

# A spectral method for estimating parameters in rainfall models

RICHARD E. CHANDLER

*Department of Statistical Science, University College, Gower Street, LONDON WC1E 6BT, UK.  
e-mail: richard@stats.ucl.ac.uk*

One of the most difficult problems in rainfall modelling is often the fitting of theoretical models to data. In this paper a spectral approach is proposed for fitting single-site models, by considering the approximate likelihood functions of collections of sample Fourier coefficients. An objective function is derived, which is the same as that used in Whittle's method for time series parameter estimation, and is shown to have an interpretation as a quasi-likelihood when the time series is non-Gaussian. The method requires knowledge of the theoretical spectral density of models to be fitted; the form of this spectral density is given for a wide class of point-process-based rainfall models. A variant of the method is also developed for when a number of independent replications of a rainfall process are available. Large-sample properties of the estimators are derived, and the method is illustrated with some data from the south-west of England.

*Keywords:* periodogram; point process; rainfall modelling; spectral analysis; Whittle criterion

## 1. Introduction

In recent years, much work has been done on building stochastic, and physically intuitive, models for rainfall at a single site; however, methods for fitting such models to data remain fairly primitive. The most widely used approach is based on 'method-of-moments' type fitting as, for example, in Rodriguez-Iturbe *et al.* (1987; 1988); however, this suffers from the disadvantage that the parameter estimates can vary greatly depending on the properties used in the fitting procedure.

The reason for the difficulty in fitting these rainfall models lies in the fact that their formulation does not lend itself easily to maximum likelihood estimation, because of the physically appealing but mathematically intractable dependencies involved between the rainfall intensities at different time points. Our solution is to work with the Fourier transform of the data, rather than the data themselves; because collections of the sample Fourier coefficients are asymptotically independent and normally distributed, we can derive an approximate likelihood function in terms of these coefficients. The method is well known in time series, where it is referred to as *Whittle's method* – see Dzhaparidze and Yaglom (1983), for example. A particular advantage of the technique is its efficient use of data (compared with, say, a 'method-of-moments' approach); for by using first- and second-order

properties of all the Fourier coefficients, we are effectively incorporating all first- and second-order properties of the data into the estimation procedure.

In Section 2 of this paper, we review the necessary theory and derive some approximate likelihood functions for estimating parameters; Section 3 gives the properties of the resulting estimators. In Section 4 the technique is illustrated on data from the south-west of England.

## 2. Theory

Let  $Y_t$  denote a stochastic process of mean rainfall intensities, recorded at a site as averages over intervals of length  $\Delta$  time units. We assume:

- (1) that  $Y_t$  is strictly stationary, with mean  $\mu_Y$ ;
- (2) that for  $k \geq 2$ , the  $k$ th-order cumulant function,  $c_k(r_1, \dots, r_{k-1})$ , of  $Y_t$  satisfies

$$\sum_{r_1=-\infty}^{\infty} \dots \sum_{r_{k-1}=-\infty}^{\infty} |c_k(r_1, \dots, r_{k-1})| < \infty. \quad (1)$$

This requires that the process  $Y_t$  has short-range rather than long-range dependence. Note that in (1), the arguments  $\{r_j\}$  are integer-valued lags; an integer lag  $r$  corresponds to a lag of  $r\Delta$  underlying time units.

### 2.1. Basic spectral theory

A particular consequence of (1) is that the second-order spectral density of the process exists and is bounded; it is defined by

$$h_Y(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} c_2(r) e^{-i\omega r \Delta} \quad \omega \in (-\pi/\Delta, \pi/\Delta), \quad (2)$$

where  $c_2(\cdot)$  is the autocovariance function of the series  $Y_t$ . Using the results from Priestley (1981, equations (4.12.24) and (7.1.12)), it is possible to show that

$$h_Y(\omega) = \sum_{k=-\infty}^{\infty} \left[ g \left\{ \omega + \frac{2k\pi}{\Delta} \right\} \times \left\{ \frac{\sin((\omega\Delta + 2k\pi)/2)}{[\omega\Delta + 2k\pi]/2} \right\}^2 \right] \quad (|\omega| \leq \pi/\Delta), \quad (3)$$

where  $g(\omega)$  is the theoretical spectral density of the underlying continuous-time rainfall intensity process from which  $Y_t$  is derived. This result follows from the fact that the discretization into bins of length  $\Delta$  is equivalent to first applying a uniform filter of width  $\Delta$  to the original process, then sampling at intervals of length  $\Delta$  (it is important to remember that  $Y_t$  is a process of rainfall *intensities* rather than depths). In practice, it is necessary to approximate the infinite sum in (3) by a finite one; however, this is not a problem as the sum is dominated by the terms for small  $k$ . A discussion of the choice  $k$  in practice is deferred to Section 4.1.

2.1.1. Form of  $g(\cdot)$  for a class of point process rainfall models.

A wide class of stochastic rainfall models is based on point processes – see, for example, Rodriguez-Iturbe *et al.* (1987; 1988) and Smith and Karr (1983). For this class we have been able to derive the most general form of the spectral density  $g(\cdot)$ ; before giving the result, however, it is necessary to establish some notation. The basic model runs as follows:

(1) The rainfall intensity at any time is made up of a superposition of ‘rain cells’, each attached to an event of a stationary, orderly point process  $D$  which we shall call the *driving process*.  $D$  has rate  $\lambda$ , and incomplete second-order spectral density  $g_D^*(\cdot)$ . Note that this excludes the usual  $\lambda/2\pi$  term, which is the contribution from a delta-function component at the origin of the covariance density. For a discussion of the spectral theory of point processes, see Bartlett (1963).

(2) Each cell is rectangular with random ‘depth’,  $X$ , representing rainfall intensity, and random duration,  $L$ .  $X$  and  $L$  are independent, and are realized independently and identically for each cell. The mean and variance of  $X$  are  $\mu_X$  and  $\sigma_X^2$  respectively, and the characteristic function of  $L$  is  $\phi_L(\cdot)$ .

This class of models is clearly stationary because  $D$  is, and because  $X$  and  $L$  are realized identically for every event of  $D$ . It may be shown that the spectral density of the continuous-time rainfall intensity process is given, for  $\omega \neq 0$ , by

$$g(\omega) = \frac{1}{2\pi\omega^2} [2\pi g_D^*(\omega)\mu_X^2 |1 - \phi_L(\omega)|^2 + 2\lambda\{\mu_X^2 + \sigma_X^2\}\{1 - \Re(\phi_L(\omega))\}] \quad (4)$$

which can be substituted into a truncated version of (3) to obtain the theoretical spectral density of the aggregated process. When  $\omega = 0$ , it is necessary to apply a limiting operation to (3) and (4) to obtain the required result. The theoretical spectral densities  $g_D^*$  for some of the most commonly used driving point processes are listed in the Appendix to this paper.

Note that (4) is applicable only to point-process rainfall models where cell depth and duration are realized independently and identically for each cell. The ‘random  $\eta$ ’ model of Rodriguez-Iturbe *et al.* (1988), for example, does not fall into this category, for the cell durations are not independent within storms in that formulation. Care should therefore be exercised in the application of (4).

## 2.2. The sample Fourier coefficients

Given data  $Y_0, \dots, Y_{N-1}$  from the process  $Y_t$ , we may compute the sample Fourier coefficients

$$A_p = \frac{2}{N} \sum_{t=0}^{N-1} Y_t \cos(\omega_p t\Delta), \quad B_p = \frac{2}{N} \sum_{t=0}^{N-1} Y_t \sin(\omega_p t\Delta), \quad (5)$$

where

$$\omega_p = \frac{2\pi p}{N\Delta}, \quad p = 0, 1, 2, \dots, \left[ \frac{N}{2} \right]. \quad (6)$$

Because a nonsingular linear transformation of the data is involved, these coefficients contain all the information about a realization of  $Y_t$ , in the sense that we can completely reconstruct the realization from the coefficients. Under the assumptions laid down previously, the coefficients have the further properties (Brillinger and Rosenblatt 1967) that they are, for large  $N$ , approximately normally distributed and mutually independent. Specifically, we have:

$$A_0 \sim N\left(2\mu_Y, \frac{8\pi h_Y(0)}{N}\right); \quad (7)$$

$$A_p \sim N\left(0, \frac{4\pi h_Y(\omega_p)}{N}\right) \quad (p \neq 0); \quad (8)$$

$$B_p \sim N\left(0, \frac{4\pi h_Y(\omega_p)}{N}\right) \quad (\omega_p \neq 0 \text{ or } \pi/\Delta); \quad (9)$$

$$B_p \equiv 0 \quad (\omega_p = 0, \pi/\Delta). \quad (10)$$

From (6), it is clear that  $\omega_p$  can only take the value  $\pi/\Delta$  if  $N$  is even; this point, along with the separate treatment required for both  $A_0$  and  $B_0$ , requires some care. A further point is that asymptotic normality of individual coefficients does not necessarily extend to joint asymptotic normality of large numbers of them.

