The Royal Society

A Poisson-cluster model of rainfall: high-order moments and extreme values

By Paul S. P. Cowpertwait

Department of Statistics, Massey University, Albany Campus, North Shore City, Auckland, New Zealand

Received 17 February 1997; accepted 9 July 1997

A conceptual stochastic model for rainfall, based on a Poisson-cluster process with rectangular pulses representing rain cells, is further developed. A method for deriving high-order moments is applied to obtain the third-moment function for the model. This is used with second-order properties to fit the model to January and July time-series taken from a site in Wellington, New Zealand. It is found that the parameter estimates may follow two solution paths converging on an optimum value over a bounded interval. The parameter estimates are used with the model to simulate 200 years of hourly data, and parametric tests used to compare simulated and observed extreme rainfalls. These show good agreement over a range of sampling intervals. The paper concludes with a discussion of the standard errors of the model parameter estimates which are obtained using a non-parametric bootstrap.

Keywords: time series; rain cells; skewness; extreme values; point process

1. Introduction

Models for rainfall based on Poisson-cluster processes have received considerable attention in the literature, starting with a discussion by LeCam (1961). Subsequently, Kavvas & Delleur (1981) introduced and fitted a Poisson-cluster process with instantaneous depths to data from Indiana, and this model was further developed by Rodriguez-Iturbe *et al.* (1984). Parameter estimation problems have been identified when fitting this model to data (see, for example, Foufoula-Georgiou & Guttorp 1987); in particular, it is clear that a model based on instantaneous rainfall depths does not have a compelling physical basis.

Rodriguez-Iturbe *et al.* (1987*a*) proposed a rectangular pulses Poisson-cluster model, which was further developed by Rodriguez-Iturbe *et al.* (1988). Empirical investigations have been reported by Rodriguez-Iturbe *et al.* (1987*b*), Onof & Wheater (1994), Islam *et al.* (1990) and Cowpertwait *et al.* (1996). A generalization of a Poisson-cluster model, allowing both convective and stratiform rain to occur within the same storm, was developed by Cowpertwait (1994).

While some of the previous work reported a good fit to extreme values, the evidence was not convincing because the standard error associated with the parameter estimates of extreme-value distributions was not accounted for, and this is expected to be large. Furthermore, it is clear that the empirical distribution, particularly the distribution tail, is unlikely to be consistently well fitted without some high-order property being included in the fitting procedure. In this paper we further develop

the model of Rodriguez-Iturbe *et al.* (1987*a*) by deriving the third moment function of the rainfall depth process. This function also applies to a special case (one cell type) of the spatial-temporal model developed by Cowpertwait (1995).

The paper is divided into three more sections as follows. In § 2 the notation is introduced and the model further developed, while § 3 reports the results of the empirical data analysis. Finally some overall conclusions and directions for future work are given in § 4.

2. Further developments

(*a*) *High-order aggregated moments*

Let $Y(t)$ be a random variable representing the rainfall intensity at time t, and $Y_h^{(i)}$ be the aggregated rainfall depth in the ith sampling interval of length h, so

$$Y_h^{(i)} = \int_{(i-1)h}^{ih} Y(t)\,\mathrm{d}t. \tag{2.1}$$

We will assume that the rainfall time series, $\{Y_h^{(i)} : i = 1, 2, \ldots\}$, is stationary so that $E\{(Y_h^{(i)})^n\} = E\{(Y_h^{(j)})^n\}$ for all $i, j = 1, 2, \ldots$. Without loss of generality, the superscripts i and j may be omitted, and the moments of $Y_h = \int_0^h Y(t)\,\mathrm{d}t$ considered. A general expression for the nth moment is then given by:

$$E(Y_h^n) = \int_0^h \int_0^h \cdots \int_0^h E\{Y(t_1)Y(t_2)\cdots Y(t_n)\}\,\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_n. \tag{2.2}$$

(*b*) *Properties of the Neyman–Scott and Bartlett–Lewis models*

Following Rodriguez-Iturbe *et al.* (1987*a*), suppose the arrival times of storm origins occur in a Poisson process with rate λ, each storm origin generating a random number C of cell origins. Let a rectangular pulse (L, X) be associated with each cell origin, where L and X are independent random variables corresponding to the lifetime and intensity of the pulse, respectively; the pulse representing a rain cell. Furthermore, let $X_{t-u}(u)$ be an independent random variable representing the rainfall intensity at time t due to a cell with starting time $t-u$, and let $\delta N(t) \equiv N(t, t+\delta)$ be the number of cell origins in the time interval $(t, t+\delta)$. The total intensity at time t, $Y(t)$, is the summation of the intensities of all cells alive at time t, and can be written as

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u)\,\mathrm{d}N(t-u). \tag{2.3}$$

Now, to evaluate (2.2), we have (from (2.3)):

$$\begin{aligned} &E\{Y(t_1)Y(t_2)\cdots Y(t_n)\} \\ &\quad = \int_{u_1=0}^{\infty}\int_{u_2=0}^{\infty}\cdots\int_{u_n=0}^{\infty} E\{X_{t_1-u_1}(u_1)X_{t_2-u_2}(u_2)\cdots X_{t_n-u_n}(u_n)\} \\ &\qquad \times E\{\mathrm{d}N(t_1-u_1)\,\mathrm{d}N(t_2-u_2)\cdots\mathrm{d}N(t_n-u_n)\}, \end{aligned} \tag{2.4}$$

because $X_{t-u}(u)$ and $\mathrm{d}N(t-u)$ are independent.

