. The diagnostic checks that we used are summarized in Section 4.4.

Professor Tong points out that $E[X_{it}|X_{js}]$ could be non-linear. We found no evidence of this in our data, but it may be true in other situations, and model (1) could easily be modified to take account of it.

Estimating d

The discussion opposes two views about the estimation of d. Our approach, which also underlies the discussions of Dr Carlin, Professor Künsch and Dr McLeod, is the traditional approach of the exact or approximate maximum likelihood estimator (MLE). However, Professor Dempster and Professor Smith point out that this amounts to using the fractional differencing term to shape spectra across the full frequency range, whereas a different value of d could be operating at the lowest frequencies. This leads to methods of estimating d based only on the lowest periodogram ordinates, such as those of Janacek (1982) and Geweke and Porter-Hudak (1983). Professor Smith suggests an ingenious way of making theoretical progress on the hitherto elusive properties of such methods by exploiting the analogy with the estimation of the tail of a probability distribution.

Li and McLeod (1986) and Hosking (1984a) report simulation results that MLE-type estimators perform much better than low-frequency-based estimators. This is valid only if the model fits reasonably well (and then is almost tautological), which is not the case for our data. We conjecture that the ARMA terms in the model determine most of the medium and higher frequency behaviour, leaving only the low frequency behaviour to be determined by the fractional differencing term. If this is true, the problem with MLE-type estimators which concerns Professor Dempster and Professor Smith is less serious.

In clarification of remarks by Dr Chatfield and Dr Yar and Professor Dempster, we did not use the AR(9) residuals to estimate d, and indeed we would not want to, for much the same reasons as Professor Dempster's. Fig. 5 is used only for the exploratory purpose of revealing the presence of long-memory dependence. This may explain the discrepancies noted by Dr Walden, whose remarks could lead to low-frequency-based estimators of d as high as $d = 2$, compared with the approximate MLE $d = 0.328$. Values such as $d = 2$ are incompatible with the behaviour of the sample means in Table 1.

Dr Carlin would welcome further justification and evaluation of our approximation to the log-likelihood. Our investigations were encouraging, although of necessity somewhat limited. For example, for simulated univariate ARIMA(0, d, 0) series of length 1000, we found that with $M = 100$ the difference between our approximate log-likelihood and the exact log-likelihood was generally less than the average contribution of a single observation. We intend to pursue these investigations, and we hope that others do likewise. In answer to Dr Carlin, we used a quasi-Newton optimization method without derivatives, with starting values found as in Section 4.2.

Professor Künsch's derivation of Whittle's approximation to the log-likelihood for the model (4.1) is a significant contribution. It is not clear that the Whittle approximation requires much less central processor unit time than the approximation that we used, but we look forward to further investigation and comparison of the two approximations.

Asymptotics

In Section 4.3 we said, 'Neither the finite-sample nor the asymptotic distribution of the maximum likelihood estimator for models such as (4.1) appears to be known'. Dr McLeod contests this, citing Li and McLeod (1986). However, their theorem applies only to the univariate case, and then only when the mean is known. It is thus far from yielding the distribution of the MLE for model (4.1), for which there may also be problems with the nugget parameter α, as pointed out by Professor Mardia. There is a further difficulty with the method of Li and McLeod (1986). They study the univariate model