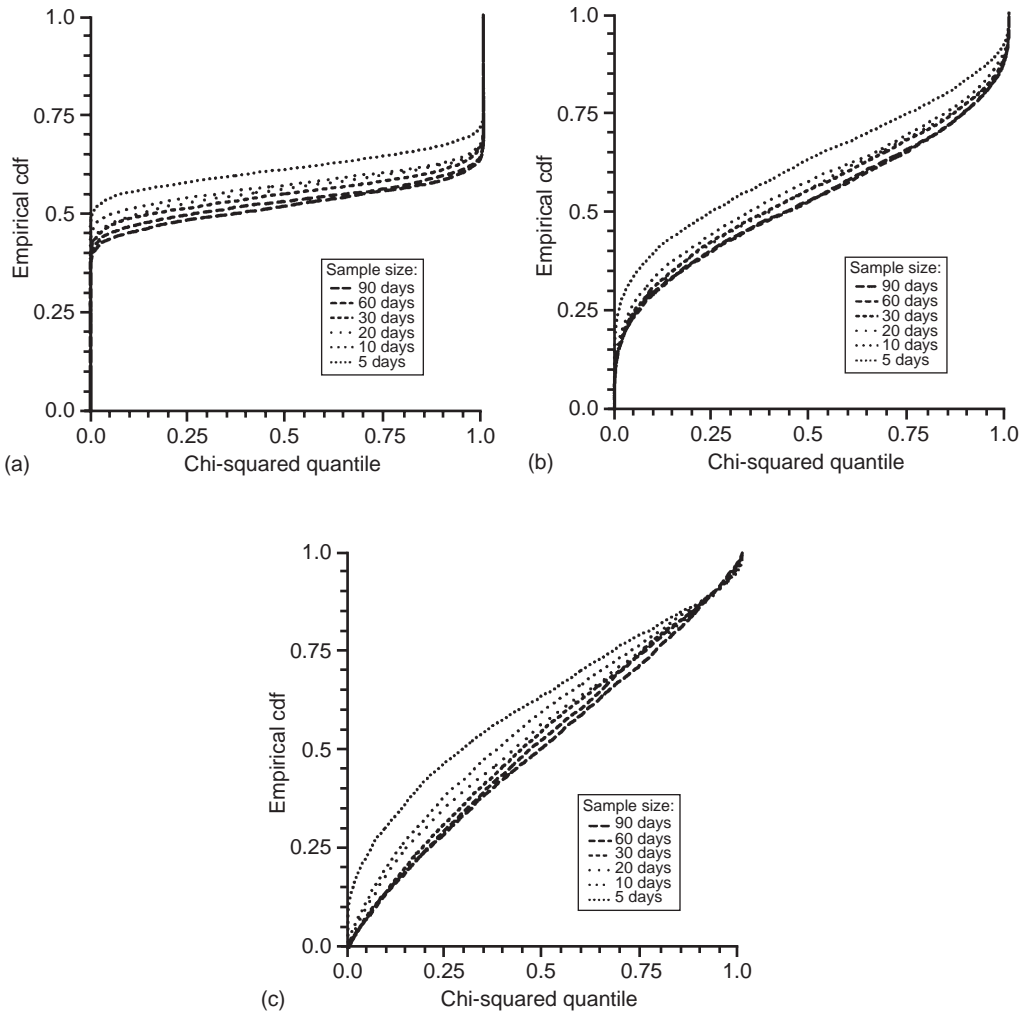
Results (7)–(10) tell us that the *periodogram* at frequency  $\omega_p$ , defined as

$$I(\omega_p) = \frac{N}{2}(A_p^2 + B_p^2), \quad (11)$$

may, when suitably scaled, be used to estimate the spectral density  $h_Y(\omega_p)$ . It is, however, inconsistent (Priestley 1981, Section 6.2) and needs to be smoothed in order to be useful. This point is touched on later.

To ascertain the extent to which the assumption of multivariate normality can be justified, data have been simulated from various different models and simple graphical tests of multivariate normality have been carried out. The test procedure used was a standard one – for details, see Krzanowski (1988), for example. For each model, 5000 realizations were simulated. The sample Fourier coefficients were calculated for each realization, and standardized to have mean 0 and variance 1, according to results (7)–(10). The sum of squares of the standardized coefficients was then computed for a subset of the coefficients, and the empirical distribution function of this sum of squares plotted on chi-squared probability paper with the appropriate degrees of freedom. If the multivariate normality assumption holds, the results should appear as a straight line through the origin, with slope 1. Figure 1 shows some typical results. The model simulated here is of the type described in Section 2.1.1, where the driving point process is a Bartlett–Lewis process with exponential inter-offspring spacing (see the Appendix to this paper for a brief outline); and cell depths and durations are exponentially distributed. The parameter values used here are typical of those found in UK rainfall modelling (as reported in the literature). The data are hourly, and the infinite sum (3) was truncated at  $k = \pm 20$ . The subsets of coefficients were chosen to be equally spaced on the frequency axis.

It is clear from Fig. 1(a) that limiting multivariate normality does not hold for all the Fourier coefficients. When only 5% of the coefficients are used (Fig. 1(b)) the



**Figure 1.** Chi-squared probability plots for testing multivariate normality: (a) using all coefficients; (b) using 5% of coefficients; (c) using 0.5% of coefficients

approximation is still inadequate; and it is only when less than 1% of the coefficients are used that the graphs for the larger sample sizes start to approach straight lines (Fig. 1(c)). Notice that in Fig. 1(c), there is very little to distinguish between sample sizes of 30, 60 and 90 days; and the results for the 20-day sample are quite close to these. A general rule of thumb, from this and other simulations, seems to be that limiting multivariate normality holds typically for collections of less than 1% of the Fourier coefficients; and that the limiting behaviour is established with data sets of more than about 20 days' worth of data

(for the parameter values used here, this corresponds to about ten storms, where a storm is defined by a cluster of cells in the model).

### 2.3. Formulation of a likelihood function

Let us now suppose that we are interested in fitting a model, parametrized by a vector  $\Theta = (\theta_1, \dots, \theta_K)'$ , to the series  $Y_0, \dots, Y_{N-1}$ . From the distributional results obtained in the preceding section, we can formulate an approximate likelihood function for  $\Theta$  in terms of collections of the Fourier coefficients; specifically, the likelihood is of the form  $L(\Theta) = L(A_0)\prod_p L(A_p)L(B_p)$ , where  $L(A_p)$ ,  $L(B_p)$  are normal probability densities. Suppose the Fourier coefficients are evaluated at frequencies  $\{\omega_p: p \in \Pi\}$ , where  $\Pi$  is a subset of the integers  $\{0, 1, \dots, [N/2]\}$  such that the multivariate normal approximation for the coefficients holds. Then the log-likelihood is given by

$$\begin{aligned} \ln L(\Theta) = & - \sum_{p \in \Pi} \left\{ 1 - \frac{1}{2} \delta_{p(N/2)} \right\} \left\{ \frac{I(\omega_p)}{4\pi h_Y(\omega_p; \Theta)} + \ln h_Y(\omega_p; \Theta) \right\} \\ & - \chi_{\Pi}(0) \left\{ \frac{1}{2} \ln h_Y(0; \Theta) + \frac{N(\bar{Y} - \mu_Y(\Theta))^2}{4\pi h_Y(0; \Theta)} \right\} + \text{constant}, \end{aligned} \quad (12)$$

where  $\bar{Y} = A_0/2$  is the sample mean of the realization,  $\chi_{\Pi}(p)$  is an indicator set function taking the value 1 if  $p \in \Pi$  and 0 otherwise, and  $\delta_{ij}$  denotes the Kronecker delta, whose appearance is a consequence of (10).

