A Generalized Point Process Model for Rainfall

Author(s): Paul S. P. Cowpertwait

Source: *Proceedings: Mathematical and Physical Sciences*, Oct. 8, 1994, Vol. 447, No. 1929 (Oct. 8, 1994), pp. 23-37

Published by: Royal Society

Stable URL: http://www.jstor.com/stable/52429

A generalized point process model for rainfall

By Paul S. P. Cowpertwait
Departments of Engineering Mathematics and Civil Engineering, University of Newcastle, Newcastle upon Tyne NE1 7RU, U.K.

This paper further develops a stochastic model for rainfall at a single site, in which storms arrive in a Poisson process, each storm generating a cluster of rain cells, with each cell having a random duration and a random intensity. The model is generalized by allowing each generated cell to be of n types. The duration of each cell is an exponential random variable that has parameter dependent on the cell type. The distribution of cell intensity is also dependent on the cell type, so that the generalized model provides a correlation between the intensities and durations of the generated rain cells. The case for two cell types is considered in some detail. The cells are categorized as either 'heavy' or 'light', where the heavy cells have a shorter expected lifetime than the light cells, which agrees with observational studies on precipitation fields. Properties are derived and used to fit the model to rainfall data at a single site. The adequacy of fit is assessed by considering properties not used in the fitting procedure but which are of interest in applications, e.g. extreme values. Further properties are derived which enable the model to be fitted to multi-site rainfall data.

1. Introduction

This paper further develops a stochastic model for rainfall at a single site, in which the parameters relate to underlying physical features observed in precipitation fields, e.g. convective cells. Precipitation areas have been classified according to their horizontal extent, which can be observed using satellites and radar. Cyclones, which have areas greater than 10^4 km^2, usually contain one or more frontal systems that generally last for several hours. Convective cells build and dissipate within the frontal systems, and have areas of the order 10–30 km^2. These cells tend to occur as clusters within areas of lighter intensity rainfall and last from several minutes to approximately $\frac{1}{2}$ h (Austin & Houze 1972). The existence of high intensity convective cells within lower intensity large mesoscale areas of precipitation has led some authors to broadly divide the precipitation field into 'heavy' convective rain and 'light' stratiform rain (e.g. Kavvas & Herd 1985).

The existence of clusters of rain cells within storms has motivated the development of clustered point process models for rainfall at a single site. Two suitable cluster models for precipitation have been studied in some detail: the Neyman–Scott Rectangular Pulses (NSRP) model (Rodriguez-Iturbe *et al.* 1987*a*, *b*; Cowpertwait 1991) and the Bartlett–Lewis Rectangular Pulses (BLRP) model (Rodriguez-Iturbe *et al.* 1987*a*, *b*, 1988). For both models, a storm origin, which can be interpreted as the time of arrival of the leading edge of a frontal system, arrives in a Poisson process. The models differ, however, in the location of the cell origins relative to the storm origin. In the BLRP model each storm origin is followed by another Poisson process of cell origins, which is terminated after a time that is exponentially distributed.

Proc. R. Soc. Lond. A (1994) **447**, 23–37
Printed in Great Britain

Thus, for the BLRP model the cell origins are independently uniformly distributed over the interval from the storm origin to the termination point (Cox & Isham 1980, p. 46). In the NSRP model the cell origins are independently displaced from each storm origin by distances that are exponentially distributed. Thus, in the NSRP model there is a greater probability that the cells will cluster near the beginning of the storm. Consequently, the NSRP model may be more plausible from the physical standpoint, because rainfall has a tendency to be more intense near the leading edge of some frontal systems (Browning 1985).

Both the NSRP and BLRP models have been extended to allow the distribution of cell durations to vary from storm to storm, providing an improved fit to the proportion of dry intervals (Rodriguez-Iturbe *et al.* 1988, Entekabi *et al.* 1989). However, a tendency to underestimate extreme values, particularly at the 1 h level of aggregation, was found. In this paper, these discrepancies are resolved by allowing different 'types' of rain cell to exist within the same storm. This is appealing because distinct precipitation areas have been identified in observational studies on rainfall fields.

2. Generalized point process model

In this section, the NSRP model is generalized by allowing multiple cell types to occur within the same storm. A Generalized NSRP model with n cell types is denoted by GNSRP(n), so a GNSRP(1) model is identical to the original NSRP model. An application of a GNSRP(2) model is considered in §4, where the two cell types are categorized as 'heavy' short-duration convective cells and 'light' long-duration stratiform cells. The generalized version of the NSRP model could be applied to fine resolution rainfall data composed of many different types of overlapping rain cells, or to multi-site rainfall data in which different cell types occur at different locations (described in §3). The GNSRP(n) model is now defined and the second-order aggregated properties derived.

Suppose that storm origins occur in a Poisson process of rate λ, and that a random number C of cell origins is associated with each storm origin. The cell origins are independently displaced from the storm origins by distances that are exponentially distributed with parameter β. Each cell origin is classed as one of n types, where α_i is the probability that the cell origin is of type $i (i = 1, \ldots, n)$. A rectangular pulse, or cell, is associated independently with each cell origin. The duration of the pulse is an independent exponential random variable with parameter η_i for type i cell origins. The intensity of the pulse is also an independent random variable and is denoted X_i for type i cell origins. The mean number of cells per storm is denoted as μ_C and, for a type i cell, the mean intensity denoted as $\mu_i (i = 1, \ldots, n)$.

Let $Y(t)$ be the total rainfall intensity at time t given by the GNSRP(n) model and let $X_{t-u}(u)$ denote the rainfall intensity at time t due to a cell with cell origin at $t-u$. Then

$$X_{t-u}(u) = \begin{cases} X_1 & \text{with probability} \quad \alpha_1 \mathrm{e}^{-\eta_1 u}, \\ X_2 & \text{with probability} \quad \alpha_2 \mathrm{e}^{-\eta_2 u}, \\ \cdot & \\ \cdot & \\ \cdot & \\ X_n & \text{with probability} \quad \alpha_n \mathrm{e}^{-\eta_n u}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

$Y(t)$ is the summation of all active cells at time t, i.e.