$$\phi(B)\nabla^d(X_t - \mu) = \theta(B)\varepsilon_t, \tag{4}$$

saying that $\{X_t\}$ has mean μ. However, it does not follow from equation (4) that $\{X_t\}$ has mean μ, since $\nabla^d\mu = 0$; indeed, equation (4) does not specify any mean for $\{X_t\}$. This is why it is important to put the del operator on the right-hand side of equations such as equation (4), as in model (4.1) and equation (1).
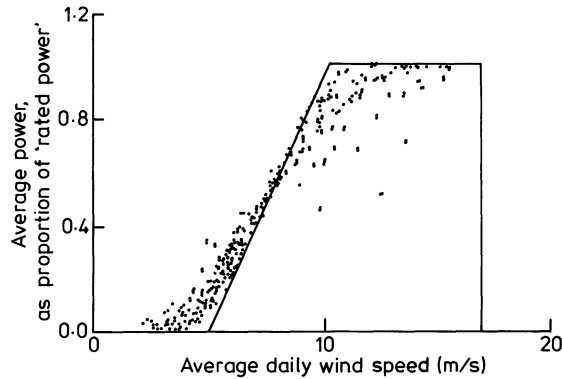
Fig. 15. Wind power generated from a given turbine, as a function of observed daily average wind speed (there is no power below 5 m/s or above 17 m/s; the data are for one year only at Belmullet): ——, power–velocity curve for the turbine

We thank Professor Stein for his authoritative comments. The only sensible asymptotics in our problem refer to N large, and, as a practical matter, we accept the inapplicability of Mardia and Marshall (1984).

Estimating μ_k

Dr Beran points out that our expression (4.10) for $\mathrm{var}(\tilde{\mu}_k)$ does not take into account the fact that d, $\phi(B)$ and $\theta(B)$ have to be estimated; this also applies to α, β and σ_X^2. However, the standard errors for these parameters appear to be small, and so it is unlikely that taking them into account would increase $\mathrm{var}(\tilde{\mu}_k)$ by much. Professor Cressie and Professor Pesarin point out that a similar comment applies to the seasonal component; we suspect that the effect of this is also small.

A more important source of variability, which we did not take into account either, is the fact that μ_i ($i \neq k$) are estimated. Because of the long-memory property, these estimates are somewhat imprecise, even with 18 years of data. Our cross-validation study was conditional on these estimates. Professor Künsch's modified estimator of μ_k and its variance do take account of this and are thus more realistic than our proposals. We suspect that the difference is slight in our application, but it may be important in other contexts.

Estimating wind power

Section 5 of the paper is rather more empirical than we would like. In particular, extrema, while critically important to the survival of the machine, as Professor Titterington and Mr Jamieson remark, are less important for power production, as Dr Lippman and Professor Mollison point out. Not only will our method overestimate the machine-specific power production, if used unthinkingly, but it is probably unnecessarily pessimistic on the question of precision. Fig. 15 helps to demonstrate Dr Lippman's point for a specific turbine, and may be contrasted with Fig. 6. The power–velocity curve relates *instantaneous* wind speed to power and shows that the machine shuts down in high winds, for safety. We apologize to Dr Glaseby for his difficulties with Fig. 6; we seem to have added a little too much scatter in preparing this diagram.

It is right that Professor Mollison should remind us that there are other approaches to this problem. He mentions two: his own interesting proposal, and the meteorologically based 'hindcasting' approach of Golding. We wonder how his nonparametric model could be extended to a multivarate study, with wind data at more than one site. This may be less important in studies of wave energy.

A further alternative, under development for some time at Risø, in Denmark (Peterson *et al.*, 1988), is based on an expert evaluation of the site in question, with regard to terrain in different directions and other similar matters. It refers not only to the hourly wind data at a local synoptic station (defined by the World Meteorological Organization as a station satisfying certain exposure criteria, at which a variety of weather data are collected at least as often as every 3 hours) but also to the 'effective geostrophic wind' at the top of the boundary layer. The method yields estimates of mean wind energy, and of the distribution of wind speeds, at the chosen site, *in advance of any data at that site*. As such it

provides a good example of the *a priori* information that we and Dr Scott feel to be so important. It does not yield explicit estimates of *a priori* precision, but very recent information provided by Liam Burke suggests that a precision of $\pm 20\%$ for mean kinetic energy has been achieved in tests at well-exposed sites in Ireland. Since this can then be complemented by new data at the site, adjusted in a manner such as we have proposed, accuracy sufficient to satisfy Professor Mollison is not impossible.