### 2.4. Inclusion of all frequencies: a quasi-likelihood interpretation

It would be desirable to include all available frequencies in the set  $\Pi$ , so as to make use of all the data and improve the efficiency of the method. On the other hand, (12) would lose its interpretation as a log-likelihood, for it has been seen that the joint normality of the Fourier coefficients can only reasonably be assumed for small collections of them. The fact that (12) involves only first- and second-order properties of the data suggests that it may be possible to interpret it as a quasi-likelihood function (McCullagh and Nelder 1989, Chapter 9). We now show that this is indeed so.

When  $N$  is odd and all frequencies except zero are included in the analysis, the objective function (12) is of the form

$$- \sum_{p=1}^P \left( \frac{I(\omega_p)}{4\pi h_Y(\omega_p; \Theta)} + \ln h_Y(\omega_p; \Theta) \right)$$

(the argument is essentially the same when  $N$  is even and when zero frequency is included, although the expressions involved are more complicated). The estimators resulting from this expression must satisfy  $\Phi(\Theta) = \mathbf{0}$ , where  $\Phi(\Theta)$  is the vector of first derivatives of the objective function which has  $i$ th component

$$\phi^{(i)}(\Theta) = 4\pi \sum_{p=1}^P \frac{\partial h_Y(\omega_p; \Theta)}{\partial \theta_i} (4\pi h_Y(\omega_p; \Theta))^{-2} (I(\omega_p) - 4\pi h_Y(\omega_p; \Theta)). \quad (13)$$

Now, to order  $T^{-1}$ , the periodogram ordinate  $I(\omega_p)$  has mean  $4\pi h_Y(\omega_p)$  and variance  $(4\pi h_Y(\omega_p))^2$ ; and the covariance between distinct periodogram ordinates is zero (Percival and Walden 1993, p. 222). Let  $\mathbf{I}$  denote the vector of periodograms,  $\nu(\Theta)$  and  $\mathbf{V}(\Theta)$  its mean vector and covariance matrix respectively, and  $\mathbf{D}$  the  $P \times k$  matrix of  $\Theta$ -derivatives of  $\nu(\Theta)$ . Then the estimator  $(\Theta)_I$ , say) satisfies

$$\mathbf{D}'\mathbf{V}^{-1}(\Theta)(\mathbf{I} - \nu(\Theta)) = \mathbf{0}. \quad (14)$$

But this expression has exactly the same form as a quasi-likelihood equation (McCullagh and Nelder 1989, p. 333). Thus the proposed estimation procedure using all Fourier coefficients can be regarded as a quasi-likelihood procedure, treating the periodogram ordinates as data. This is important, for the standard toolkit of likelihood-based inference becomes available.

Functions of the form (12) were first introduced by Whittle (1953) as a means of approximating the exact likelihood for Gaussian time-series models; the motivation, as here, was that the spectral density of a model is easy to obtain whereas an exact likelihood is not. The application of Whittle's method has been extensively discussed in the time-series literature, although it is usually assumed that the series is Gaussian. Results for the consistency and asymptotic normality of the Whittle estimators, when the series is not Gaussian, were provided by Rice (1979). Large-sample properties of the estimators are discussed in Section 3 of the present paper.

### 2.5. Improving the computational efficiency

In general, (12) will have to be maximized numerically, which can be time-consuming. The main reason for this is that, when all frequencies are included, the spectral density  $h_Y(\omega_p; \Theta)$  must be calculated at  $[N/2] + 1$  frequencies for each iteration. We now show that, by suitably smoothing the periodogram and working at a coarser grid of frequencies, the computation time can be substantially reduced while preserving our likelihood-based approach.

Consider splitting the data sequence into a number of segments of equal length. Providing these segments are long enough, they will be approximately independent by assumption (a result of (1)), so we can treat them as if they were independent replications of the same stationary stochastic process. Specifically, let the number of segments be  $S$ ; then the length of each segment is  $[N/S] = M$ , say. Fourier coefficients can be calculated for each segment, over the grid of frequencies  $\{\omega_{S_p}: p = 0, 1, \dots, [M/2]\}$ , where

$$\omega_{S_p} = \frac{2\pi p}{M\Delta}. \quad (15)$$

This grid of frequencies is coarser than that in (6), reflecting the fact that we are now dealing with shorter realizations. Denote by  $\bar{I}(\omega_{S_p})$  the mean of the  $S$  periodograms at frequency  $\omega_{S_p}$ , and by  $\bar{A}_0$  the mean Fourier coefficient at zero frequency. Then the approximate log-likelihood obtained by considering the Fourier coefficients of each segment separately is

$$\ln L_S(\Theta) = -S \sum_{p=1}^{\lfloor M/2 \rfloor} \left\{ 1 - \frac{1}{2} \delta_{p(M/2)} \right\} \left\{ \frac{\bar{I}(\omega_{S_p})}{4\pi h_Y(\omega_{S_p}; \Theta)} + \ln h_Y(\omega_{S_p}; \Theta) \right\} \\ - \frac{S}{2} \ln h_Y(0; \Theta) - \frac{M}{4\pi h_Y(0; \Theta)} \sum_{s=1}^S (\bar{Y}_s - \mu_Y(\Theta))^2 + \text{constant}. \quad (16)$$

This expression requires  $\lfloor M/2 \rfloor + 1$  terms to be evaluated at each iteration of the numerical optimization process, as opposed to  $\lfloor N/2 \rfloor + 1$  for the original likelihood (12); hence, if a large value of  $S$  can be chosen, substantial computational savings will result. Some comment on the choice of  $S$  is called for: the argument is the same as that for spectral smoothing, and may be found in Priestley (1981, Section 6.2.3) in greater detail. In order for the asymptotic theory to hold, we require both the number of segments,  $S$ , and the length of the segments,  $\lfloor N/S \rfloor$  to increase with  $N$ . A convenient way to do this is to set

$$S = \lfloor (N\Delta)^\alpha \rfloor \quad (17)$$

for some  $\alpha \in (0, 1)$  (the  $\Delta$  is included here so as to standardize between data at different temporal resolutions). The choice of  $\alpha$  is arbitrary; however, an important consideration is that, given a finite sample of data, the individual segments should be long enough that the sampling theory for the Fourier coefficients holds. In Section 4 models are fitted to a month's rainfall data from the UK, and results using various different values of  $\alpha$  are compared.

In Section 1 it was claimed that the spectral method makes efficient use of the available data. If the data set is split then the individual segments can be completely reconstructed from their Fourier coefficients, but information on the ordering of the segments has been lost. However, as the segments are assumed to be approximately independent, there is very little information to be had here in any case; hence when a replication is subdivided, the loss in efficiency is small, and we would expect little variation in the estimates. In the data analysis reported in Section 4, this is indeed the case.

The periodogram average  $\bar{I}(\omega_{S_p})$  in (16) is in fact the spectral estimate obtained by smoothing the original periodogram using Barlett's window (Priestley 1981, p. 439), and this suggests the use of other smoothed spectral estimates in place of  $\bar{I}(\omega_{S_p})$  in (16). This matter will not be pursued further here; our aim is to preserve as much information as possible, and the procedure described above achieves this in a straightforward way while maintaining a likelihood-based interpretation.

## 2.6. Multiple replications

It is convenient at this stage to extend the theory to the case when we have available a set of replications of a rainfall intensity process which can, for practical purposes, be regarded as independent. An example would be the case of a raingauge with several years' data; in this case, the years could be regarded as replications if we were interested in modelling the rainfall in one particular month. Where  $S$  replications are available (and assuming that they are all of the same length  $N$ ), it is easily shown that the likelihood function corresponding to (12) is (when all frequencies are used)



$$\ln L_S(\Theta) = -S \sum_{p=1}^{\lfloor N/2 \rfloor} \left\{ 1 - \frac{1}{2} \delta_{p(N/2)} \right\} \left\{ \frac{\bar{I}(\omega_p)}{4\pi h_Y(\omega_p; \Theta)} + \ln h_Y(\omega_p; \Theta) \right\} - \frac{S}{2} \ln h_Y(0; \Theta) - \frac{N}{4\pi h_Y(0; \Theta)} \sum_{s=1}^S (\bar{Y}_s - \mu_Y(\Theta))^2 + \text{constant}, \quad (18)$$

where now  $\bar{I}(\omega_p)$  denotes the mean periodogram at frequency  $\omega_p$  obtained from the  $S$  replications, and  $\bar{Y}_s$  is the sample mean rainfall intensity from the  $s$ th replication). The periodograms may also be used to assess the validity of the assumption that the replications are independent and identically distributed – see Diggle and Al-Wasel (1994), for example.