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u)\,\mathrm{d}N(t-u), \tag{2.2}$$

where
$$\mathrm{d}N(t-u) = \begin{cases} 1, & \text{if there is a cell origin at } t-u, \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$E\{Y(t)\} = \int_0^{\infty} \sum_{i=1}^{n} [E(X_i)\,\alpha_i\,\mathrm{e}^{-\eta_i u}]\,\lambda\mu_{\mathrm{C}}\,\mathrm{d}u = \lambda\mu_{\mathrm{C}} \sum_{i=1}^{n} \mu_i\,\alpha_i/\eta_i. \tag{2.3}$$

The lag τ autocovariance of the intensity process $Y(t)$ is given by

$$\begin{aligned} C_Y(\tau) &= \operatorname{Cov}\{Y(t), Y(t+\tau)\} \\ &= \int_{v=0}^{\infty}\int_{u=0}^{\infty} E\{X_{t-u}(u)\,X_{t+\tau-v}(v)\}\operatorname{Cov}\{\mathrm{d}N(t-u), \mathrm{d}N(t+\tau-v)\}, \end{aligned} \tag{2.4}$$

where, for the GNSRP(n) model,

$$E\{X_{t-u}(u)\,X_{t+\tau-v}(v)\} = \begin{cases} \sum_{i=1}^{n} (\mu_i\,\alpha_i\,\mathrm{e}^{-\eta_i u}) \sum_{j=1}^{n} (\mu_j\,\alpha_j\,\mathrm{e}^{-\eta_j v}), v \neq u+\tau \\ \sum_{i=1}^{n} E(X_i^2)\,\alpha_i\,\mathrm{e}^{-\eta_i v}, \quad v = u+\tau. \end{cases} \tag{2.5}$$

The autocovariance of the Neyman–Scott point process is

$$\operatorname{Cov}\{\mathrm{d}N(t-u), \mathrm{d}N(t+\tau-v)\} = \lambda\mu_{\mathrm{C}}[\delta(|u+\tau-v|) + \tfrac{1}{2}\mu_{\mathrm{C}}^{-1}E(C^2-C)\,\beta\,\mathrm{e}^{-\beta|u+\tau-v|}]\,\mathrm{d}u\,\mathrm{d}v, \tag{2.6}$$

where δ denotes the Dirac delta function (Cox & Isham 1980). Hence, substituting (2.5) and (2.6) into (2.4) gives, for $\tau \geqslant 0$,

$$C_Y(\tau) = \operatorname{Cov}\{Y(t), Y(t+\tau)\} = \sum_{j=1}^{n} C_j\,\mathrm{e}^{-\eta_j\tau} + C_\beta\,\mathrm{e}^{-\beta\tau}, \tag{2.7}$$

where

$$C_j = \frac{\lambda\mu_{\mathrm{C}}\,\alpha_j\,E(X_j^2)}{\eta_j} + \frac{\lambda\beta^2\mu_j\,\alpha_j\,E(C^2-C)}{\beta^2-\eta_j^2}\sum_{i=1}^{n}\frac{\mu_i\,\alpha_i}{\eta_i+\eta_j}, \quad j = 1, \ldots, n \tag{2.8}$$

and

$$C_\beta = -\tfrac{1}{2}\lambda\beta E(C^2-C)\sum_{i=1}^{n}\sum_{j=1}^{n}\{\mu_i\,\mu_j\,\alpha_i\,\alpha_j/[(\beta-\eta_j)(\beta+\eta_i)]\}. \tag{2.9}$$

Rainfall data are usually only available in aggregated form, e.g. as historical records of daily or hourly totals, so that aggregated properties of the model are needed to estimate the model parameters. Let $Y_i^{(h)}$ be the aggregated rainfall depth in the ith time interval of length h, so

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)\,\mathrm{d}t. \tag{2.10}$$

The series $\{Y_i^{(h)}: i = 1, 2, \ldots\}$ is thus a rainfall time series at the h level of aggregation. The second-order properties of the aggregated process can be obtained from

$$\left.\begin{aligned} E(Y_i^{(h)}) = hE\{Y(t)\}, \quad \operatorname{Var}(Y_i^{(h)}) = 2\int_0^h (h-u)\, C_Y(u)\,\mathrm{d}u, \\ \operatorname{Cov}(Y_i^{(h)}, Y_{i+k}^{(h)}) = \int_{-h}^{h} C_Y(kh+v)\,(h-|v|)\,\mathrm{d}v. \end{aligned}\right\} \tag{2.11}$$

(see, for example, Rodriguez-Iturbe *et al.* 1987*a*, §2.3). Hence, substituting (2.3) and (2.7) into (2.11),

$$E(Y_i^{(h)}) = h\lambda\mu_C \sum_{j=1}^{n} \alpha_j \mu_j/\eta_j, \tag{2.12}$$

$$\operatorname{Var}\{Y_i^{(h)}\} = \sum_{j=1}^{n}\left[\frac{2C_j}{\eta_j^2}(h\eta_j + \mathrm{e}^{-\eta_j h} - 1)\right] + \frac{2C_\beta}{\beta^2}(h\beta + \mathrm{e}^{-\beta h} - 1), \tag{2.13}$$

and for $k \geqslant 1$,

$$\operatorname{Cov}\{Y_i^{(h)}, Y_{i+k}^{(h)}\} = \sum_{j=1}^{n}\left[\frac{C_j}{\eta_j^2}\mathrm{e}^{-\eta_j(k-1)h}(1-\mathrm{e}^{-\eta_j h})^2\right] + \frac{C_\beta}{\beta^2}\mathrm{e}^{-\beta(k-1)h}(1-\mathrm{e}^{-\beta h})^2, \tag{2.14}$$

where C_j and C_β are defined in (2.8) and (2.9) respectively.

To obtain an expression for the probability of no rain in an arbitrary interval of length h, i.e. $\operatorname{pr}\{Y_i^{(h)} = 0\}$ denoted by $\phi(h)$, consider a single storm origin at time zero. Let W_i be the waiting time after the origin for a type i cell to terminate given that the cell origin is in $(0, t)$. Then

$$\operatorname{pr}\{W_i < t\} = \omega_i = 1 - \{\beta(\mathrm{e}^{-\beta t} - \mathrm{e}^{-\eta_i t})/[(\eta_i - \beta)(1-\mathrm{e}^{-\beta t})]\}, \quad i = 1, \ldots, n. \tag{2.15}$$

Now suppose there are k cells in $(0, t)$. Then

$$\operatorname{pr}\{\text{all } k \text{ cells terminate in } (0,t)\} = \left(\sum_{i=1}^{n} \alpha_i \omega_i\right)^k = \omega^k, \quad \text{say.}$$