To evaluate (2.4) some assumptions need to be made about the location of cell origins (the starting times of the rain cells) relative to their storm origin and the distribution of cell lifetime L. Following Rodriguez-Iturbe *et al.* (1987), we assume

that L is exponentially distributed with mean η^{-1}. For a model based on a Neyman–Scott point process, we take the waiting time for a cell origin after a storm origin to be independently exponentially distributed with mean β^{-1}, no cell origin occurring at the storm origin. Alternatively, we could use a Bartlett–Lewis point process, in which case the waiting time between successive cell origins can be taken to be independently exponentially distributed with mean κ^{-1}, the process of cell origins terminating after a random time that is exponentially distributed. However, we note here that under these conditions, both the Neyman–Scott and Bartlett–Lewis models have equivalent second-order properties, provided the same distribution is used for C. These properties, which essentially follow from (2.4) and (2.2), can be written as

$$\mu_h = E(Y_h) = \lambda\mu_C\mu_X h/\eta, \tag{2.5}$$

$$\begin{aligned}\gamma_{h,l} = \mathrm{Cov}\{Y_h^{(i)}, Y_h^{(i+l)}\} = \lambda\eta^{-3}A(h,l)\{2\mu_C E(X^2) + \mu_X^2\beta^2 E(C^2-C)/(\beta^2-\eta^2)\}\\ -\lambda\mu_X^2 B(h,l)E(C^2-C)/\{\beta(\beta^2-\eta^2)\},\end{aligned} \tag{2.6}$$

where $A(h,0) = (h\eta + \mathrm{e}^{-\eta h} - 1)$, $B(h,0) = (h\beta + \mathrm{e}^{-\beta h} - 1)$; and, for l a positive integer, $A(h,l) = \frac{1}{2}(1-\mathrm{e}^{-\eta h})^2\mathrm{e}^{-\eta h(l-1)}$ and $B(h,l) = \frac{1}{2}(1-\mathrm{e}^{-\beta h})^2\mathrm{e}^{-\beta h(l-1)}$. For the Bartlett–Lewis process, with no cell origin at the storm origin and $C+1$ distributed geometrically, we have $E\{C^2-C\} = 2\mu_C^2$ and $\kappa = \mu_C\beta$ in (2.6). If the Bartlett–Lewis process is to have the first cell origin occurring at the storm origin, then C is a geometric random variable and $E\{C(C-1)\} = 2\mu_C(\mu_C-1)$ and $\kappa = (\mu_C-1)\beta$ in (2.6).

(*c*) *The third moment function*

To evaluate a third moment we consider the Neyman–Scott process, although the same method could be applied to the Bartlett–Lewis process.

Suppose cell origins occur in a Neyman–Scott point process, and let $t_1 < t_2 < t_3$ be the starting times of three cells chosen at random, i.e. the positions of three cell origins. Then, extending the method described by Cox & Isham (1980, p. 78), the three cells must either: (i) all belong to different storm origins; (ii) all belong to the same storm origin; or (iii) one cell belongs to a different storm origin and two cells belong to the same storm origin. Hence, in the case of the Neyman–Scott process,

$$\begin{aligned}E\{\delta N(t_1)\delta N(t_2)\delta N(t_3)\} &= \lambda^3\mu_C^3\delta_1\delta_2\delta_3\\ &\quad + \tfrac{1}{2}\lambda^2\beta\mu_C E(C^2-C)\{\mathrm{e}^{-\beta(t_3-t_1)} + \mathrm{e}^{-\beta(t_3-t_2)} + \mathrm{e}^{-\beta(t_2-t_1)}\}\delta_1\delta_2\delta_3\\ &\quad + \tfrac{1}{3}\lambda\beta^2 E\{C(C-1)(C-2)\}\mathrm{e}^{-\beta(t_3+t_2-2t_1)}\delta_1\delta_2\delta_3 + o(\delta_1\delta_2\delta_3).\end{aligned} \tag{2.7}$$

Now $X_{t_1-u_1}(u_1)$ and $X_{t_1+\tau_1-u_2}(u_2)$ are independent if, and only if, $t_1 - u_1 \neq t_1 + \tau_1 - u_2$, i.e. $u_2 \neq \tau_1 + u_1$. Hence,

$$\begin{aligned}&E\{X_{t_1-u_1}(u_1)X_{t_1+\tau_1-u_2}(u_2)X_{t_1+\tau_1+\tau_2-u_3}(u_3)\}\\ &= \begin{cases} E(X^3)\mathrm{e}^{-\eta(u_1+\tau_1+\tau_2)}, & \text{when } t_1-u_1 = t_1+\tau_1-u_2 = t_1+\tau_1+\tau_2-u_3,\\ \mu_X E(X^2)\mathrm{e}^{-\eta(u_1+u_2+\tau_2)}, & \text{when } u_3 = u_2+\tau_2, \quad u_2 \neq u_1+\tau_1,\\ \mu_X E(X^2)\mathrm{e}^{-\eta(u_1+u_2+\tau_1+\tau_2)}, & \text{when } u_3 = u_1+\tau_1+\tau_2, \quad u_2 \neq u_1+\tau_1,\\ \mu_X E(X^2)\mathrm{e}^{-\eta(u_1+u_3+\tau_1)}, & \text{when } u_2 = u_1+\tau_1, \quad u_3 \neq u_1+\tau_1+\tau_2,\\ \mu_X^3\mathrm{e}^{-\eta(u_1+u_2+u_3)}, & \text{when } u_2 \neq u_1+\tau_1, \quad u_3 \neq u_1+\tau_1+\tau_2,\\ 0, & \text{otherwise,}\end{cases}\end{aligned} \tag{2.8}$$

where $\tau_1, \tau_2 > 0$. Letting $t_2 = t_1 + \tau_1$ and $t_3 = t_2 + \tau_2$ in (2.7) followed by evaluating (2.4) and (2.2) with $n = 3$, gives, after considerable algebra, the third (central) moment for the Neyman–Scott model:

$$\begin{aligned}\xi_h = E[\{Y_h - E(Y_h)\}^3] &= 6\lambda\mu_C E(X^3)(\eta h - 2 + \eta h \mathrm{e}^{-\eta h} + 2\mathrm{e}^{-\eta h})/\eta^4 \\ &+3\lambda\mu_X E(X^2)E\{C(C-1)\}f(\eta,\beta,h)/\{2\eta^4\beta(\beta^2-\eta^2)^2\} \\ &+\lambda\mu_X^3 E\{C(C-1)(C-2)\}g(\eta,\beta,h)/\{2\eta^4\beta(\eta^2-\beta^2)(\eta-\beta)(2\beta+\eta)(\beta+2\eta)\},\end{aligned} \tag{2.9}$$

where the functions $f(\eta, \beta, h)$ and $g(\eta, \beta, h)$ are listed below.