### 2.7. Comparing models

The motivation behind this method of parameter estimation was to develop a likelihood-based approach to the problem. A particular advantage of this kind of approach is that it allows different models to be compared objectively using standard techniques. Although neither (12), (16) nor (18) can strictly be regarded as a log-likelihood unless a very small proportion of the Fourier coefficients is used, it has been shown that they are interpretable as quasi-likelihoods, whence standard techniques such as likelihood ratio tests and the Akaike information criterion (AIC) may be used to assess model fit. The AIC is defined (Akaike 1974), when (12) is used to estimate  $\Theta$ , by

$$AIC(\Theta) = -2\ln L(\Theta) + 2K \quad (19)$$

(recall that  $K$  is the number of parameters being estimated) and, where it is used, the aim is to choose the model with the lowest AIC value (if either (16) or (18) is used to estimate  $\Theta$  then  $L_s(\cdot)$  is used instead of  $L(\cdot)$  in (19)). The principle behind the AIC is to minimize the expected Kullback–Leibler distance between the fitted and true models for the data (Akaike, 1974, Section IV). In fact the conclusions reported below do not depend critically on the use of AIC in preference to various other similar criteria that have been suggested in the literature.

In the present context, the use of either likelihood ratio tests or the AIC to compare models implicitly involves a comparison between the fitted and actual joint distributions of the sample Fourier coefficients. Hence, if (16) is used to estimate the parameters of each model, the same value of  $\alpha$  should be used in (17) for all models under consideration. Further, because the objective functions considered here are quasi-likelihoods rather than true likelihoods, the standard  $\chi^2$  null distributions for likelihood ratio testing must be modified (Kent 1982). This also implies some correction to the AIC if its information-theoretic interpretation is strictly to be maintained, for its derivation makes use of these null distributions. Determining the required correction to the AIC is beyond the scope of this paper; however, note that it is needed essentially because the two alternative forms of the Fisher information matrix (the expected score derivative matrix and the expected squared score matrix, respectively) are equal only when the underlying family of joint densities for

the Fourier coefficients is correctly specified. If these two matrices are approximately equal, the uncorrected AIC is still a reasonable criterion for model comparison.

### 3. Properties of the estimators

The properties of estimators arising from a maximization of (12) are well known in the time-series field, so will only be sketched here. For a fuller account the reader should refer to Rice (1979). Properties of estimators arising from (16) and (18) can be obtained in a similar manner. It will be convenient to denote the true value of  $\Theta$  by  $\Theta_0$ ; in addition, we will denote the estimator arising from (12) by  $\Theta_I$ , that arising from (16) by  $\Theta_{II}$  and that arising from (18) by  $\Theta_{III}$ .

#### 3.1. Consistency

Under mild conditions, it is straightforward to show that all three classes of estimators are consistent. Specifically, we have:

**Lemma 3.1.** *Providing the mixing condition (1) holds, the set  $\Omega$  of possible  $\Theta$ -values is compact, and the spectral density  $h_Y$  is continuous and unique for a given parametrization of the model,*

$$\Theta_I \xrightarrow{\text{Pr}} \Theta_0 \quad \text{as } N \rightarrow \infty, \quad (20)$$

with similar results holding for  $\Theta_{II}$  and  $\Theta_{III}$ .

The result for  $\Theta_I$  here is just that presented as Lemma 2.1 of Rice (1979); proofs of the results for  $\Theta_{II}$  and  $\Theta_{III}$  follow in exactly the same way.

#### 3.2. Convergence of the estimators

There are few results available for the large-sample properties of the estimators proposed here; in this section, following Rice (1979), we examine the speed of convergence of the estimators to the true value. The derivation is for  $\Theta_{II}$ ; the result for  $\Theta_I$  appears in Rice (1979), and that for  $\Theta_{III}$  follows similarly (for this estimator, we assume that the number of replications  $S$  stays constant, so that asymptotics are with respect to  $N$  only).

For large  $N$ ,  $\Theta_{II}$  is close to  $\Theta_0$ , by Lemma 3.1. Define  $\phi^{(i)}(\Theta) = \partial \ln L_S(\Theta) / \partial \theta_i$  ( $i = 1, \dots, K$ ), and define  $\Phi(\Theta) = (\phi^{(1)}(\Theta), \dots, \phi^{(K)}(\Theta))'$ . Then

$$\phi^{(i)}(\Theta_0) = \phi^{(i)}(\Theta_{II}) + \sum_{j=1}^K \frac{\partial \phi^{(i)}}{\partial \theta_j} \Big|_{\Theta=\Theta'} \left( \Theta_{II}^{(j)} - \Theta_0^{(j)} \right), \quad (21)$$

where  $\Theta_{II}^{(j)}$  is the  $j$ th component of the vector  $\Theta_{II}$ , and  $\Theta'$  is on the line segment joining  $\Theta_0$  and  $\Theta_{II}$ . Note that  $\phi^{(i)}(\Theta_{II}) = 0$  by definition; hence, combining all  $K$  equations (21) gives

$$\Phi(\Theta_0) = A(\Theta_{II} - \Theta_0), \tag{22}$$

where  $A$  is a  $K \times K$  matrix whose  $(i, j)$ th element is

$$A_{ij} = \left. \frac{\partial \phi^{(i)}}{\partial \theta_j} \right|_{\Theta = \Theta_0} = \left. \frac{\partial^2 \ln L_S(\Theta)}{\partial \theta_i \partial \theta_j} \right|_{\Theta = \Theta_0}. \tag{23}$$

The expansion (22) holds with probability tending to 1 as  $N \rightarrow \infty$ .

Rearranging (22) gives

$$(\Theta_{II} - \Theta_0) = A^{-1} \Phi(\Theta_0). \tag{24}$$

In order to make use of result (24), it is necessary to establish large-sample approximations for the right-hand side of the expression. We consider the case when  $M = [N/S]$  is odd – when  $M$  is even, the same procedure clearly applies. First, note that when  $M$  is odd the log-likelihood (16) can be written

$$\ln L_S(\Theta) = -S \sum_{p=1}^{[N/2S]} \left( \frac{\bar{I}(\omega_{S_p})}{4\pi h_Y(\omega_{S_p}; \Theta)} + \ln h_Y(\omega_{S_p}; \Theta) \right) + \Lambda(\Theta) + \text{constant}, \tag{25}$$

where  $\Lambda(\Theta)$  is the contribution from terms at zero frequency. Hence

$$\phi^{(i)}(\Theta) = \frac{\partial \ln L_S(\Theta)}{\partial \theta_i} = -S \sum_{p=1}^{[N/2S]} \frac{\partial h_Y(\omega_{S_p}; \Theta)}{\partial \theta_i} (h_Y(\omega_{S_p}; \Theta))^{-1} X(p, \Theta) + \frac{\partial \Lambda(\Theta)}{\partial \theta_i} \tag{26}$$

and

$$\begin{aligned} \frac{\partial^2 \ln L_S(\Theta)}{\partial \theta_i \partial \theta_j} &= S \sum_{p=1}^{[N/2S]} (h_Y(\omega_{S_p}; \Theta))^{-2} \left[ \frac{\partial h_Y(\omega_{S_p}; \Theta)}{\partial \theta_i} \frac{\partial h_Y(\omega_{S_p}; \Theta)}{\partial \theta_j} (2X(p, \Theta) - 1) \right. \\ &\quad \left. - X(p, \Theta) h_Y(\omega_{S_p}; \Theta) \frac{\partial^2 h_Y(\omega_{S_p}; \Theta)}{\partial \theta_i \partial \theta_j} \right] + \frac{\partial^2 \Lambda(\Theta)}{\partial \theta_i \partial \theta_j}, \end{aligned} \tag{27}$$

where

$$X(p, \Theta) = -\frac{\bar{I}(\omega_{S_p})}{4\pi h_Y(\omega_{S_p}; \Theta)} + 1. \tag{28}$$