Assuming the number of cells per storm C is a geometric random variable with parameter ν,

$$\begin{aligned} &\operatorname{pr}\{\text{no rain in } (t, t+h) \text{ due to a storm origin at time zero}\} \\ &= \sum_{j=1}^{\infty}\sum_{k=0}^{j}\binom{j}{k}(1-\mathrm{e}^{-\beta t})^k \omega^k\, \mathrm{e}^{-\beta(t+h)(j-k)}(1-\nu)^{j-1}\nu \\ &= \frac{\nu[\mathrm{e}^{-\beta(t+h)} + \omega(1-\mathrm{e}^{-\beta t})]}{1-(1-\nu)[\mathrm{e}^{-\beta(t+h)} + \omega(1-\mathrm{e}^{-\beta t})]} \\ &= p_h(t), \quad \text{say.} \end{aligned}$$

So $p_0(t)$ is the probability of no rain at time t due to a storm origin at time zero, and $p_t(0)$ is the probability of no rain in the interval $(0, t)$. Then the probability of no rain in an arbitrary interval of length h is given by

$$\phi(h) = \exp\left\{-\lambda\int_0^\infty [1-p_h(t)]\,\mathrm{d}t\right\}\exp\left\{-\lambda\int_0^h [1-p_t(0)]\,\mathrm{d}t\right\}, \tag{2.16}$$

where the first expression is the probability of no rain in interval h due to storm origins preceding the interval, and the second expression is the probability of no rain in the interval due to storm origins within the interval (Cowpertwait 1991, §3).

The first expression usually requires numerical evaluation. For the second expression, $p_t(0)$ may be substituted to give

$$\phi(h) = \exp\left\{-\lambda \int_0^\infty [1-p_h(t)]\,\mathrm{d}t\right\} \exp\left\{-\lambda h + \frac{\lambda\nu}{\beta(1-\nu)} \ln\left[\frac{1}{\nu} + \left(1-\frac{1}{\nu}\right)\mathrm{e}^{-\beta h}\right]\right\}. \tag{2.17}$$

Let the probability that an arbitrary interval of length h is wet given the previous interval of length h is wet be denoted by $\phi_{\mathrm{ww}}(h)$ and let the probability that an arbitrary interval of length h is dry given the previous interval of length h is dry be denoted by $\phi_{\mathrm{dd}}(h)$; i.e.

$$\phi_{\mathrm{ww}}(h) = \mathrm{pr}\{Y_{i+1}^{(h)} > 0 | Y_i^{(h)} > 0\} \quad \text{and} \quad \phi_{\mathrm{dd}}(h) = \mathrm{pr}\{Y_{i+1}^{(h)} = 0 | Y_i^{(h)} = 0\}.$$

These will be called the wet and dry spell transition probabilities respectively, and can be derived by considering adjacent time intervals $[0,h]$ and $[h,2h]$:

$$\mathrm{pr}\{[h,2h]\,\mathrm{dry}\,|\,[0,h]\,\mathrm{dry}\} = \mathrm{pr}\{[0,2h]\,\mathrm{dry}\}/\mathrm{pr}\{[0,h]\,\mathrm{dry}\},$$

i.e.

$$\phi_{\mathrm{dd}}(h) = \phi(2h)/\phi(h) \tag{2.18}$$

and

$$\mathrm{pr}\{[h,2h]\,\mathrm{dry}\} = \mathrm{pr}\{[h,2h]\,\mathrm{dry}\,|\,[0,h]\,\mathrm{dry}\}\,\mathrm{pr}\{[0,h]\,\mathrm{dry}\} + pr\{[h,2h]\,\mathrm{dry}\,|\,[0,h]\,\mathrm{wet}\}\,\mathrm{pr}\{[0,h]\,\mathrm{wet}\},$$

i.e.

$$\phi(h) = \phi_{\mathrm{dd}}(h)\,\phi(h) + [1-\phi_{\mathrm{ww}}(h)]\,[1-\phi(h)].$$

Hence,

$$\phi_{\mathrm{ww}}(h) = [1-2\phi(h)+\phi(2h)]/[1-\phi(h)]. \tag{2.19}$$

3. Multi-site model

Consider a catchment containing m sites. Suppose that cells arrive according to the GNSRP(n) process of §2, so that each cell origin is classed as one of n types, where α_i is the probability that the cell origin is of type $i (i = 1, \dots, n; \Sigma_{i=1}^{n} \alpha_i = 1)$. Each cell type is associated with one or more of the sites. Hence, at a single site having p associated cell types, the cells arrive in a GNSRP(p) process. However, note that $\Sigma_{i=1}^{p} \alpha_i$ does not equal one when $p < n$, so that, for some storm origins, no cells will be generated at some of the sites. This is likely to agree with empirical experience as summer convective storms are often highly localized.

Let $Y_i(t)$ be the total rainfall intensity at time t due to type i cells, and let $X_{t-u}^{(i)}(u)$ denote the rainfall intensity at time t due to a type i cell with cell origin at $t-u$, so that

$$Y_i(t) = \int_{u=0}^\infty X_{t-u}^{(i)}(u)\,\mathrm{d}N^{(i)}(t-u), \tag{3.1}$$

where

$$\mathrm{d}N^{(i)}(t-u) = \begin{cases} 1, & \text{if there is a type } i \text{ cell origin at } t-u \\ 0, & \text{otherwise.} \end{cases}$$

Suppose, without loss of generality, that p cell types, with probabilities $\alpha_i (i = 1, \dots, p)$, are associated with site 1. Then the total rainfall intensity at time t for site 1, denoted $Y^{(1)}(t)$, is the summation of the rainfall intensities due to all cell types that can occur at that site, i.e.

$$Y^{(1)}(t) = \sum_{i=1}^{p} Y_i(t). \tag{3.2}$$

The aggregated process is given by

$$S^{(1)}_{(h)}(a) = \int_{t=(a-1)h}^{ah} Y^{(1)}(t)\,\mathrm{d}t = \int_{(a-1)h}^{ah} \sum_{i=1}^{p} Y_i(t)\,\mathrm{d}t. \tag{3.3}$$

It will now be assumed that a pair of sites will have r cell types in common. The total intensity at site 2, which has cell types $p-r+1, p-r+2, \dots, p$ in common with site 1, is given by

$$Y^{(2)}(t) = \sum_{i=p-r+1}^{q} Y_i(t), \tag{3.4}$$

where $q \leqslant n$, and the single site rainfall time series at site 2 is given by a GNSRP$(q-p+r)$ model.