$$\begin{aligned}f(\eta,\beta,h) = &-2\eta^3\beta^2\mathrm{e}^{-\eta h} - 2\eta^3\beta^2\mathrm{e}^{-\beta h} + \eta^2\beta^3\mathrm{e}^{-2\eta h} + 2\eta^4\beta\mathrm{e}^{-\eta h} + 2\eta^4\beta\mathrm{e}^{-\beta h} \\ &+2\eta^3\beta^2\mathrm{e}^{-(\eta+\beta)h} - 2\eta^4\beta\mathrm{e}^{-(\eta+\beta)h} - 8\eta^3\beta^3 h + 11\eta^2\beta^3 - 2\eta^4\beta \\ &+2\eta^3\beta^2 + 4\eta\beta^5 h + 4\eta^5\beta h - 7\beta^5 - 4\eta^5 + 8\beta^5\mathrm{e}^{-\eta h} - \beta^5\mathrm{e}^{-2\eta h} \\ &-2h\eta^3\beta^3\mathrm{e}^{-\eta h} - 12\eta^2\beta^3\mathrm{e}^{-\eta h} + 2h\eta\beta^5\mathrm{e}^{-\eta h} + 4\eta^5\mathrm{e}^{-\beta h},\end{aligned} \tag{2.10}$$

$$\begin{aligned}g(\eta,\beta,h) = &\,12\eta^5\beta\mathrm{e}^{-\beta h} + 9\eta^4\beta^2 + 12\eta\beta^5\mathrm{e}^{-\eta h} + 9\eta^2\beta^4 + 12\eta^3\beta^3\mathrm{e}^{-(\eta+\beta)h} \\ &-\eta^2\beta^4\mathrm{e}^{-2\eta h} - 12\eta^3\beta^3\mathrm{e}^{-\beta h} - 9\eta^5\beta - 9\eta\beta^5 - 3\eta\beta^5\mathrm{e}^{-2\eta h} \\ &-\eta^4\beta^2\mathrm{e}^{-2\beta h} - 12\eta^3\beta^3\mathrm{e}^{-\eta h} + 6\eta^5\beta^2 h - 10\beta^4\eta^3 h + 6\beta^5\eta^2 h \\ &-10\beta^3\eta^4 h + 4\beta^6\eta h - 8\beta^2\eta^4\mathrm{e}^{-\beta h} + 4\beta\eta^6 h + 12\beta^3\eta^3 \\ &-8\beta^4\eta^2\mathrm{e}^{-\eta h} - 6\eta^6 - 6\beta^6 - 2\eta^6\mathrm{e}^{-2\beta h} - 2\beta^6\mathrm{e}^{-2\eta h} \\ &+8\eta^6\mathrm{e}^{-\beta h} + 8\beta^6\mathrm{e}^{-\eta h} - 3\beta\eta^5\mathrm{e}^{-2\beta h}.\end{aligned} \tag{2.11}$$

When $C \equiv 1$ the model reduces to the special case of a Poisson rectangular pulses process (without clustering), the third moment of which is given by the first term in (2.9). The first term in (2.9) will also appear in the third central moment for the Bartlett–Lewis model. However, (2.10) and (2.11) will differ for the Bartlett–Lewis model; in particular, the substitutions $\kappa = \mu_C\beta$ and $\kappa = (\mu_C - 1)\beta$ cannot be made to obtain the third moment for the Bartlett–Lewis process and evaluation must proceed from first principles.

3. Empirical analysis

(*a*) *Objective*

Our objective in this empirical study was to assess goodness-of-fit to properties up to third order and extreme values. A further related objective was to note the potential lack-of-fit to high-order properties that could result from omitting the third moment function from the fitting procedure.

We are not proposing that the fitting procedure used here is necessarily the best over all possible fitting procedures. However, it seems reasonable to postulate that the method should give a better fit to the tail of the empirical distribution than any method based only on second-order properties, and this includes some approximate likelihood methods (e.g. those based on the spectral density function (see Chandler 1997)).

(*b*) *Model parameters*

We first need to assign distributions to the cell intensity (X) and the number of cells (C) in a storm. To ensure at least one cell occurs in each storm, C was taken to

be a geometric random variable with mean μ_C, so that $\mathrm{pr}(C = i) = (1 - \mu_C^{-1})^{i-1}\mu_C^{-1}$ and $E\{C(C-1)\} = 2\mu_C(\mu_C - 1)$ and $E\{C(C-1)(C-2)\} = 6\mu_C(\mu_C - 1)^2$ in equations (2.6) and (2.9).

The gamma distribution with parameters (α, θ) was selected for the random variable X, which must take positive values, so that X has probability density function $\theta^{-\alpha}x^{\alpha-1}\mathrm{e}^{-x/\theta}/\Gamma(\alpha)$, $x \geqslant 0$. The moments of X, which appear in equations (2.5), (2.6) and (2.9), are then given by $E(X^r) = \theta^r\Gamma(\alpha + r)/\Gamma(\alpha)$.

A slightly more parsimonious approach would be to use an exponential distribution for X. However, past studies have reported a lack-of-fit to extreme values under this distribution, and so it is not considered further here.

The model thus has the six unknown parameters: λ, β, η, μ_C, α and θ. To estimate these parameters requires at least six equations:

$$z_i(\lambda, \beta, \eta, \mu_C, \alpha, \theta) = \hat{z}_i, \quad i = 1, \ldots, 6, \tag{3.1}$$

where z_i is a moment or autocovariance function of the Poisson cluster model (given by equations (2.5), (2.6) or (2.9)), and $\hat{z}_i$ is the equivalent sample estimate taken from the observed data.

By rewriting equation (2.5) to express one of the parameters in terms of the mean μ_1 and the remaining parameters, the problem immediately reduces to solving five simultaneous equations. The choice of which parameter to make the subject of the equation in the mean is arbitrary and in this study θ was selected.