Now standard results for the sampling theory of smoothed periodograms (for example, Brillinger 1975, Section 5.6) imply that, providing  $h_Y$  is bounded below and  $\partial h_Y / \partial \theta_i$  is bounded above for all  $\omega_{S_p}$ ,  $E\{\phi^{(i)}(\Theta_0)\} = O(S)$  and  $\text{var}\{\phi^{(i)}(\Theta_0)\} = O(N^2 S^{-1})$ ; hence  $\phi^{(i)}(\Theta_0) = O_p(S) + O_p(NS^{-1/2})$ . An additional consequence of these results is that, for large  $N$ ,  $X(p, \Theta') = o_p(1)$  uniformly in  $p$ , providing  $h_Y$  is bounded below for all  $\omega_{S_p}$  and is continuous, with continuous first  $\Theta$ -derivatives in the neighbourhood of  $\Theta_0$ . The first term in (27) is thus dominated, with probability tending to 1 as  $N \rightarrow \infty$ , by the quantity

$$-S \sum_{p=1}^{[N/2S]} (h_Y(\omega_{S_p}; \Theta))^{-2} \frac{\partial h_Y(\omega_{S_p}; \Theta)}{\partial \theta_i} \frac{\partial h_Y(\omega_{S_p}; \Theta)}{\partial \theta_j}.$$

In addition,  $\bar{Y} \xrightarrow{\text{Pr}} \mu_Y(\Theta_0)$  as  $N \rightarrow \infty$  so, providing  $\partial\mu_Y/\partial\theta_i$  and  $\partial^2\mu_Y/\partial\theta_i\partial\theta_j$  are bounded, the behaviour of  $\Lambda(\Theta)$  and its derivatives is dominated by that of  $\ln h_Y(0; \Theta)$ . Thus we can approximate

$$\frac{\partial\Lambda(\Theta)}{\partial\theta_i} \approx -\frac{S}{2h_Y(0; \Theta)} \frac{\partial h_Y(0; \Theta)}{\partial\theta_i} \tag{29}$$

$$\frac{\partial^2\Lambda(\Theta)}{\partial\theta_i\partial\theta_j} \approx \frac{S}{2(h_Y(0; \Theta))^2} \frac{\partial h_Y(0; \Theta)}{\partial\theta_i} \frac{\partial h_Y(0; \Theta)}{\partial\theta_j}, \tag{30}$$

and, providing  $\partial^2 h_Y/\partial\theta_i\partial\theta_j$  is bounded above for all  $\omega_{S_p}$ , we can approximate

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L_S(\Theta)}{\partial\theta_i\partial\theta_j} &\approx -\frac{S}{2N} \left( \frac{\partial}{\partial\theta_i} \ln h_Y(0; \Theta) \right) \left( \frac{\partial}{\partial\theta_j} \ln h_Y(0; \Theta) \right) \\ &- \frac{S}{N} \sum_{p=1}^{\lfloor N/2S \rfloor} \left( \frac{\partial}{\partial\theta_i} \ln h_Y(\omega_{S_p}; \Theta) \right) \left( \frac{\partial}{\partial\theta_j} \ln h_Y(\omega_{S_p}; \Theta) \right) \\ &\approx -\frac{\Delta}{2\pi} \int_{\omega=0}^{\pi/\Delta} \left( \frac{\partial}{\partial\theta_i} \ln h_Y(\omega; \Theta) \right) \left( \frac{\partial}{\partial\theta_j} \ln h_Y(\omega; \Theta) \right) d\omega = \alpha_{ij}, \text{ say,} \end{aligned} \tag{31}$$

as other terms in (27) will be small in comparison. The error involved in replacing the sum by an integral is  $O(S/N)^2$ , providing the second  $\omega$ -derivative of the integrand is bounded on the interval  $(0, \pi/\Delta)$ .

So we can approximate the matrix  $\mathbf{A}$  by  $N\alpha$ , where  $\alpha$  is a  $K \times K$  matrix with  $(i, j)$ th element  $\alpha_{ij}$ . Providing  $\alpha$  is invertible, we thus have  $\mathbf{A}^{-1} \approx N^{-1}\alpha^{-1}$ ; hence, from (24) and the order of magnitude of  $\Phi(\Theta_0)$ ,  $\Theta_{II} - \Theta_0 = O_p(S/N + S^{-1/2})$ , and we have shown the result for  $\Theta_{II}$  in the following:

**Lemma 3.2.** *Under the conditions of Lemma 3.1, and providing that:*

- (1) *the matrix  $\alpha$  in (31) is non-singular;*
- (2)  *$h_Y(\cdot; \Theta)$  does not vanish in the neighbourhood of  $\Theta_0$ ;*
- (3)  *$h_Y(\cdot; \Theta)$  is twice  $\Theta$ -differentiable in the neighbourhood of  $\Theta_0$ , and that all its first and second  $\Theta$ -derivatives are bounded in this neighbourhood;*
- (4) *the first and second  $\Theta$ -derivatives of  $\mu_Y(\Theta)$  exist and are bounded in the neighbourhood of  $\Theta_0$ ;*

*then, with probability tending to 1 as  $N \rightarrow \infty$ ,  $\Theta_I - \Theta_0 = O(N^{-1/2})$ ;  $\Theta_{II} - \Theta_0 = O(S/N + S^{-1/2})$ ; and  $\Theta_{III} - \Theta_0 = O(N^{-1/2})$ .*

The result derived for the estimator  $\Theta_{II}$  here could clearly be strengthened in principle, for setting  $\alpha = 0$  (17) yields the estimator  $\Theta_I$ . The reason for the apparent anomaly is that the standard results on the sampling theory of smoothed periodograms do not make any assumptions on the frequencies at which the periodograms are calculated; in particular, the covariance between periodogram ordinates is much less than those results would suggest when we restrict ourselves to frequencies in the set  $\{\omega_{S_p}\}$ . Unfortunately, it has not proved

possible to obtain specific results for the magnitude of these covariances, so the weaker result must stand at present.

It is also possible to show that, for large  $N$ , the estimators  $\Theta_I$ ,  $\Theta_{II}$  and  $\Theta_{III}$  are approximately normally distributed. The argument relies on the fact that, by (24), the error matrix  $(\Theta_{II} - \Theta_0)$  is a linear combination of the  $\{\phi^{(i)}\}$ . These are, in turn, each a weighted sum of smoothed periodograms, which themselves have a limiting normal distribution (Brillinger and Rosenblatt 1967). Hence the error matrix  $(\Theta_{II} - \Theta_0)$  tends to a linear combination of normal random variables, and is itself normally distributed. The covariance matrix of the quantity  $A(\Theta_{II} - \Theta_0)$  may be deduced from that of the vector  $\Phi(\Theta_0)$ , and depends on the fourth-order spectrum of the rainfall intensity process.

In fact, such distributional results are not very useful when applied to many of the models in current use. Apart from the difficulty in obtaining expressions for fourth-order spectra, there are often problems with parameter identifiability with these models, in the sense that widely separated regions of the parameter space can give rise to similar objective function values. This is a phenomenon which has been reported in the literature in connection with other methods of parameter estimation (see, for example, Rodriguez-Iturbe *et al.* 1988, p. 288; Onof and Wheeler 1993), and appears to be a feature of this type of model.