Miscellaneous comments

Professor Künsch and Dr Katz both cast doubt on our recommendation that wind speed data be collected at a much denser grid of locations, perhaps using simple anemometers attached to existing electricity and telephone poles. Dr Katz's reservations are based on the debate between long memory and non-stationarity, on which we have already commented. Professor Künsch rightly points out that such information will be useful only if the records are much longer than at the site of interest; our recommendation is that they be collected permanently, if perhaps infrequently, as a supplement to the synoptic data. The question of optimally siting such new locations, or wind farms, remains, as Dr Scott points out, an open and difficult question.

Dr Chatfield and Dr Yar take us to task for not smoothing the periodograms in Fig. 5. Interest there focuses on a small number of low frequency ordinates and on the narrow peak at the annual frequency, and we felt that smoothing would obscure rather than highlight these features, which are already clear from the raw periodograms.

Professor Cressie and Professor Pesarin ask whether the data are available for reanalysis. They may be obtained by sending electronic mail to Adrian Raftery at raftery@entropy.ms.washington.edu or raftery%entropy.ms@beaver.cs.washington.edu; they occupy about half a megabyte of storage.

We are grateful to Julian Besag, Liam Burke, Michael Newton, Paul Sampson and Richard Smith for helpful discussions during the preparation of this reply.

References in the Discussion

Akaike, H. (1979) Likelihood and Bayes procedure. In *Bayesian Statistics* (eds J. M. Bernard *et al.*). Valencia: University Press.
Bartlett, M. S. (1971) Physical nearest-neighbour models and non-linear time-series. *J. Appl. Prob.*, **8**, 222–232.
——(1975) *The Statistical Analysis of Spatial Pattern*. London: Chapman and Hall.
Bates, D. M., Lindstrom, M. J., Wahba, G. and Yandell, B. S. (1987) GCVPACK—routines for generalized cross validation. *Commun. Statist.* B, **16**, 263–297.
Beran, J. (1986) Estimation, testing, and prediction for self-similar and related processes. Unpublished doctoral dissertation, ETH, Zurich.
Besag, J. E. (1972) Nearest-neighbor systems and the auto-logistic binary data. *J. R. Statist. Soc.* B, **34**, 75–83.
Bhattacharya, R. N., Gupta, V. K. and Waymire, E. C. (1983) The Hurst effect under trends. *J. Appl. Prob.*, **20**, 649–662.
Bookstein, F. L. (1978) The measurement of biological shape and shape change. *Lect. Notes Biomath.*, **24**.
British Wind Energy Association (1987) *Wind Power for the UK*. London: British Wind Energy Association.
Camacho, F., McLeod, A. I. and Hipel, K. W. (1987a) Contemporaneous bivariate time series. *Biometrika*, **74**, 103–113.
——(1987b). Multivariate contemporaneous ARMA models with hydrological applications. *Stochast. Hydrol. Applicn*, **1**, 141–154.
Carlin, J. B. (1987) Seasonal analysis of economic time series. Unpublished PhD thesis, Department of Statistics, Harvard University.
Carlin, J. B. and Dempster, A. P. (1989) Sensitivity analysis of seasonal adjustments: empirical case studies. *J. Amer. Statist. Ass.*, to be published.
Carlin, J. B., Dempster, A. P. and Jonas, A. B. (1985) On models and methods for Bayesian time series analysis. *J. Economet.*, **30**, 67–90.
Carlin, J. B. and Haslett, J. (1982) The probability distribution of wind power from a dispersed array of wind turbine generators. *J. Appl. Meteorol.*, **21**, 303–313.
Carter, D. J. T. and Draper, L. (1988) Has the north-east Atlantic become rougher? *Nature*, **332**, 494.
Chan, W. Y. T. and Wallis, K. F. (1978) Multiple time series modelling: another look at the mink-muskrat interaction. *Appl. Statist.*, **27**, 168–175.
Cox, D. D., Koh, E., Wahba, G. and Yandell, B. S. (1988) Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.*, **16**, 113–119.
Cox, D. R. (1984) Long-range dependence: a review. In *Statistics: an Appraisal* (eds H. A. David and H. T. David), pp. 55–74. Iowa State University Press.
Cressie, N. (1985) Fitting variogram models by least squares. *Math. Geol.*, **17**, 563–586.