It is almost impossible to obtain an exact solution to (3.1), because the functions are highly non-linear in the parameters. However, a 'near' solution can be obtained by minimizing:

$$Z = \sum_{i=1}^{5}(1 - z_i/\hat{z}_i)^2, \tag{3.2}$$

subject to the estimates lying in some feasible region: $0.0001 < \lambda < 0.05$ (h^{-1}); $0.05 < \beta < 0.99$ (h^{-1}); $0.5 < \eta < 60$ (h^{-1}); $1 \leqslant \mu_C < 500$; $0.01 < \alpha < 20$.

The optimization has to be constrained because the parameters must all be positive. Furthermore, it would be unreasonable to allow $\beta < \lambda$, because this would imply that cell origins are more highly dispersed than storm origins. By constraining the parameters to lie within upper and lower bounds we can further ensure that the estimates do not enter physically unrealistic parameter spaces, although some judgement is required to set these bounds. However, the above feasible region can be justified on the basis that the bounds are wide and cell parameters estimated directly from spatial data lie within them (see, for example, Mellor & O'Connell 1996).

Clearly there are many functions and sample estimates that could be used in (3.2). However, the parameters can be divided into two sets: 'storm parameters' $\{\lambda, \beta, \mu_C\}$, which relate to properties of a storm, and 'cell parameters' $\{\eta, \alpha, \theta\}$, which relate to properties of a cell. To estimate the cell parameters we need to include functions taken over intervals comparable to the lifetime of a cell (e.g. functions of Y_1); whilst to estimate the storm parameters we need functions taken over intervals comparable to the lifetime of a storm (e.g. functions of Y_{24}). The following functions were therefore selected: (i) the 1 h variance $(\gamma_{1,0})$, third central moment (ξ_1), and lag 1 autocovariance function $(\gamma_{1,1})$; (ii) the 24 h variance $(\gamma_{24,0})$ and lag 1 autocovariance function $(\gamma_{24,1})$.

We denote the above set of functions as 'FIT 1'. By replacing ξ_1 with $\gamma_{12,0}$, a second fit 'FIT 2' was considered to determine whether the introduction of the third

Table 1. *Summary statistics for Wellington data (1955–1995): observed standard errors (in parentheses) and fitted values*

(The standard errors were found using a non-parametric bootstrap by resampling whole years (with replacement) selected at random from the 41 years of data; the estimate is based on a sample of 500 evaluations.)

statistic	month	observed	FIT 1	FIT 2
1 h mean (mm)	January	0.106 (0.011)	0.106	0.106
μ_1	July	0.194 (0.011)	0.193	0.194
1 h variance (mm^2)	January	0.441 (0.070)	0.443	0.442
$\gamma_{1,0}$	July	0.539 (0.046)	0.537	0.538
1 h skewness	January	11.7 (0.74)	11.6	8.84
$\xi_1/\gamma_{1,0}^{3/2}$	July	6.92 (0.36)	6.92	5.25
1 h correlation	January	0.620 (0.022)	0.621	0.623
$\gamma_{1,1}/\gamma_{1,0}$	July	0.690 (0.014)	0.689	0.702
24 h variance (mm^2)	January	61.8 (13.8)	62.1	62.8
$\gamma_{24,0}$	July	80.2 (9.1)	80.0	83.3
24 h correlation	January	0.290 (0.041)	0.290	0.285
$\gamma_{24,1}/\gamma_{24,0}$	July	0.273 (0.030)	0.273	0.262

moment is likely to improve the goodness-of-fit to observed high-order properties, particularly extreme values.

The replacement $\gamma_{12,0}$ is needed to avoid equations (3.1) becoming degenerate. This choice is again quite arbitrary, and moments taken at other aggregation levels (e.g. $\gamma_{2,0}$) may give a better fit to the high-order properties. It is beyond the scope of this study to consider the many permutations of second-order properties that may potentially give a better fit to the high-order values; the results are thus only a broad indication of the lack-of-fit that might be obtained if ξ_1 is omitted.

(*c*) *Data and fitted model*

The New Zealand National Institute of Water and Atmospheric Research (NIWA) provided their longest complete record of hourly data for use in this study. This was a 41-year record taken from a site in Wellington for the period 1955 to 1995. January and July were selected to be representative months for summer and winter in Wellington. All available January data (a sample of $41\times31\times24$ hours) were pooled and sample estimates of the moments evaluated. The same procedure was carried out for July (see table 1).

For each of the two months, the parameters were estimated by minimizing (3.2) using the simplex method developed by Nelder & Mead (1965). The algorithm was terminated when the standard deviation of the function values of Z at each point in the simplex was less than 0.0001 (see, for example, O'Neil (1985) for details). For each month, 10 different starting values were tried; these were randomly taken from a uniform distribution between the constraints on the parameters. The smallest optimum of the 10 runs was selected as the best fit, although this tended to be close

Table 2. *Parameter estimates*

	FIT 1		FIT 2	
parameter	January	July	January	July
λ	0.00591	0.0162	0.00514	0.0135
β	0.0647	0.0646	0.0677	0.0717
μ_C	22.4	17.8	6.15	5.01
η	1.15	0.835	1.17	0.830
α	0.265	0.258	9.43	9.96
θ	3.49	2.17	0.417	0.238

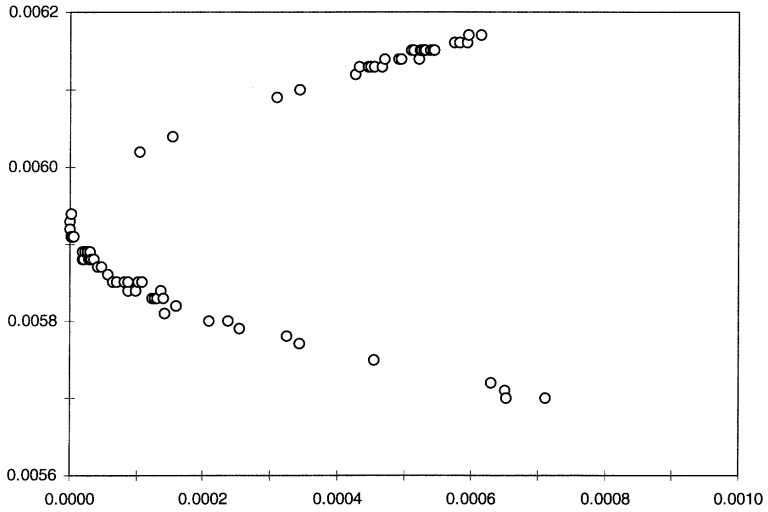

Figure 1. Estimates of λ against Z. Parameter estimates, on the y-axis, plotted against their resulting sum of squares Z (equation (3.2)), on the x-axis, obtained from starting values randomly selected over the bounded intervals.

to the other best fitting values (figures 1–6 are shown as an example for FIT 1–January). The parameter estimates and fitted values are given in tables 1 and 2.