The lag τ cross-covariance of the intensity processes is given by

$$\begin{aligned} \operatorname{Cov}\{Y^{(1)}(t), Y^{(2)}(t+\tau)\} &= \operatorname{Cov}\left\{\sum_{i=1}^{p} Y_i(t), \sum_{j=p-r+1}^{q} Y_j(t+\tau)\right\} \\ &= \sum_{i=1}^{p} \sum_{j=p-r+1}^{q} \operatorname{Cov}\{Y_i(t), Y_j(t+\tau)\}. \end{aligned} \tag{3.5}$$

Now, for $i \neq j$,

$$\operatorname{Cov}\{Y_i(t), Y_j(t+\tau)\} = \int_{v=0}^{\infty} \int_{u=0}^{\infty} E\{X^{(i)}_{t-u}(u)\, X^{(j)}_{t+\tau-v}(v)\} \operatorname{Cov}\{\mathrm{d}N^{(i)}(t-u), \mathrm{d}N^{(j)}(t+\tau-v)\}, \tag{3.6}$$

where

$$\operatorname{Cov}\{\mathrm{d}N^{(i)}(t-u), \mathrm{d}N^{(j)}(t+\tau-v)\} = \begin{cases} \frac{1}{2}\lambda\beta\alpha_i\,\alpha_j E(C^2-C)\,\mathrm{e}^{-\beta(u+\tau-v)}\,\mathrm{d}u\,\mathrm{d}v, v < u+\tau \\ \frac{1}{2}\lambda\beta\alpha_i\,\alpha_j E(C^2-C)\,\mathrm{e}^{-\beta(v-u-\tau)}\,\mathrm{d}u\,\mathrm{d}v, v > u+\tau, \end{cases} \tag{3.7}$$

and

$$E\{X^{(i)}_{t-u}(u)\, X^{(j)}_{t+\tau-v}(v)\} = \begin{cases} E(X_i)\,E(X_j), & \text{with probability } \mathrm{e}^{-\eta_i u}\,\mathrm{e}^{-\eta_j v} \\ 0, & \text{otherwise,} \end{cases} \tag{3.8}$$

so that, substituting (3.7) and (3.8) into (3.6),

$$\begin{aligned} &\operatorname{Cov}\{Y_i(t), Y_j(t+\tau)\} \\ &= \tfrac{1}{2}\lambda\beta\alpha_i\,\alpha_j E(C^2-C)\,E(X_i)\,E(X_j)\left[\frac{2\beta\,\mathrm{e}^{-\eta_j\tau}}{(\beta^2-\eta_j^2)\,(\eta_i+\eta_j)} - \frac{\mathrm{e}^{-\beta\tau}}{(\beta-\eta_j)\,(\beta+\eta_i)}\right]. \end{aligned} \tag{3.9}$$

Writing

$$\sum_{i=1}^{p}\sum_{j=p-r+1}^{q}\operatorname{Cov}\{Y_i(t),Y_j(t+\tau)\}=\sum_{i\neq j}\operatorname{Cov}\{Y_i(t),Y_j(t+\tau)\}+\sum_{i=j}\operatorname{Cov}\{Y_i(t),Y_j(t+\tau)\}$$

gives (from 3.5)

$$\begin{aligned}\operatorname{Cov}\{Y^{(1)}(t),Y^{(2)}(t+\tau)\}\\ &=\sum_{i\neq j}\tfrac{1}{2}\lambda\beta\alpha_i\alpha_j E(C^2-C)E(X_i)E(X_j)\left[\frac{2\beta\,\mathrm{e}^{-\eta_j\tau}}{(\beta^2-\eta_j^2)(\eta_i+\eta_j)}-\frac{\mathrm{e}^{-\beta\tau}}{(\beta-\eta_j)(\beta+\eta_i)}\right]\\ &\quad+\sum_{i=j}\left[\frac{\lambda\mu_{\mathrm{C}}\alpha_i E(X_i^2)\,\mathrm{e}^{-\eta_i\tau}}{\eta_i}+\frac{\lambda\beta^2\mu_i^2\alpha_i^2E(C^2-C)\,\mathrm{e}^{-\eta_i\tau}}{2\eta_i(\beta^2-\eta_i^2)}-\tfrac{1}{2}\lambda\beta E(C^2-C)\frac{\mu_i^2\alpha_i^2\,\mathrm{e}^{-\beta\tau}}{\beta^2-\eta_i^2}\right].\end{aligned}\tag{3.10}$$

Using (3.10), the lag k cross-covariance of the aggregated processes for $k\geqslant 1$ is given by

$$\begin{aligned}\operatorname{Cov}\{S_{(h)}^{(1)}(a),S_{(h)}^{(2)}(a+k)\}\\ &=\operatorname{Cov}\left\{\int_{(a-1)h}^{ah}Y^{(1)}(t)\,\mathrm{d}t,\int_{(a+k-1)h}^{(a+k)h}Y^{(2)}(s)\,\mathrm{d}s\right\}\\ &=\int_{(a-1)h}^{ah}\int_{(a+k-1)h}^{(a+k)h}\operatorname{Cov}\{Y^{(1)}(t),Y^{(2)}(s)\}\,\mathrm{d}s\,\mathrm{d}t\\ &=\int_{(a-1)h}^{ah}\int_{(a+k-1)h}^{(a+k)h}\sum_{i=1}^{p}\sum_{j=p-r+1}^{q}\operatorname{Cov}\{Y_i(t),Y_j(s)\}\,\mathrm{d}s\,\mathrm{d}t\\ &=\tfrac{1}{2}\lambda\beta E(C^2-C)\sum_{i\neq j}\alpha_i\alpha_jE(X_i)E(X_j)\left[\frac{2\beta(1-\mathrm{e}^{-\eta_jh})^2\,\mathrm{e}^{-\eta_j(k-1)h}}{\eta_j^2(\beta^2-\eta_j^2)(\eta_i+\eta_j)}-\frac{(1-\mathrm{e}^{-\beta h})^2\,\mathrm{e}^{-\beta(k-1)h}}{\beta^2(\beta-\eta_j)(\beta+\eta_i)}\right]\\ &\quad+\sum_{i=j}\left\{\left[\frac{\lambda\mu_{\mathrm{C}}\alpha_iE(X_i^2)}{\eta_i^3}+\frac{\lambda\beta^2\mu_i^2\alpha_i^2E(C^2-C)}{2\eta_i^3(\beta^2-\eta_i^2)}\right]\mathrm{e}^{-\eta_i(k-1)h}(1-\mathrm{e}^{-\eta_ih})^2\right.\\ &\quad\left.-\frac{\lambda E(C^2-C)\mu_i^2\alpha_i^2\,\mathrm{e}^{-\beta(k-1)h}(1-\mathrm{e}^{-\beta h})^2}{2\beta(\beta^2-\eta_i^2)}\right\}.\end{aligned}\tag{3.11}$$