## 4. Application to data

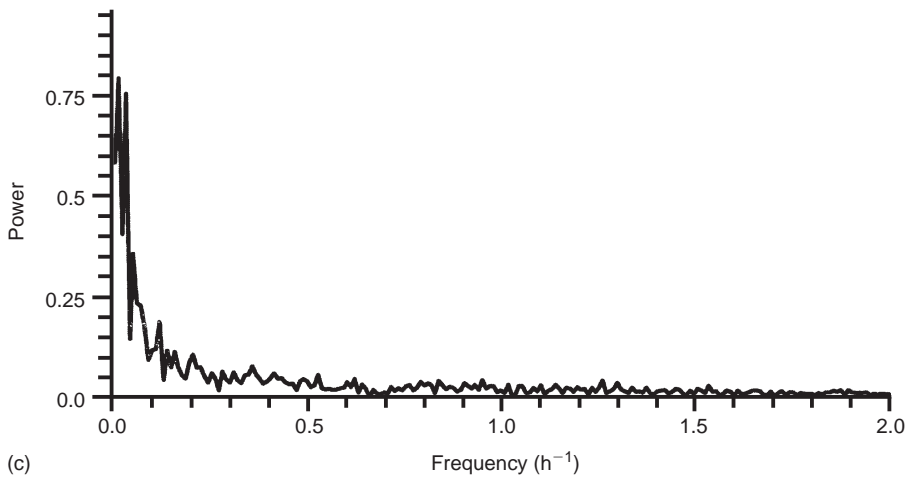
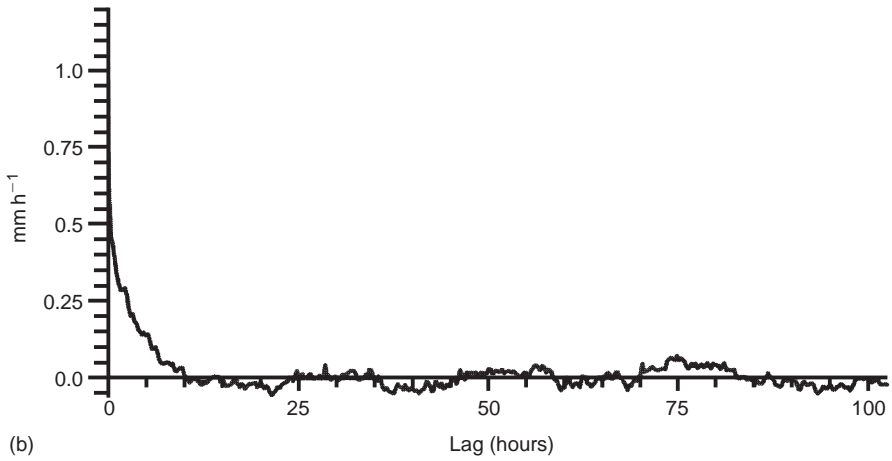
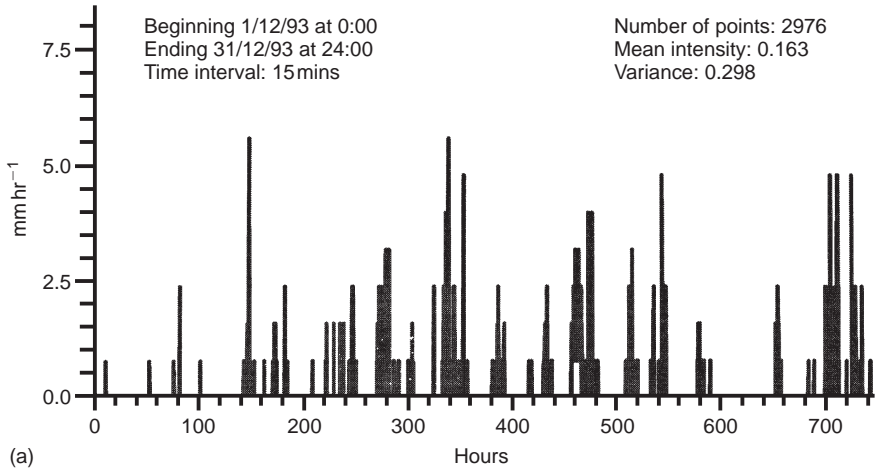
The techniques described above are now illustrated using data, shown in Fig. 2, from a raingauge in the south-west of England. They consist of 15-minute rainfall intensities (so that  $\Delta = 0.25$ ) for December 1993. The gauge is of 'tipping-bucket' type, with buckets of size 0.2 mm. The mean rainfall intensity in any 15-minute interval is therefore obtained by counting the number of tips in that interval, and multiplying by 0.8 (to yield an intensity in millimetres per hour). Figure 2 also shows the autocorrelation function, and estimated spectral density, for these data. The spectral density was estimated using the smoothed periodogram ( $\bar{I}(\omega_{S_p})$  in (16)); the value of  $\alpha$  used in (17) to obtain this was 0.3. Model fitting was carried out using both (12) and (16). Two separate values of  $\alpha$ , 0.3 and 0.5, were tried in the latter. When calculating the theoretical spectral density of the model, the infinite sum in (3) was truncated at  $k = \pm 5$ .

The choice of  $\alpha$ -values for the smoothed periodogram estimate was motivated by a desire to assess the robustness of the method in cases when the sample size is not large enough to claim that limiting behaviour has been attained. In particular, when  $\alpha = 0.5$ , in our case the data are being split up into segments of 27 hours' duration, which is clearly well below the length needed to claim that the large-sample approximations hold. Even with  $\alpha = 0.3$ , the segments are only 106 hours long; a comparison of these results with those obtained when  $\alpha = 0$  will give an idea of the sensitivity of the method to such departures from assumptions.

An inspection of Fig. 2 reveals that the series appears stationary, consisting of 12–13

---

**Figure 2.** Rainfall intensity series used in Section 4, together with autocorrelation function and periodograms: (a) rainfall intensity series for gauge 3; (b) autocorrelation function; (c) estimated spectral density



'storms' interspersed with dry periods. Of these storms about 7 or 8 are quite major events, with smaller events occurring in the first 100 hours and around the 400- and 600-hour periods. There is some quasi-periodicity in the data, indicated by a sinusoidal autocorrelation function and a peak in the smoothed periodogram away from the origin. The mean rainfall intensity is  $0.163 \text{ mm h}^{-1}$ .

Ten separate models, of the point-process type described in Section 2.1.1, were fitted to the data set. Each is specified by its driving point process, together with the distributions of cell depth and duration. The specifications for the various models are shown in Table 1 (note that the fitting method cannot handle depth distributions with more than two parameters, as the properties of the cell depths enter the spectral density (4) only through their mean and variance). For a discussion of the Poisson, Bartlett–Lewis and Neyman–Scott point-process models, see Cox and Isham (1980, Chapter 3); for the Markov renewal model, see Smith and Karr (1983). For a brief summary of all these models, see the Appendix to the present paper.

For the 'clustering' models 3–8, the number of cells per storm (that is, the number of offspring per parent in the point process) was taken to have a geometric distribution, and parent events were not included in the process (there is no cell at the 'storm origin'). For the Bartlett–Lewis models, the distribution of inter-cell times within a storm was taken to be exponential. For the Markov renewal model, an exponential distribution of dry period lengths was used. The two different Neyman–Scott models – one with a Gaussian scatter of cells about a storm origin, and the other with an exponential scatter – were included as they will give rise to rather different storm profiles; the Gaussian scatter will produce symmetrical storms, while the exponential scatter will produce storms which start intensely and then gradually decay.

Experience, both of this and of other methods of model fitting, has shown that for the type of model we are considering, numerical optimization algorithms are frequently hampered by the presence of many local optima in the function to be optimized. It is therefore necessary to perform the optimization several times from different starting points, and to select the best set of parameter estimates from the results.

In our work, the fitting strategy was as follows: for models 1, 3, 5, 7 and 9, initial estimates were obtained from tables of parameters in Onof (1992). Random perturbations

**Table 1.** Specification of models fitted to the data set used in Section 4

Model	Driving process	Depth distribution	Duration distribution
1	Poisson	Exponential	Exponential
2	Poisson	2-parameter	Gamma
3	Bartlett–Lewis	Exponential	Exponential
4	Bartlett–Lewis	2-parameter	Gamma
5	Neyman–Scott, Gaussian scatter	Exponential	Exponential
6	Neyman–Scott, Gaussian scatter	2-parameter	Gamma
7	Neyman–Scott, Exponential scatter	Exponential	Exponential
8	Neyman–Scott, Exponential scatter	2-parameter	Gamma
9	Markov renewal	Exponential	Exponential
10	Markov renewal	2-parameter	Gamma

were added to these initial estimates to obtain 20 different sets of starting values for the numerical optimization procedure. The Fortran NAG routine E04JAF was then used to carry out the optimization from each of these starting values. The best sets of parameter estimates thus obtained were used as initial estimates for the remaining models (which are simply generalized versions of the exponential models), and perturbed to give 20 different starting values for each model.

#### 4.1. Results of the fitting procedure

The results reported in this section represent, for each model, the best of the 20 sets of parameter estimates obtained. In assessing model fit, the Akaike information criterion (19) was used – although to maintain its information-theoretic interpretation, it should strictly be corrected for the fact that the estimates are not true likelihood estimates (recall Section 2.7). For the present discussion, it seems a reasonable (and simple) criterion to use merely for illustrative purposes – its information-theoretic interpretation may be doubtful, but it still provides a quick and simple means of penalizing overparametrized models, and (as will be seen below) the results it produces are sensible and agree with those reported elsewhere in the literature (for example, Rodriguez-Iturbe *et al.* 1987; 1988).