The estimates tend to follow two solutions paths which converge to an optimum value within the feasible region. Figures 1–6 illustrate the solution paths that the estimates follow and were obtained by taking 100 random starting values over the bounded intervals; only those estimates which had $Z < 0.001$ are plotted in the figures. Thus, even though highly non-linear functions (particularly the third moment function) are used in the estimation procedure the resulting solution is almost unique over the bounded interval. A refinement of the estimation procedure would be to terminate the algorithm when Z is small, rather than when the standard deviation of the evaluated simplex points is small. However, the latter approach can be used to generate the figures and gain confidence in the resulting optimum solution.

As expected, a good fit is obtained to those sample estimates used in the optimization; all fitted values are within one standard error of the observed sample estimates. However, it is clear from table 1 that FIT 2 under-estimates the observed 1 h skew-

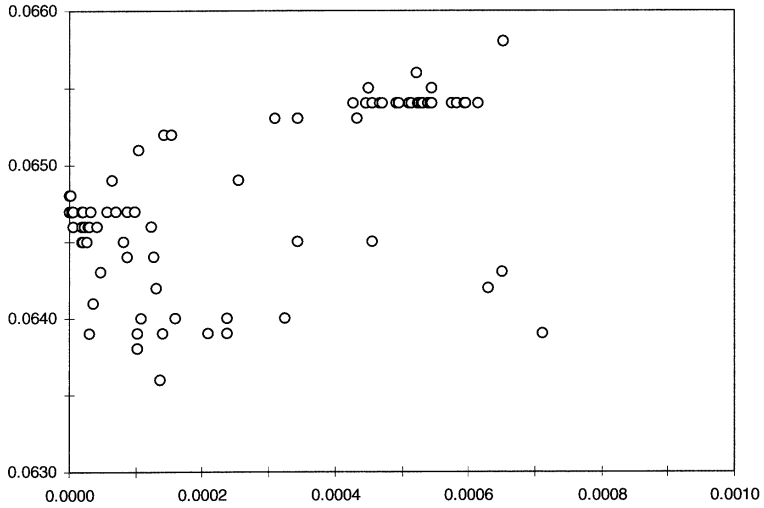

Figure 2. Estimates of β against Z. Parameter estimates, on the y-axis, plotted against their resulting sum of squares Z (equation (3.2)), on the x-axis, obtained from starting values randomly selected over the bounded intervals.

ness by more than two standard errors, providing evidence to support the use of the third moment in fitting the model.

The parameter estimates in table 2, suggest that application of the third moment results in a model with more expected cells per storm μ_C, with each cell having a lower expected intensity ($\alpha\theta$). The estimates of λ, β and η are close for both FIT 1 and FIT 2, suggesting that these parameters are well determined by the second-order properties.

(*d*) *Extreme value analysis*

Fisher & Tippett (1928) showed that under some fairly general conditions the distribution of the maxima Q in a random sample tends to a type I extreme value (Gumbel) distribution taking the form

$$\mathrm{pr}\{Q < q\} = \exp\{-\mathrm{e}^{-(q-\nu)/\varphi}\}, \quad -\infty < q < \infty. \tag{3.3}$$

Plots of the ordered maxima against the standardized Gumbel variate for a range of aggregation levels were considered, and confirmed that the Gumbel distribution was a reasonable choice for the maximum rainfalls. Consequently, to test goodness-of-fit to the extremes, it seemed appropriate to perform parametric tests on fitted Gumbel distributions.

For each of the parameter sets, 200 years of hourly data were simulated. Data were considered over a range of aggregation levels, from 1 h up to 192 h. Maximum likelihood estimates of the Gumbel parameters (ν, φ) and their asymptotic standard errors were found and are given in tables 3 and 4.

From tables 3 and 4, it is clear that the FIT 1 Gumbel parameter estimates compare favourably to the observed estimates (within one standard error of the observed values), whilst greater discrepancies are evident for FIT 2, particularly for the July estimates. The point estimates of the Gumbel parameters clearly depend on the point estimates of the model parameters and consequently on the fitted skewness. Whilst these comparisons are not all independent, they do provide convincing evi-