Deely, J. J. and Lindley, D. V. (1981) Bayes empirical Bayes. *J. Amer. Statist. Ass.*, **76**, 833–841.

Dowd, P. A. (1982) Lognormal kriging—the general case. *J. Math. Geol.*, **14**, 475–499.

Dubey, S. D. (1967) Normal and Weibull distributions. *Nav. Logist. Res. Q.*, **14**, 69–79.

Fox, R. and Taqqu, M. S. (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, **14**, 517–532.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.*, **4**, 221–238.

Golding, B. (1980) Computer calculations of waves from wind fields. In *Power from Sea Waves* (ed. B. Count), pp. 115–134. London: Academic Press.

Halford, D. (1968) A general mechanical model for $|f|^\alpha$ spectral density random noise with special reference to flicker noise $1/|f|$. *Proc. IEEE*, **56**, 251–258.

Hall, P. (1982) On some simple estimates of an exponent of regular variation. *J. R. Statist. Soc.* B, **44**, 37–42.

Hall, P. and Welsh, A. H. (1984) Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.*, **12**, 1079–1084.

——(1985) Adaptive estimates of parameters of regular variation. *Ann. Statist.*, **13**, 331–341.

Halliday, J. A. (1983) Wind speed statistics of 14 dispersed U.K. Met Stations with special regard to wind energy. *Report RL-83-124*. Didcot: Science and Engineering Research Council.

Hannan, E. J. and Thompson, P. J. (1971) The estimation of coherence and group delay. *Biometrika*, **58**, 469–481.

Haslett, J. and Kelledy, E. (1979) The assessment of actual wind power availability in Ireland. *En. Res.*, **3**, 333–348.

Hosking, J. R. M. (1984) Modeling persistence in hydrological time series using fractional differencing. *Wat. Resour. Res.*, **20**, 1898–1908.

Hui, Y. V. and Li, W. K. (1988) *On Fractionally Differenced Periodic Processes*. Chinese University of Hong Kong and University of Hong Kong.

Hutchinson, M. F. (1984) A summary of some surface fitting and contouring programs for noisy data. *Technical Report ACT84/6*. Division of Mathematics and Statistics, CSIRO.

Janacek, G. J. (1982) Determining the degree of differencing for time series via the log spectrum. *J. Time Ser. Anal.*, **3**, 177–183.

Johnson, N. L. and Kotz, S. (1970) *Continuous Univariate Distributions, Distributions in Statistics*, vols. 2 and 3. Boston: Houghton Mifflin.

Jones, R. H. (1975) Estimating the variance of time averages. *J. Appl. Meteorol.*, **14**, 159–163.

Kramer, W. and Donninger, C. (1987) Spatial autocorrelation among errors and the relative efficiency of ols in the linear regression model. *J. Amer. Statist. Ass.*, **82**, 577–589.

Lawrance, A. J. and Kottegoda, N. T. (1977) Stochastic modeling of river-flow time series. *J. R. Statist. Soc.* A, **140**, 1–47.

Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika*, **73**, 217–221.

Loader, C. and Switzer, P. (1988) Non-stationary spatial covariances estimated from monitoring data. *Technical Report 115*. Stanford University.

Madden, R. A. (1976) Estimates of the natural variability of time-averaged sea level pressure. *Month. Weath. Rev.*, **104**, 942–952.

Mallows, C. L. (1980) Some theory of nonlinear smoothers. *Ann. Statist.*, **8**, 694–715.