Table 2 shows the AIC values achieved by the different models, for the different values of the smoothing parameter  $\alpha$ . It is immediately clear that the Poisson models 1 and 2 provide the poorest fit, with the Markov renewal models 9 and 10 doing rather better in AIC terms. However, it is the clustering models 3–8 which have the lowest AIC values and hence provide the best fit. It is interesting to note the difference between the Neyman–Scott models with Gaussian cell scatter and those with exponential scatter, with the exponential scatter models performing rather better. The conclusion from this is that storms are not symmetrical, and that rainfall data are not time-reversible; this is in line with results reported by other authors, among them Rodriguez-Iturbe *et al.* (1988, p. 295). There is

**Table 2.** AIC values for models fitted to this data set (asterisks indicate best-fitting models)

Model	No. of parameters	AIC value		
		$\alpha = 0.0$	$\alpha = 0.3$	$\alpha = 0.5$
1	3	–18 001.4	–1734.80	–292.298
2	5	–18 099.3	–1744.32	–293.147
3*	5	–18 412.8	–1793.19	–305.771
4	7	–18 410.9	–1789.29	–301.972
5	5	–18 383.0	–1788.57	–304.491
6	7	–18 381.6	–1785.35	–300.526
7*	5	–18 415.1	–1793.48	–305.974
8	7	–18 411.4	–1789.48	–301.974
9	5	–18 186.5	–1749.58	–290.039
10	7	–18 216.1	–1751.00	–289.527



hardly any difference between the Neyman–Scott exponential scatter and Bartlett–Lewis models.

Turning now to the specification of cell depth and duration, it is seen that the exponential distribution seems an adequate choice for both of these quantities; the use of two-parameter distributions improves the fit only for the worst-fitting models (Poisson and Markov renewal).

The conclusions regarding model fit are affected little by the choice of the smoothing parameter  $\alpha$ . In fact, the rankings of the models by AIC value are exactly the same for  $\alpha = 0.0$  as they are for  $\alpha = 0.3$ . The rankings for  $\alpha = 0.5$  differ slightly, this being due mainly to the fact that there are far fewer periodogram ordinates contributing to the likelihood in this case, so the ability to distinguish between models is lost if they perform similarly.

It is to be expected, from Lemma 3.2 and from the simulation results reported in Section 2.2, that the choice of  $\alpha$  may affect the values of the parameter estimates themselves. In fact, the estimates obtained using a value of 0.3 were generally very similar to those using a value of zero; those resulting from  $\alpha = 0.5$  were substantially different, however. The general pattern was that estimates were more robust to changes in  $\alpha$  for the better-fitting models. The least robust parameters were: the cell depth parameters for models with two-parameter depth distributions; and the mean number of cells per storm in all clustering models, which tended to increase with  $\alpha$  (this being compensated for by a decrease in the storm arrival rate).

For the two best-fitting models, 3 and 7, the parameter estimates are given in Table 3. For reasons that have already been discussed, the estimates for  $\alpha = 0.0$  are preferable to those when  $\alpha = 0.3$ , and from here on we confine our observations to these estimates, noting simply that there is very little difference in practical terms between the two sets, and that the second set was produced around seven times faster than the first.

There is broad agreement in Table 3 when we compare the models, in particular with regard to the properties of individual cells. Both fitted models yield a mean rainfall intensity of  $0.164 \text{ mm h}^{-1}$ ; the expected number of storms in the month is 8.0 for the Bartlett–Lewis model, and 6.5 for the Neyman–Scott. These results tie in well with our earlier assessment of the data, although it is clear that the method has effectively ignored

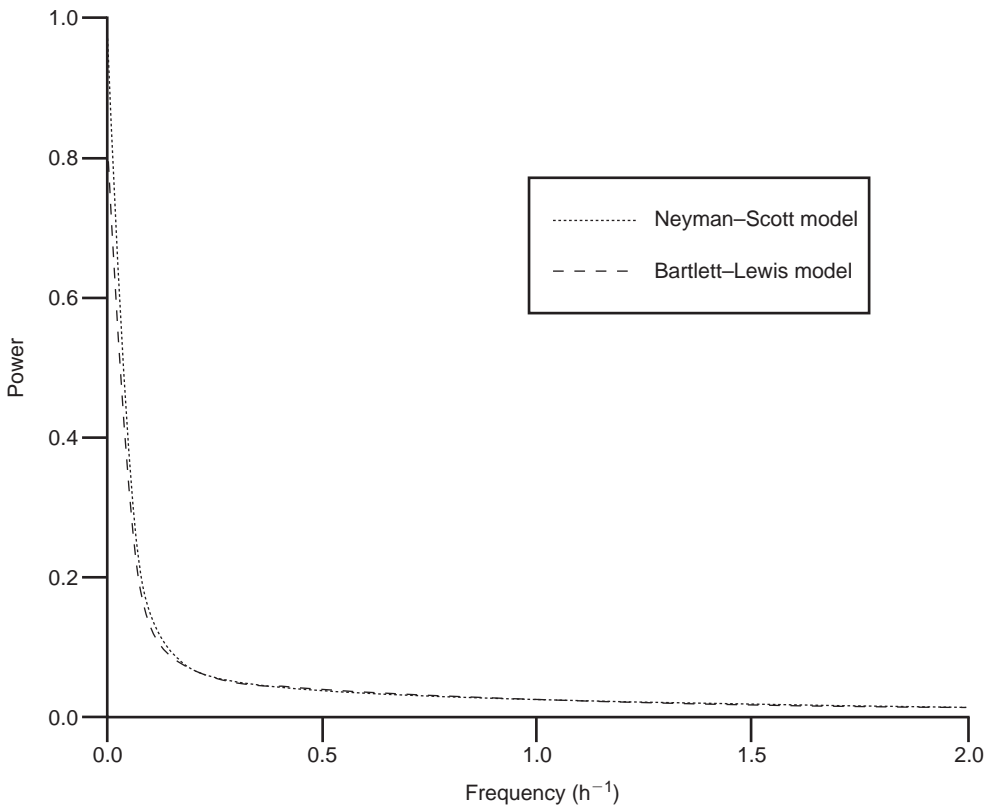
**Table 3.** Parameter estimates for best-fitting models ( $\alpha = 0.0$  and 0.3)

Parameter	Bartlett–Lewis model		Neyman–Scott model	
	$\alpha = 0.0$	$\alpha = 0.3$	$\alpha = 0.0$	$\alpha = 0.3$
Mean cell depth ( $\text{mm h}^{-1}$ )	3.79	3.73	3.76	3.69
Mean cell duration (h)	0.0986	0.107	0.0928	0.103
Storm arrival rate ( $\text{h}^{-1}$ )	0.0107	0.0088	0.0087	0.0074
Mean no. of cells/storm	40.9	46.6	53.8	58.1
Parameter of inter-cell dbn	9.96	9.87		N/A
Parameter of dispersal dbn		N/A	0.224	0.206

the smaller events. The theoretical spectra  $h_Y(\cdot)$  are shown for both models in Fig. 3, whence it is seen that neither model captures the observed quasi-periodicity (the spectra are both maximal at the origin).

In this work, the infinite sum at (3) was truncated at  $k = \pm 5$ . Similar fitting exercises have been carried out for different values of  $k$ . It was found that, for  $|k| \geq 2$ , parameter estimates changed very little, so for 15-minute rainfall data it probably suffices to truncate (3) at  $k = \pm 2$ ; for data at coarser temporal resolutions, this must clearly be increased accordingly.

There seems little point in using the theory developed in Section 3 to generate confidence intervals for the parameter estimates, because there are so many local maxima to the likelihood function. Indeed, there appear to be widely separated regions of the parameter space that are consistently favoured over others; in particular, estimates of the mean number of cells per storm and storm arrival rate lie in distinct ‘bands’. This may indicate that there is another layer of clustering in the data which has not been captured by the models, or that there is more than one type of storm.