For $k=0$, the cross-covariance of the aggregated processes is given by

$$\begin{aligned}\operatorname{Cov}\{S_{(h)}^{(1)}(a),S_{(h)}^{(2)}(a)\}\\ &=\tfrac{1}{2}\lambda\beta E(C^2-C)\sum_{i\neq j}\alpha_i\alpha_jE(X_i)E(X_j)\left[\frac{2\beta(h\eta_j+\mathrm{e}^{-\eta_jh}-1)}{\eta_j^2(\beta^2-\eta_j^2)(\eta_i+\eta_j)}-\frac{(h\beta+\mathrm{e}^{-\beta h}-1)}{\beta^2(\beta-\eta_j)(\beta+\eta_i)}\right.\\ &\quad\left.+\frac{2\beta(h\eta_i+\mathrm{e}^{-\eta_ih}-1)}{\eta_i^2(\beta^2-\eta_i^2)(\eta_j+\eta_i)}-\frac{(h\beta+\mathrm{e}^{-\beta h}-1)}{\beta^2(\beta-\eta_i)(\beta+\eta_j)}\right]\\ &\quad+2\lambda\sum_{i=j}\left\{\left[\frac{\mu_{\mathrm{C}}\alpha_iE(X_i^2)}{\eta_i^3}+\frac{\beta^2\mu_i^2\alpha_i^2E(C^2-C)}{2\eta_i^3(\beta^2-\eta_i^2)}\right](h\eta_i+\mathrm{e}^{-\eta_ih}-1)\right.\\ &\quad\left.-\frac{E(C^2-C)\mu_i^2\alpha_i^2(h\beta+\mathrm{e}^{-\beta h}-1)}{2\beta(\beta^2-\eta_i^2)}\right\}.\end{aligned}\tag{3.12}$$

As a special case of the above, a simple multi-site model for m sites might consist of $(m+1)$ cell types, where cell types $1, 2, \ldots, m$ occur only at sites $1, 2, \ldots, m$ respectively and cell type $(m+1)$ occurs at all m sites. The $(m+1)$ cell type might be interpreted as 'light' stratiform rain, and the remaining cell types as 'heavy' localized convective rain. For any single site the GNSRP(2) model could then be applied, and aggregated cross-covariances fitted between pairs of sites using (3.11) and (3.12).

Broadly speaking, such an approach might be appropriate when the distances between pairs of sites are greater than the diameter of a convective cell, i.e. greater than about 4 km. When the distance between a pair of sites is less than the diameter of a convective cell it would be appropriate to have a proportion of convective cells occurring simultaneously at those sites. However, the approach described above may be suitable for many practical problems in which sites separated by small spatial distances can be assumed to receive the same rainfall time series.

4. Single site application

(a) *Fitting procedure*

A single site application of the GNSRP(2) model is considered, where type 1 cells are interpreted as 'heavy' short-duration convective cells and type 2 cells as 'light' long-duration stratiform cells. For the purpose of fitting the GNSRP(2) model, the intensities of the heavy and light rain cells, X_1 and X_2 respectively, are taken to be exponential random variables with parameters ξ_1 and ξ_2 respectively, giving a total of eight parameters.

If the model is fitted to each calendar month of a rainfall record, which would take account of seasonality, then 12 estimates per parameter would be needed, i.e. a total of 96 estimates per site. A large number of estimates per site can be difficult to interpret, particularly when the 12 estimates of a parameter show random variation and are correlated with the estimates of other parameters. An approach that ensures a smooth seasonal variation in the parameter estimates and also reduces the number of estimates per site is to assume that the 12 estimates for a parameter vary across the year according to harmonic relationships.

If $\hat{\psi}_i$ is a parameter estimate of the GNSRP(2) model for the ith calendar month (where $i = 1, 2, \ldots, 12$), so $\psi \in \{\lambda, \beta, \eta_1, \eta_2, \nu, \xi_1, \xi_2, \alpha\}$, then it is assumed that

$$\hat{\psi}_i = \hat{m}_\psi + \hat{A}_\psi \sin[(2\pi i/12) + \hat{\theta}_\psi], \tag{4.1}$$

where $\hat{m}_\psi$, $\hat{A}_\psi$, and $\hat{\theta}_\psi$ are estimates of harmonic variables. Instead of estimating each GNSRP model parameter for each calendar month, the harmonic variables can be estimated directly, resulting in three estimates per parameter, by minimizing the following sum of squares:

$$SS = \sum_{i=1}^{12} \sum_{h \in H} \sum_{f_i \in F} \left[1 - \frac{f_i(h)}{\hat{f}_i(h)}\right]^2, \tag{4.2}$$

subject to $m_\psi > 0, A_\psi \geqslant 0, 2\pi \geqslant \theta_\psi \geqslant 0$, where F is a set of aggregated properties for the GNSRP(2) model (e.g. given by 2.12, 2.13, 2.14, 2.17 and 2.19), $\hat{f}_i$ denotes the sample estimate of f_i for the ith calendar month, which is obtained by pooling all available data for the month, and H is a set of aggregation levels.

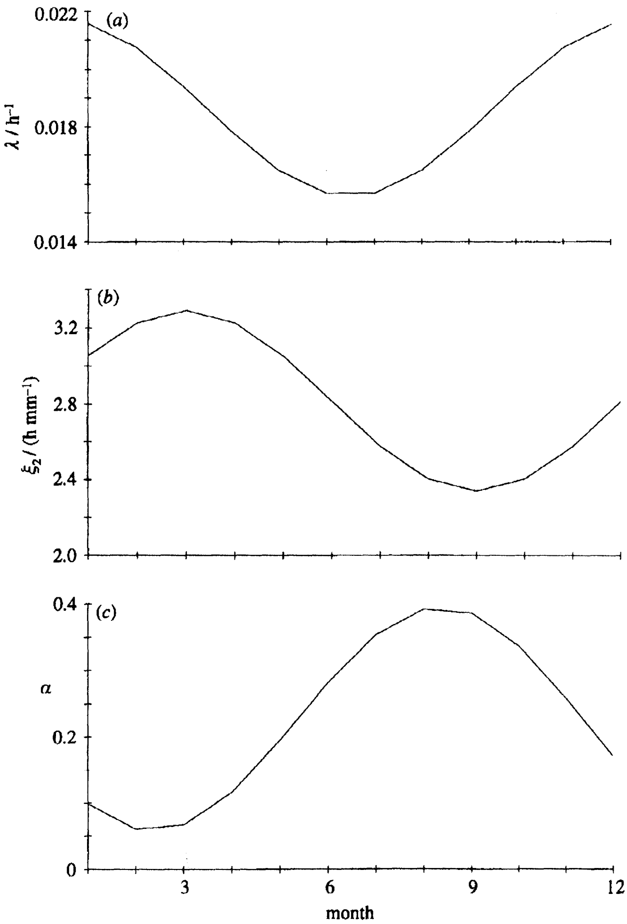

Figure 1. Parameter estimates against month. (*a*) λ; (*b*) ξ_2; (*c*) α.