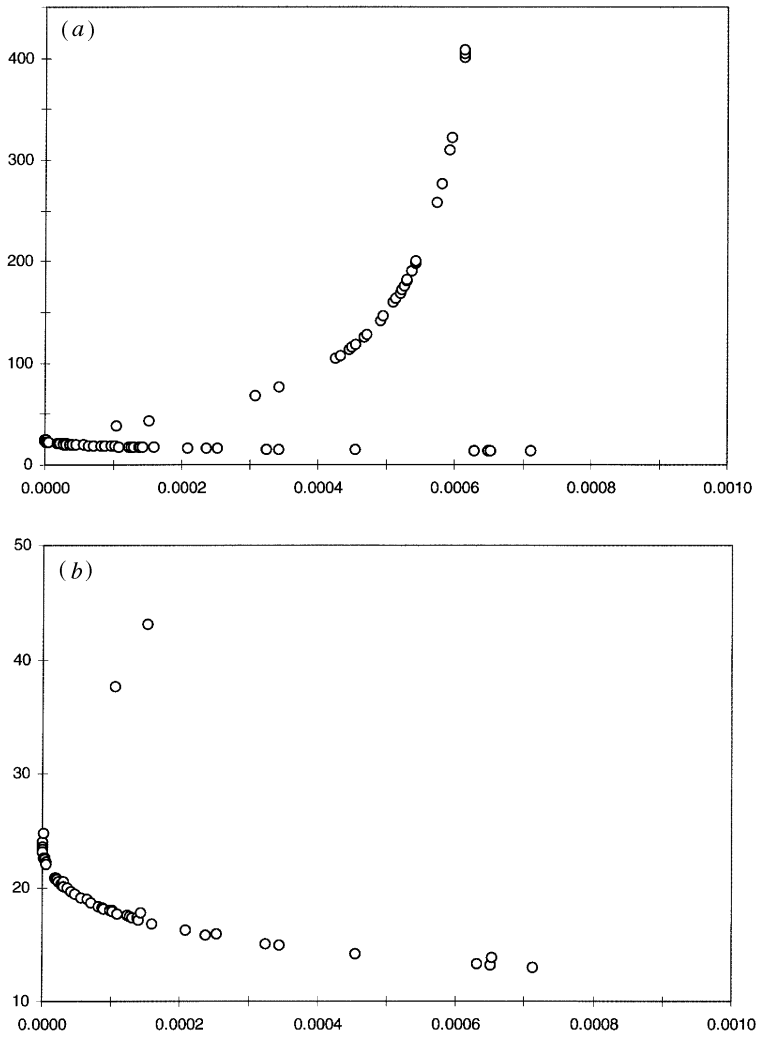

Figure 3. (*a*) Estimates of μ_C against Z. (*b*) Estimates of μ_C against Z (restricted values of μ_C). Parameter estimates, on the y-axis, plotted against their resulting sum of squares Z (equation (3.2)), on the x-axis, obtained from starting values randomly selected over the bounded intervals.

dence that including the third moment in the fitting procedure provides a good fit to the observed extreme values over a range of sampling intervals (tables 3 and 4).

(*e*) *Fit to dry periods*

Using the probability of a dry interval in fitting the model to data can lead to bias, because the estimate of the proportion of dry intervals is sensitive to the accuracy of the data measurements and the depths assigned to trace values. However, we note in this section that a reasonable fit to the proportion of intervals below a small threshold is obtained for FIT 1 even though this function is not used in the fitting procedure.

Sample estimates of 'dry' probabilities ($\mathrm{pr}\{Y_{24} < \delta\}$ for small δ) are given in table 5. Note that the sample estimates of $\mathrm{pr}\{Y_{24} < 0.001\}$ and $\mathrm{pr}\{Y_{24} < 0.1\}$ are

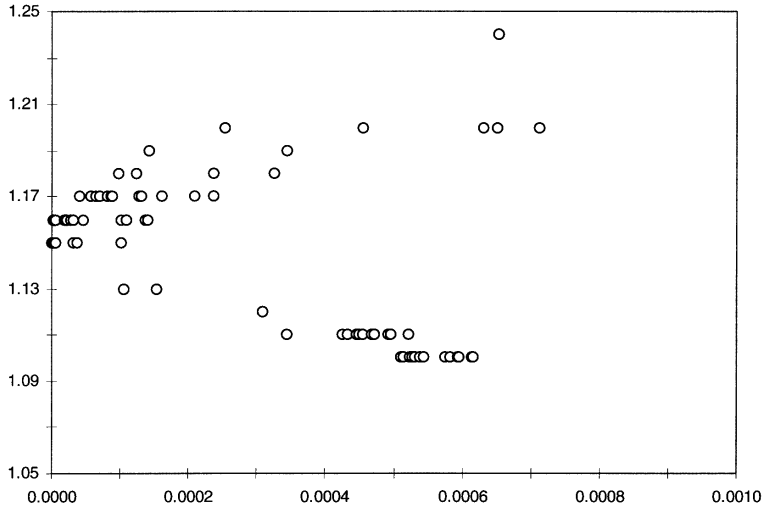

Figure 4. Estimates of η against Z. Parameter estimates, on the y-axis, plotted against their resulting sum of squares Z (equation (3.2)), on the x-axis, obtained from starting values randomly selected over the bounded intervals.

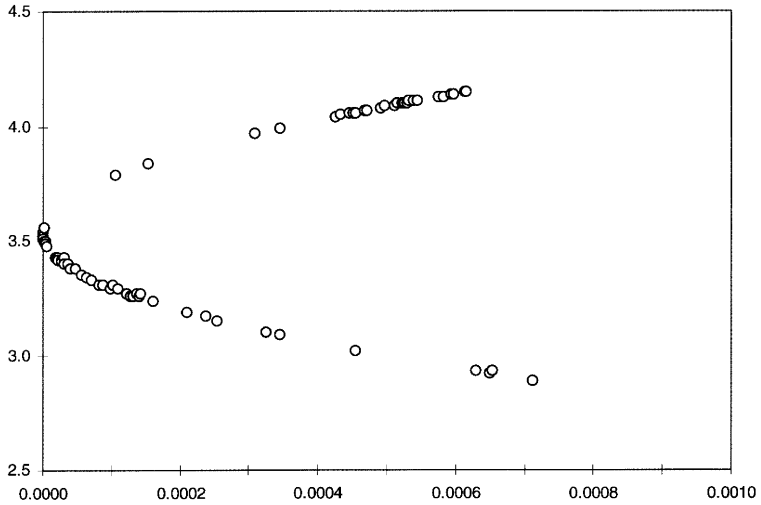

Figure 5. Estimates of θ against Z. Parameter estimates, on the y-axis, plotted against their resulting sum of squares Z (equation (3.2)), on the x-axis, obtained from starting values randomly selected over the bounded intervals.

all the same for the observed data, because these data are recorded to the nearest 0.1 mm. The FIT 1 and FIT 2 sample estimates of $\mathrm{pr}\{Y_{24} < \delta\}$ get progressively closer to the observed sample estimates as δ increases; again FIT 1 gives better results when compared with FIT 2.

(*f*) *Bootstrap standard errors*

To assess the standard errors of the parameter estimates, FIT 1 for January is considered. The same method was also applied to FIT 1–July and gave slightly smaller standard errors.