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

McWilliams, B., Newmann, M. M. and Sprevak, D. (1979) The probability distribution of wind speed and direction. *Wind Engng*, **3–4**, 269–273.

Mollison, D. (1980) The prediction of device performance. In *Power from Sea Waves* (ed. B. Count), pp. 135–172. London: Academic Press.

——(1986) Wave climate and the wave power resource. In *Proc. IUTAM Symp. Hydrodynamics of Ocean Wave-energy Utilisation* (eds D. V. Evans and A. F. de O. Falcao), pp. 135–156. Heidelberg: Springer.

Montroll, E. W. and Schlesinger, M. F. (1982) On $1/f$ noise and other distributions with long tails. *Proc. Natl Acad. Sci. USA*, **79**, 3380–3383.

Morris, C. N. (1983) Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Ass.*, **78**, 47–65.

Ogata, Y. (1988) Comment on Space-time modelling with long-memory dependence: assessing Ireland's wind power resources (by J. Haslett and A. Raftery). *Res. Mem.*, No. 351, 1–5.

Ogata, Y. and Katsura, K. (1988) Likelihood analysis of spatial inhomogeneity for marked point patterns, *Ann. Inst. Statist. Math.*, **40**, 29–40.

Parzen, E. (1983) Time series model identification by estimating information, memory and quantiles. *Technical Report*. Department of Statistics, Texas A&M University.

Petersen, E. L., Troen, I. and Mortensen, N. G. (1988) The European wind energy resources. *Proc. 1988 European Community Wind Energy Conf.* To be published.

Raftery, A. E., Haslett, J. and McColl, E. (1982) Wind power: a space–time process? In *Time Series Analysis: Theory and Practice 2* (ed. O. D. Anderson), pp. 191–202.

Ramsey, F. L. (1974) Characterization of the partial autocorrelation function. *Ann. Statist.*, **2**, 1296–1303.

Reinsch, C. H. (1967) Smoothing by spline functions. *Numer. Math.*, **10**, 177–183.

Renshaw, E. (1984) Competition experiments for light in a plant monoculture: an analysis based on two-dimensional spectra. *Biometrics*, **40**, 717–728.

Sampson, P. D. (1986) Spatial covariance estimation by scaled-metric scaling and biorthogonal grids. *SIMS Technical Report 102*. University of British Columbia.

Shukla, J. and Gutzler, D. (1983) Interannual variability and predictability of 500 mb geopotential heights over the Northern Hemisphere. *Month. Weath. Rev.*, **111**, 1273–1279.

Stein, M. L. (1987) Minimum norm quadratic estimation of spatial variograms. *J. Amer. Statist. Ass.*, **82**, 765–772.

——(1988) Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.*, **16**, 55–63.

Taam, W. (1988) A semi-parametric approach to spatially correlated data. Unpublished PhD Thesis, University of Wisconsin–Madison.

Tiao, G. C. and Box, G. E. P. (1981) Modeling multiple time series with applications. *J. Amer. Statist. Ass.*, **76**, 802–816.

Tjøstheim, D. (1978) Statistical spatial series modelling. *Adv. Appl. Prob.*, **10**, 130–154.

Trenberth, K. E. (1985) Potential predictability of geopotential heights over the Southern Hemisphere. *Month. Weath. Rev.*, **113**, 54–64.

Wahba, G. (1983) Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. R. Statist. Soc.* B, **45**, 133–150.

——(1985) A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.*, **13**, 1378–1402.

Watkins, A. J. (1988) Some aspects of statistical inference in the spatial linear model. PhD Thesis, Department of Statistic, University of Leeds.

Whittle, P. (1953) The analysis of multiple stationary time series. *J. R. Statist. Soc.* B, **15**, 125–139.

Wigley, T. M. L. and Jones, P. D. (1981) Detecting CO_2-induced climatic change. *Nature*, **292**, 205–208.

Yandell, B. S. (1988) Block diagonal smoothing splines. *Statist. Prob. Lett.*, **6**, in the press.