Table 1. *Estimates of harmonic variables*

parameter ψ	$\hat{m}_\psi$	$\hat{A}_\psi$	$\hat{\theta}_\psi$/rad
λ/h^{-1}	0.019	0.0031	1.3
η_1/h^{-1}	3.0	0	—
ν	0.20	0	—
$\xi_2/\mathrm{h\ mm}^{-1}$	2.8	0.48	0
α	0.23	0.17	3.5

Clearly, there are many possible choices for F and H. The choice made here is the 1 h mean, and the 1, 3, 6, 12 and 24 h variances, proportion of dry intervals, and wet spell transition probabilities. In many ways this choice is similar to that made by other authors. The exception is the use of the wet spell transition probability, which is used in preference to the autocorrelation because it tends to have a smaller sampling error.

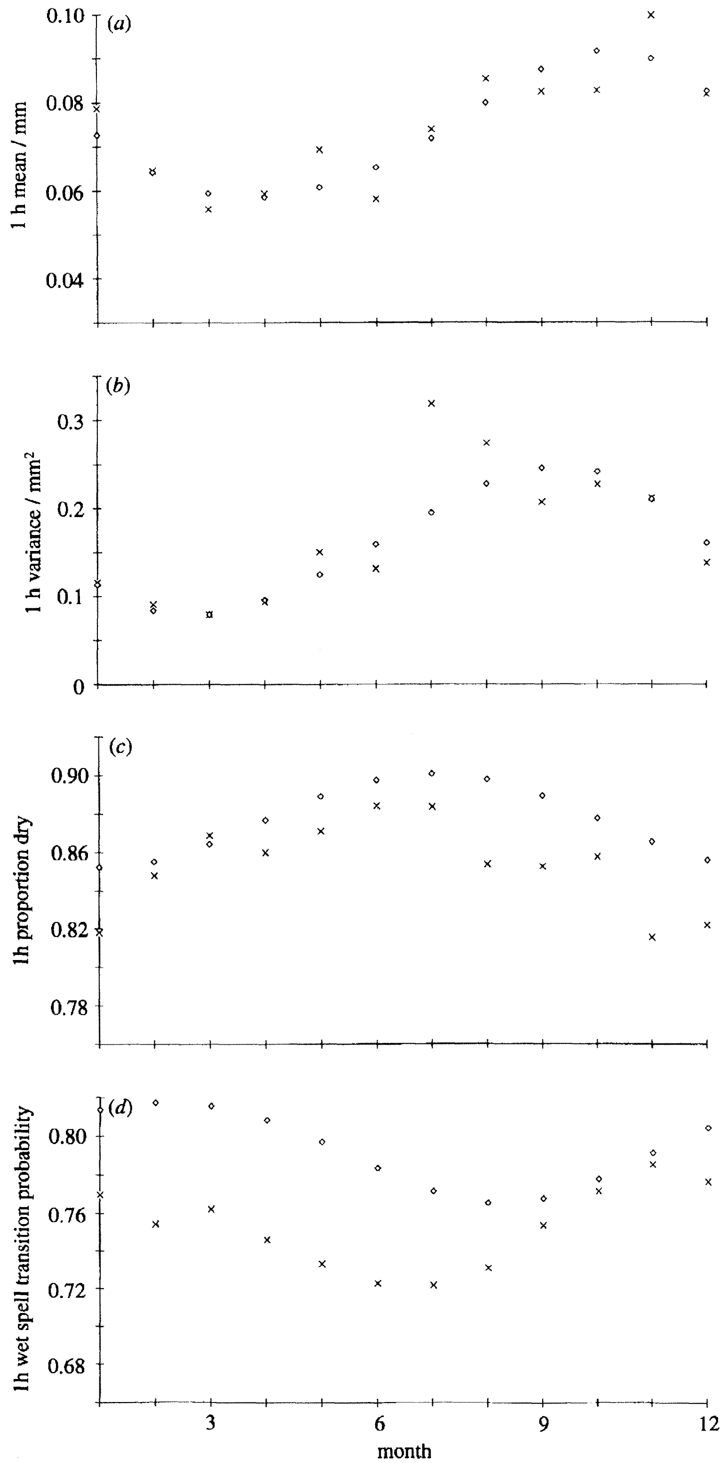

Figure 2. Historical ($\times$) and fitted ($\diamond$) properties at the 1 h and 24 h levels of aggregation against month. (*a*) 1 h mean; (*b*) 1 h variance; (*c*) 1 h proportion dry; (*d*) 1 h wet spell transition probability; (*e*) 24 h variance; (*f*) 24 h proportion dry; (*g*) 24 h wet spell transition probability.

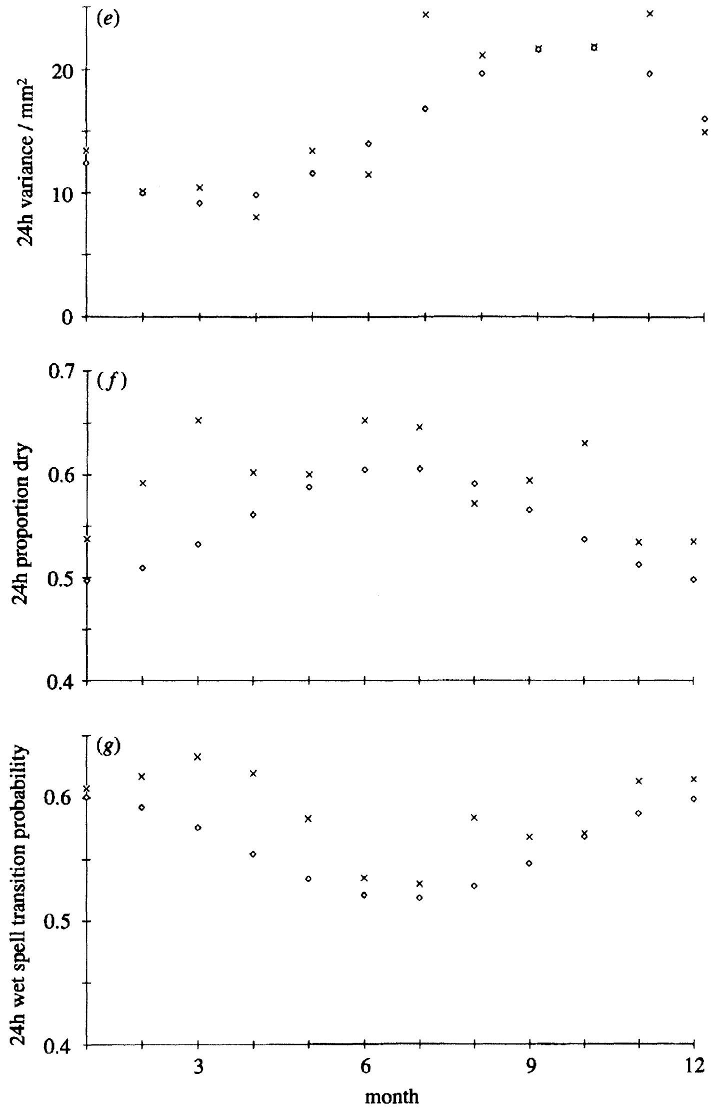

Figure 2. For legend see opposite.