A non-parametric bootstrap standard error was obtained by resampling (with replacement) whole years from the 41 years of observed data and fitting the stochastic

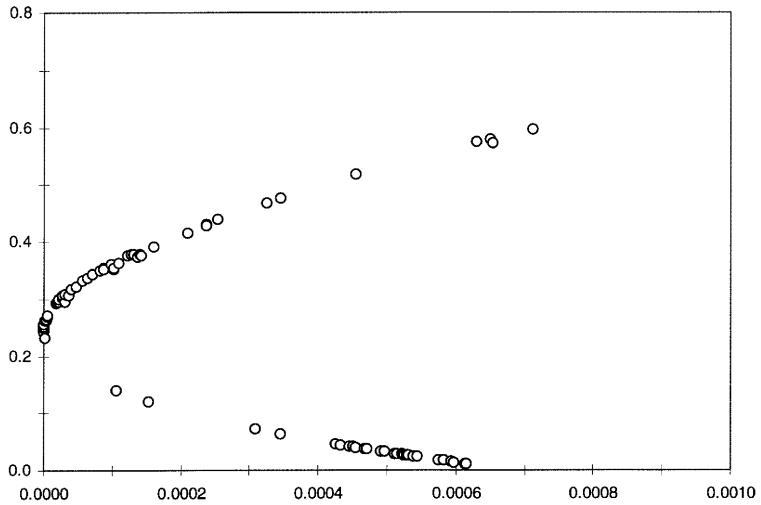

Figure 6. Estimates of α against Z. Parameter estimates, on the y-axis, plotted against their resulting sum of squares Z (equation (3.2)), on the x-axis, obtained from starting values randomly selected over the bounded intervals.

Table 3. *Gumbel parameter estimates and standard errors (in parentheses) for January*

aggregation level (h)	observed		FIT 1		FIT 2	
	ν	φ	ν	φ	ν	φ
1	6.44 (0.59)	3.60 (0.44)	6.04 (0.26)	3.43 (0.19)	5.66 (0.22)	2.92 (0.16)
2	8.93 (0.85)	5.17 (0.63)	8.85 (0.40)	5.36 (0.30)	8.51 (0.36)	4.77 (0.26)
6	13.4 (1.4)	8.77 (1.1)	13.3 (0.65)	8.66 (0.48)	13.3 (0.63)	8.45 (0.47)
12	17.0 (1.8)	11.2 (1.4)	16.7 (0.84)	11.3 (0.62)	17.0 (0.85)	11.5 (0.63)
24	20.9 (2.4)	14.8 (1.8)	20.8 (1.1)	14.3 (0.79)	21.0 (1.1)	14.7 (0.81)
48	25.0 (3.0)	18.5 (2.3)	24.1 (1.3)	17.5 (0.96)	25.1 (1.4)	18.2 (1.0)
96	26.7 (3.5)	21.1 (2.6)	26.2 (1.5)	20.7 (1.1)	29.0 (1.6)	22.0 (1.2)
192	30.8 (4.1)	25.2 (3.1)	28.3 (1.8)	24.3 (1.3)	30.5 (2.0)	26.2 (1.4)

model to the resampled data. This was repeated 100 times giving a random sample of 100 estimates for each model parameter.

The bootstrap coefficients of variations (CVs) for the estimates of λ, β and η were found to be 15%, 29% and 15%, respectively, whilst the CVs for the estimates of μ_C, α and θ were 65%, 99% and 16%. An informal comparison of the CVs for μ_C and α with the standard errors of the sample moments in table 1 suggests that these errors are larger than can be attributed by the sampling variation in the data. Furthermore, some of the estimates are related as illustrated in figure 7.

Figure 7 suggests that the fitted properties do not clearly distinguish between storms that have large μ_C and small μ_X, with those that have small μ_C and large μ_X. The high CVs and non-linear relation between the estimates raise doubts about the parametrization of the model. In the case of an empirical statistical model (e.g. a Markov chain), these results would suggest a reduction in model parameters was

Table 4. *Gumbel parameter estimates and standard errors (in parentheses) for July*

aggregation level (h)	observed		FIT 1		FIT 2	
	ν	φ	ν	φ	ν	φ
1	6.60 (0.44)	2.68 (0.33)	6.33 (0.17)	2.32 (0.13)	5.50 (0.12)	1.65 (0.091)
2	10.3 (0.70)	4.24 (0.52)	10.2 (0.31)	4.13 (0.23)	8.93 (0.21)	2.80 (0.15)
6	17.5 (1.2)	7.46 (0.91)	16.9 (0.55)	7.45 (0.41)	16.0 (0.45)	6.01 (0.33)
12	22.9 (1.7)	10.6 (1.3)	21.8 (0.76)	10.2 (0.56)	21.0 (0.61)	8.23 (0.45)
24	27.4 (2.1)	12.9 (1.6)	27.9 (1.0)	13.7 (0.75)	27.3 (0.89)	11.9 (0.66)
48	35.6 (2.7)	16.3 (2.0)	34.8 (1.3)	16.9 (0.93)	33.8 (1.1)	15.0 (0.83)
96	43.2 (3.8)	23.1 (2.8)	42.1 (1.6)	21.6 (1.2)	40.0 (1.4)	18.9 (1.0)
192	51.7 (5.0)	30.3 (3.7)	51.0 (2.2)	29.7 (1.6)	46.9 (1.8)	24.8 (1.4)

Table 5. *Observed and simulated estimates for the proportion of 'dry' days*

(The simulated data were recorded to three significant figures, whilst the observed data are recorded to one decimal place. An approximate bootstrap standard error for these estimates is 0.02.)