**Figure 3.** Theoretical spectra for clustering models fitted to this data set

## 5. Discussion

The data analysis in the previous section demonstrates that this method of fitting models is feasible, and that it gives sensible results. The use of the smoothed periodogram in the fitting procedure gives rise to increased bias in the estimates for fixed sample sizes; but the considerable computational savings more than compensate for this, especially in view of the number of local optima in the objective function. The presence of local optima may appear disturbing, but it is a common feature of the type of model being considered here that different sets of parameters can produce very similar behaviour; the phenomenon is also encountered with the more usual moment-based approach, and is essentially due to the simplicity of the models rather than any defect on the part of the estimation procedure.

The method is particularly useful when applied to models for which certain theoretical properties are hard to derive. The only theoretical knowledge required of a model is its mean and second-order spectral density; this allows us to remove certain assumptions (regarding the exponential distribution of cell durations, for example) that are typically made in order to fit models using methods of moments. In addition, it is possible to drop the assumption that cell depth and duration are independent, and still obtain tractable expressions for theoretical spectra.

Another advantage of the technique is that, because of its efficient use of the data, it can be used with relatively small data sets (compared with, say, methods of moments). However, because of its reliance on first- and second-order properties the fitted models tend, whatever the sample size, to be poor at capturing features such as the wet and dry interval properties of the data. In applications, it is often more important to model such features than to capture fully the low-order moment structure of the data; hence this is a potential disadvantage. A more detailed investigation of this, and other issues of an applied nature, is currently in preparation.

Although the objective functions (12), (16) and (18) were derived as likelihood functions by treating the sample Fourier coefficients as data, it has been seen that in general these functions cannot strictly be regarded as likelihoods. However, it is possible to view them as quasi-likelihood functions if we consider the data to be the vector of periodogram ordinates; this perspective allows models to be compared using standard likelihood-based techniques. A possible refinement of the method would involve the inclusion of terms which are  $O(T^{-1})$  in the covariance matrix of the periodograms.

This work is currently being extended to fit spatiotemporal rainfall models to rain radar data.

## Appendix: spectral densities of some common point-process models

The spectral densities listed are the *incomplete* spectral densities ( $g_D^*(\cdot)$  in the notation of Section 2.1.1). We give the general formula and, where necessary, the value at zero frequency. A fuller account of most models may be found in Cox and Isham (1980, Chapter 3).

*Poisson process*

A renewal process whose intervals are exponential random variables with parameter  $\lambda$ .

$$g_D^*(\omega) = 0 \quad \forall \omega. \quad (32)$$

*Renewal process*

Intervals between events are i.i.d. with mean  $\nu$ ; and their characteristic function is  $\phi(\cdot)$ .

$$g_D^*(\omega) = \frac{1}{\nu\pi} \Re \left( \frac{\phi(\omega)}{1 - \phi(\omega)} \right) \quad (33)$$

$$g_D^*(0) = 0. \quad (34)$$

*Bartlett–Lewis process*

‘Parent’ events form a Poisson process of rate  $\rho$ ; each parent gives rise to a random number,  $Z$ , of offspring, which form a finite renewal process starting at the position of the parent.  $Z$  has mean  $\mu_Z$  and probability generating function  $\Pi_Z(\cdot)$ . The inter-offspring distribution has characteristic function  $\phi(\cdot)$ . Parents may or may not be included in the final process; the results here are for when they are not.

$$g_D^*(\omega) = \frac{\rho}{\pi} \Re \left[ \frac{\phi(\omega)}{1 - \phi(\omega)} \left( \mu_Z - \frac{1 - \Pi_Z(\phi(\omega))}{1 - \phi(\omega)} \right) \right] \quad (35)$$

$$g_D^*(0) = \frac{\rho}{2\pi} E\{Z(Z - 1)\}. \quad (36)$$

*Neyman–Scott process*

As Bartlett–Lewis, but now offspring are independently and identically scattered about the parent event according to some distribution with characteristic function  $\phi(\cdot)$ . Again, the results are for when parents are not included.

$$g_D^*(\omega) = \frac{\rho}{2\pi} E\{Z(Z - 1)\} |\phi(\omega)|^2 \quad \forall \omega. \quad (37)$$

*Generalized Neyman–Scott process (Chandler 1993, Section 2.2.6)*

As Neyman–Scott, but without the restriction that the parent process be Poisson. We assume it has incomplete spectral density  $g_p^*(\cdot)$ . Parents are not included.

$$g_D^*(\omega) = \frac{|\phi(\omega)|^2}{2\pi} [\rho E\{Z(Z - 1)\} + 2\pi\mu_Z^2 g_p^*(\omega)] \quad \forall \omega. \quad (38)$$

Markov renewal process (Smith and Karr 1983)

‘Wet’ periods and ‘dry’ periods alternate. The length of wet periods is exponentially distributed with mean  $\gamma$ ; and the length of ‘dry’ periods has mean  $\mu$  and characteristic function  $\phi(\cdot)$ . The length of the current period is independent of all previous periods. During dry periods, no events occur; during wet periods, events occur in a Poisson process with rate  $\rho$ .

$$g_D^*(\omega) = \frac{\rho^2 \gamma}{\pi(\gamma + \mu)} \Re \left[ \frac{i(\phi(\omega) - 1)}{\omega(1 - i\omega\gamma - \phi(\omega))} \right] \quad (39)$$

$$g_D^*(0) = 0, \quad (40)$$

the latter result following from the fact the process can be represented as a renewal process.

## Acknowledgements

Thanks are due to Anastasia Kakou, for thoroughly checking much of this work; to Kostas Skouras, for some extremely enlightening discussions regarding asymptotic theory; to Valerie Isham and David Cox for their helpful comments and encouragement; and to the referees for some stimulating and constructive criticism.

The research was supported by the Natural Environment Research Council, as part of the HYREX project.

## References

- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Automat. Control*, **AC-19**(6), 716–723.
- Bartlett, M.S. (1963) The spectral analysis of point processes. *J. Roy. Stat. Soc. Ser. B*, **25**, 264–296.
- Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart & Winston.
- Brillinger, D.R. and Rosenblatt, M. (1967) Asymptotic theory of estimates of  $k$ th-order spectra. In B. Harris (ed.), *Spectral Analysis of Time Series*, pp. 153–188. New York: Wiley.
- Chandler, R.E. (1993) The application of spectral methods to the analysis of point process data. PhD thesis, University of Manchester Institute of Science and Technology.
- Cox, D.R. and Isham, V.S. (1980) *Point Processes*. London: Chapman & Hall.
- Diggle, P.J. and Al-Wasel, I. (1994) Spectral analysis of replicated time series. Technical Report MA 94/97, Department of Mathematics, Lancaster University.
- Dzhaparidze, K.O. and Yaglom, A.M. (1983) Spectrum parameter estimation in time series analysis. In P.R. Krishnaiah, (ed.), *Developments in Statistics*, Vol. 4. New York: Academic Press.
- Kent, J.T. (1982) Robust properties of likelihood ratio tests. *Biometrika*, **69**, 19–27.
- Krzanowski, W.J. (1988) *Principles of Multivariate Analysis: A User's Perspective*. Oxford: Oxford University Press.
- McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman & Hall.
- Onof, C. (1992) Stochastic modelling of British rainfall using Poisson processes. PhD thesis, Department of Civil Engineering, Imperial College, London.

- Onof, C. and Wheater, H.S. (1993) Modelling of British rainfall using a random parameter Bartlett–Lewis rectangular pulse model. *J. Hydrology*, **149**, 67–95.
- Percival, D.B. and Walden, A.T. (1993) *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge: Cambridge University Press.
- Priestley, M.B. (1981) *Spectral Analysis and Time Series*. New York: Academic Press.
- Rice, J. (1979) On the estimation of the parameters of a power spectrum. *J. Multivariate Anal.*, **9**, 378–392.
- Rodriguez-Iturbe, I., Cox, D.R. and Isham, V.S. (1987) Some models for rainfall based on stochastic point processes. *Proc. Roy. Soc. London Ser. A*, **410**, 269–288.
- Rodriguez-Iturbe, I., Cox, D.R. and Isham, V.S. (1988) A point process model for rainfall: further developments. *Proc. Roy. Soc. London Ser. A*, **417**, 283–298.
- Smith, J.A. and Karr, A.F. (1983) A point process model of summer seasonal rainfall occurrences. *Water Resources Res.*, **19**(1), 95–103.
- Whittle, P. (1953) Estimation and information in stationary time series. *Ark. Mat.*, **2**, 423–434.

Received March 1995 and revised July 1996