(b) *Fitting to hourly rainfall data*

Rainfall data were available for a large number of sites scattered throughout the UK. A regression analysis, which first involved fitting the model to all available sites, was used to explore regional and seasonal variations in the parameter estimates (a detailed account of this will appear elsewhere). The analysis showed that it was reasonable to treat three of the parameters, β, η_2, and ξ_1, as constants. Consequently, by fixing the parameters, β, η_2, and ξ_1, at their mean values of 0.24 (h^{-1}), 0.53 (h^{-1}) and 0.26 (h mm^{-1}) respectively, a further reduction in the total number of estimates per site was possible.

Of the available data, the longest record of hourly data, which was a 30 year record taken from a site in Farnborough was selected to test the model. The harmonic

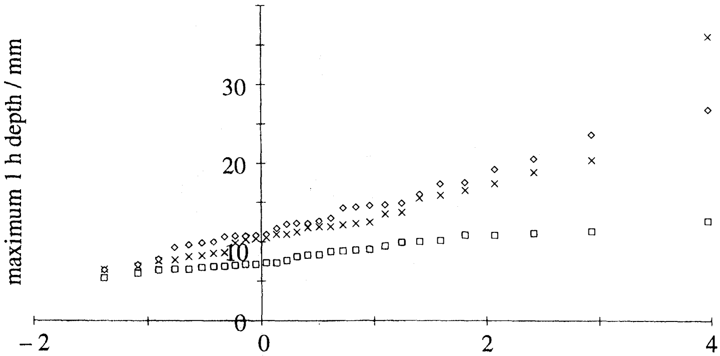

Figure 3. Comparison of historical (×) and simulated annual maximum 1 h depths. Simulated depths generated using the GNSRP(2) model (◇) and NSRP model (□).

variables m_ψ, A_ψ, θ_ψ for $\psi \in \{\lambda, \eta_1, \nu, \xi_2, \alpha\}$ were estimated for the site by minimizing (4.2) and are given in table 1. In addition, the parameter estimates of λ, ξ_2 and α were evaluated using (4.1) and are plotted against each month in figure 1.

The parameter estimates take physically realistic values. Over the summer months, the rate of storm origin arrivals (λ) decreases, the proportion of heavy cells (α) increases, and the expected intensity (ξ_2^{-1}) of the light cells increases (figure 1). Thus, the summer storms generated by the GNSRP(2) model will be heavier and contain more convective rainfall than the winter storms. This agrees with the physical interpretation that winter rainfall in the UK tends to be the result of persistant frontal systems, whereas summer rainfall tends to be the result of heavy convective storms.

Another feature of the plots is that the rate of storm origin arrivals reaches a minimum during midsummer, i.e. June and July (figure 1*a*), whereas the proportion of heavy cells reaches a maximum towards the end of summer, i.e. August and September (figure 1*c*). The physical interpretation of this is that summer thunderstorms, which are the result of rising air masses, tend to increase towards the end of summer after the land has been heated during the summer period.

The estimates for ν and η_1 were found to be constant throughout the year at values of 0.20 and 3.0 (h^{-1}) respectively, so that, on average, there will be 5.0 cells per storm with the heavy cells lasting for 0.33 h (table 1). It is worth noting that the estimates of the mean duration for the heavy and light cells, 0.33 h and 1.9 h respectively, are in broad agreement with the observational studies of Austin & Houze (1972), who found that convective cells generally last from a few minutes to approximately 0.5 h and tend to occur within small mesoscale areas of precipitation that last for approximately 2 h.

The fitted values at the 1, 3, 6, 12 and 24 h levels of aggregation were compared with the sample estimates taken from the historical data. Some lack of fit was evident, the worst case being a tendency for the model to overestimate the 1 h wet spell transition probabilities shown in figure 2*d*. The fitted values to the sample means and variances compared favourably and showed smooth seasonal variation across the year. Figure 2 illustrates the results at the 1 h and 24 h levels of aggregation. Overall, the fit of the model to the data seemed reasonable given the reduction in the total number of estimates per site obtained by using the harmonic relationships.

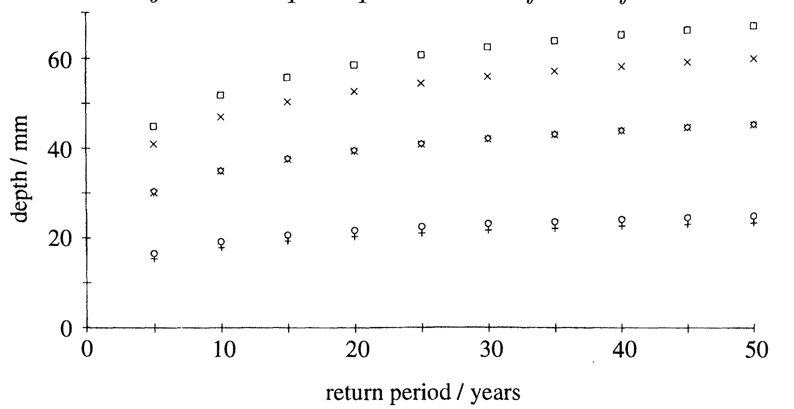

Figure 4. Predicted annual maximum depths. 1 h aggregation level: + historical, ○ simulated; 6 h aggregation level: ✱ historical, ◇ simulated; 24 h aggregation level: × historical, □ simulated.

(c) *Model validation*

The practical implication of any consistent error, such as the overestimation in the 1 h wet spell transition probability, should be assessed before the model is applied to problems in engineering hydrology. However, some confidence in the model can be gained by examining properties that are not used in the fitting procedure but which are of interest in applications, such as extreme values.