dry probability	month	observed	FIT 1	FIT 2
$\mathrm{pr}\{Y_{24} < 0.001\}$	January	0.691	0.686	0.760
	July	0.432	0.353	0.546
$\mathrm{pr}\{Y_{24} < 0.01\}$	January	0.691	0.696	0.760
	July	0.432	0.371	0.547
$\mathrm{pr}\{Y_{24} < 0.1\}$	January	0.691	0.720	0.764
	July	0.432	0.414	0.553
$\mathrm{pr}\{Y_{24} \leqslant 1\}$	January	0.777	0.784	0.786
	July	0.561	0.529	0.602
$\mathrm{pr}\{Y_{24} \leqslant 2\}$	January	0.814	0.816	0.805
	July	0.626	0.599	0.646

needed. Conversely, for a deterministic model the large standard errors would suggest that the fitting procedure does not adequately determine these parameters. However, the model is conceptual–stochastic and has parameters that relate to cells that occur in rainfall fields. In the rainfall process, cells may overlap and merge (see, for example, Browning 1985) and consequently can be difficult to identify, or separate out, even with fine resolution radar data. Hence, the large errors for μ_C and μ_X, and the relation between these parameters, are probably due to the difficulty in determining whether an event was the result of a single large cell or of several smaller overlapping or merging cells. Hence, from the physical standpoint, the present model parametrization could be retained, particularly as the special case of no-clustering (i.e. the Poisson rectangular pulses model for $\mu_C = 1$) did not occur in any of the 100 estimates, i.e. the Neyman–Scott parametrization was always required to provide an adequate fit to the resampled data (figure 7).

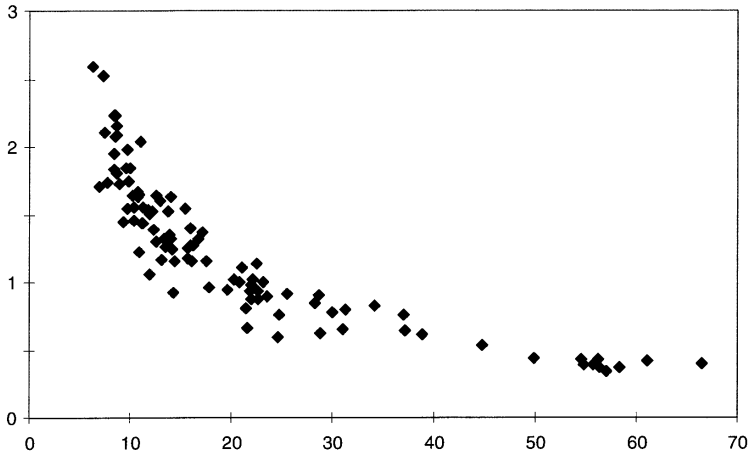

Figure 7. Estimates of μ_X against μ_C.

4. Conclusion

A method was given for deriving high-order moments of a stochastic Poisson cluster model of rainfall and applied to obtain the third central moment. When this was included in fitting the model to data taken from a site in Wellington, New Zealand, a good fit to the observed extreme values, over a range of time scales, was found. Lack-of-fit was evident when the third moment was excluded from the fitting procedure. Consequently, the derivation of the third moment function seems well justified and the function should prove useful if the model is to be used in long-term simulation studies or in the design of hydraulic structures.

Some of the cell parameter estimates had large standard errors and were related. This is partially due to the difficulty in identifying cells in the physical process, and implies that the model has limited value in making inferences about the cellular structure of storms; the exception is the cell duration parameter which can be estimated with reasonable confidence.

There is more scope for further analysis of the extreme values. Further work might include incorporating the standard errors of the model parameters into the extreme-value analysis, though this would not change the point estimates of the Gumbel parameters. A more valuable exercise would be to compare simulated and observed data using the well-known 'peaks over thresholds' approach in hydrology (see, for example, Davison & Smith (1990) for a recent account), or the L-moments approach developed by Hosking (1990). As yet, these methods have not been considered in the context of the Poisson-cluster rectangular pulses models.

I thank the New Zealand National Institute of Water and Atmospheric Research (NIWA) for supplying the data. Sir David Cox is also gratefully acknowledged for his correspondence on a preprint of this paper.

References

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

Chandler, R. E. 1997 A spectral method for estimating parameters in rainfall models. *Bernoulli* **3**, 301–322.

Cowpertwait, P. S. P. 1994 A generalized point process model for rainfall. *Proc. R. Soc. Lond.* A **447**, 23–37.

Cowpertwait, P. S. P. 1995 A generalized spatial–temporal model of rainfall based on a clustered point process. *Proc. R. Soc. Lond.* A **450**, 163–175.

Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V. & Mawdsley, J. 1996 Stochastic point process modelling of rainfall. *J. Hydrol.* **175**, 17–65.

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

Davison, A. C. & Smith, R. L. 1990 Models for exceedances over high thresholds. *J. R. Statist. Soc.* B **52**, 393–442.

Fisher, R. A. & Tippett, L. H. C. 1928 Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180–190.

Foufoula-Georgiou, E. & Guttorp, P. 1987 Assessment of class of Neyman–Scott Models for temporal rainfall. *J. Geophys. Res.* **92**, 9679–9682.

Hosking, J. R. M. 1990 L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc.* B **52**, 105–124.

Islam, S., Entekhabi, D., Bras, R. L. & Rodriguez-Iturbe, I. 1990 Parameter estimation and sensitivity analysis for the modified Bartlett–Lewis rectangular pulses model of rainfall. *J. Geophys. Res.* **95**, 2093–2100.

Kavvas, M. L. & Delleur, J. W. 1981 A stochastic cluster model of daily rainfall sequences. *Water Resour. Res.* **17**, 1151–1160.

LeCam, L. 1961 A stochastic description of precipitation. In *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability* **3**, 165–186.

Mellor, D. & O'Connell, P. E. 1996 The Modified Turning Bands (MTB) model for space-time precipitation. II. Estimation of raincell parameters. *J. Hydrol.* **175**, 129–159.

Nelder, J. A. & Mead, R. 1965 A simplex method for function minimization. *Comp. J.* **7**, 308–313.

O'Neil, R. 1985 Function minimization using a Simplex procedure. Algorithm AS 47. In *Applied statistics algorithms* (ed. P. Griffiths & I. D. Hill). London: Ellis Horwood.

Onof, C. & Wheater, H. 1994 Improvements to the modelling of British rainfall using a modified random parameter Bartlett–Lewis rectangular pulse model. *J. Hydrol.* **157**, 177–195.

Rodriguez-Iturbe, I., Gupta, V. J. & Waymire, E. 1984 Scale considerations in the modelling of temporal rainfall. *Water Resour. Res.* **20**, 1611–1619.

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.