Using the parameter estimates in table 1, 30 years of hourly rainfall time series were simulated and the maximum total at the 1 h level of aggregation extracted from each year in the simulated record. These totals were then ordered and plotted against the standardized Gumbel variate, together with ordered annual maximum 1 h totals taken from the historical record (figure 3).

Figure 3 shows that the GNSRP(2) model has a good fit to the historical maximum 1 h totals. As the ordered maxima lay approximately on a straight line, it seemed appropriate to assume that they followed a Gumbel probability distribution and perform an extreme value analysis. Maximum likelihood estimates of the location and scale parameters of the Gumbel distribution were found for the annual maxima at the 1, 6 and 24 h aggregation levels for both the historical and simulated data. Rainfall depths at the 1, 6 and 24 h aggregation levels were then predicted for return periods in the range 5 to 50 years using the Gumbel distribution function with the estimated location and scale parameters (figure 4).

The simulated predicted depths show remarkably good agreement with the historical values at the 1 and 6 h levels of aggregation (figure 4). At the 24 h level of aggregation the simulated predicted depths overestimate the historical depths.

In some hydrological applications (e.g. storm sewer design), 'events' are extracted from a time series record and used to predict discharge. Thus, some properties of the distribution of the duration and depth of an event were considered in a further validation exercise. An event was defined to be a sequence of wet hours that had at least one dry hour immediately before and after the sequence. The duration and total depth were found for each event in the historical and simulated records. Only those events with total depths exceeding 5 mm were considered in the analysis, i.e. the heavier events, which are of more interest in engineering applications. Some properties of the distribution of the depth and duration of the events are given in tables 2–4.

Table 2. *Properties of the distribution of event duration/h*

	historical	GNSRP(2)	NSRP
number of events	1120	1171	1046
mean	8.1	8.7	6.1
standard deviation	4.6	4.3	2.8
75th percentile	10	12	7.0
90th percentile	14	15	10
95th percentile	17	17	12
99th percentile	24	22	15

Table 3. *Properties of the distribution of event depth/mm*

	historical	GNSRP(2)	NSRP
number of events	1120	1171	1046
mean	11	11	10
standard deviation	7.0	6.4	5.4
90th percentile	18	19	16
95th percentile	23	24	20
99th percentile	37	35	32

Table 4. *Correlation between depth and duration of the events*

historical	GNSRP(2)	NSRP
0.55	0.52	0.36

The model shows a good fit to the historical values (tables 2 and 3). In particular, the fit to the high percentiles is remarkably good, given that the parameters of the model relate to underlying physical features in the rainfall field and not directly to properties of extreme events. The fit to the correlation between the depth and duration is also good (table 4).

(*d*) *Comparison with original* NSRP *model*

The original NSRP model with the cell intensity also distributed as an independent exponential random variable was fitted to the same data using the same properties in the fitting procedure, and a time series of 30 years of hourly data was simulated. The ordered annual maximum 1 h totals were found and are also plotted in figure 3 where it can be seen that the original NSRP model substantially underestimates the historical maxima. Furthermore, the original NSRP model showed a poorer fit to the properties of the events when compared with the GNSRP(2) model (tables 2–4).

Alternative approaches could be used to improve the fit of the original NSRP model to the historical extreme values. For example, the cell intensity in the original NSRP model could be a random variable from a long-tailed distribution, such as the gamma or Weibull. However, this would ignore the physical arguments in § 1 and so may lead to a poorer fit in other properties.

5. Summary and conclusions

The generalized Neyman–Scott rectangular pulses (GNSRP) model is a physically realistic representation of rainfall at a single site because it allows there to be more than one cell type within the same storm.

The GNSRP model can also be used for multi-site rainfall, where different cell types can be allocated to one or more sites over a catchment area. One fairly simple approach is to assume that all the sites receive the same light rain cells simultaneously, with heavy convective cells being associated with individual sites. This approach may be applicable when the distances between pairs of sites are greater than the diameter of a convective cell, i.e. greater than about 4 km. However, when the distance between a pair of sites is smaller than the diameter of a convective cell, it would be appropriate to assume that a proportion of convective cells occur simultaneously at those sites.

When fitting the GNSRP(2) model to hourly data, the total number of estimates per parameter was reduced by allowing the parameter estimates to vary over the year according to harmonic relationships. This also produced a smooth seasonal variation in the parameter estimates and fitted moments. The estimates for the mean heavy and light cell duration for the GNSRP(2) model were 0.33 h and 1.9 h respectively, which were in broad agreement with the observational studies of Austin & Houze (1972). Some consistent lack of fit to the data remained; the worst case being the fit to the 1 h wet spell transition probability. Such errors need to be assessed for their practical implication before the model is applied to problems in engineering hydrology.

The GNSRP(2) model showed a good fit to properties not used in the fitting procedure. Overall, the GNSRP model was found to be superior to the original NSRP model, both in terms of relating the parameters to underlying physical features in the rainfall field and in the fit to extreme values.

Useful discussions with Enda O'Connell, Geoff Pegram and Valerie Isham are gratefully acknowledged. The work described in §4 was coordinated by the Water Research Centre, Swindon and funded by the UK Foundation for Water Research.

References

Austin, P. M. & Houze, R. A. 1972 Analysis of the structure of precipitation patterns in New England. *J. appl. Met.* **11**, 926–935.

Browning, K. A. 1985 Conceptual models for precipitation systems. *Met. Mag., Lond.* **114**, 293–319.

Cowpertwait, P. S. P. 1991 Further developments of the Neyman–Scott clustered point process for modeling rainfall. *Wat. Resour. Res.* **27**, 1431–1438.

Cox, D. R. & Isham, V. 1980 *Point processes.* London: Chapman & Hall.

Entekhabi, D., Rodriguez-Iturbe, I. & Eagleson, P. S. 1989 Probabilistic representation of the temporal rainfall process by a modified Neyman–Scott rectangular pulses model: Parameter estimation and validation. *Wat. Resour. Res.* **25**, 295–302.

Kavvas, M. L. & Herd, K. R. 1985 A radar-based stochastic model for short-time-increment rainfall. *Wat. Resour. Res.* **21**, 1437–1455.

Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. 1987*a* Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond.* A **410**, 269–288.

Rodriguez-Iturbe, I., Febres De Power, B. & Valdes, J. B. 1987*b* Rectangular pulses point process models for rainfall: analysis of empirical data. *J. geophys. Res.* **92**, 9645–9656.

Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. 1988 A point process model for rainfall: further developments. *Proc. R. Soc. Lond.* A **417**, 283–298.

Received 4 November 1993; accepted 23 February 1994