

Point Process Models of Rainfall: Developments for Fine-Scale Structure

Author(s): Paul Cowpertwait, Valerie Isham and Christian Onof

Source: *Proceedings: Mathematical, Physical and Engineering Sciences*, Oct. 8, 2007, Vol. 463, No. 2086 (Oct. 8, 2007), pp. 2569-2587

Published by: Royal Society

Stable URL: <https://www.jstor.org/stable/20209332>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/20209332?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Royal Society is collaborating with JSTOR to digitize, preserve and extend access to *Proceedings: Mathematical, Physical and Engineering Sciences*

Point process models of rainfall: developments for fine-scale structure

BY PAUL COWPERTWAIT^{1,*}, VALERIE ISHAM² AND CHRISTIAN ONOF³

¹*Institute of Information and Mathematical Sciences, Massey University,
Auckland 0745, New Zealand*

²*Department of Statistical Science, University College London,
London WC1E 6BT, UK*

³*Department of Civil and Environmental Engineering, Imperial College London,
London SW7 2AZ, UK*

A conceptual stochastic model of rainfall is proposed in which storm origins occur in a Poisson process, where each storm has a random lifetime during which rain cell origins occur in a secondary Poisson process. In addition, each cell has a random lifetime during which instantaneous random depths (or ‘pulses’) of rain occur in a further Poisson process. A key motivation behind the model formulation is to account for the variability in rainfall data over small (e.g. 5 min) and larger time intervals. Time-series properties are derived to enable the model to be fitted to aggregated rain gauge data. These properties include moments up to third order, the probability that an interval is dry, and the autocovariance function. To allow for distinct storm types (e.g. convective and stratiform), several processes may be superposed. Using the derived properties, a model consisting of two storm types is fitted to 60 years of 5 min rainfall data taken from a site near Wellington, New Zealand, using sample estimates taken at 5 min, 1 hour, 6 hours and daily levels of aggregation. The model is found to fit moments of the depth distribution up to third order very well at these time scales. Using the fitted model, 5 min series are simulated, and annual maxima are extracted and compared with equivalent values taken from the historical record. A good fit in the extremes is found at both 1 and 24 hour levels of aggregation, although at the 5 min level there is some underestimation of the historical values. Proportions of time intervals with depths below various low thresholds are extracted from the simulated and historical series and compared. A tendency for underestimation of the historical values is evident at some time scales, with a close fit being obtained as the threshold is increased.

Keywords: precipitation; Poisson cluster process; Bartlett-Lewis process; stochastic models; time series; extreme values

1. Introduction

In the literature on stochastic models of rainfall based on Neyman–Scott (NS) and Bartlett–Lewis (BL) processes (Rodriguez-Iturbe *et al.* 1987; Wheater *et al.* 2005), rainfall intensity is treated as a random variable that remains constant

* Author for correspondence (p.s.cowpertwait@massey.ac.nz).

Received 4 April 2007
Accepted 4 June 2007

2569

This journal is © 2007 The Royal Society

throughout the lifetime of a rain cell, so that rain cells are modelled using rectangular profiles. While rectangular profiles are unrealistic in continuous time, they provide a suitable approximation to discrete rainfall series that are aggregated over time intervals of 1 hour or more, and there are many examples in the literature where the NS or BL rainfall models provide good fits to aggregated series (Cowpertwait 1994; Verhoest *et al.* 1997; Onof *et al.* 2000).

Rainfall series sampled over 1 hour (or higher) time intervals are sufficient for many hydrological applications. However, some catchment studies—the analysis of urban drainage systems, for example—require rainfall series at a much finer resolution. When data are sparse or limited, one approach to this problem is to simulate 1 hour series using a BL or NS model, and to disaggregate the simulated series to a finer resolution using a stochastic disaggregation model (Cowpertwait *et al.* 1996*a,b*; Onof *et al.* 2005). Rather than combining two stochastic models in this way, which increases model uncertainty, it is more satisfactory to have a single model with an appropriate fine-scale structure that can be applied at any time scale.

Our objective in this paper is to develop a stochastic model of rainfall, based on the original BL rainfall model, with more realistic cell profiles. The result is that a good fit to various statistical properties of rainfall data can be achieved for a range of time scales from a fine resolution upwards, thus removing the need for disaggregation. We use a BL occurrence process of cell starting times; the methodology could also be applied to a NS process or other point processes of cell starting times.

2. Model formulation

In this section, we first review the formulation of the original BL rainfall model, and follow this with a description of the new model formulation.

The original BL model for the continuous process of rainfall intensity (Rodriguez-Iturbe *et al.* 1987) has the following structure, where we follow the terminology and notation of previous papers. Storm origins, $\{T_i\}$ say, occur in a Poisson process of rate λ . Each storm has a random lifetime of length D_i , where the D_i are independent and exponentially distributed with parameter γ , and terminates at time $T_i + D_i$. During the lifetime of the i th storm, cell origins, $\{T_{ij}\}$ say, are generated in a Poisson process of rate β . This process terminates at the end of the storm lifetime, i.e. $T_i < T_{ij} < T_i + D_i$. Without loss of generality, it is assumed that there is a rain cell starting at the storm origin. The point process $\{T_{ij}\}$ is a BL process.

The random durations of the cells, L_{ij} , are assumed to be exponentially distributed with parameter η , and each rain cell has an independent random intensity, X_{ij} , that remains constant over the duration of the cell. Both storms, and cells within storms, are assumed to evolve independently, and the rainfall process has a temporal Markov structure that simplifies the derivation of its properties. Both storms and cells can overlap and the rainfall intensity at any time t is the sum of the intensities of all cells whose lifetimes overlap t . In this model, once the rain cell has been initiated, it will continue for its full lifetime, regardless of whether or not the storm to which it belongs has terminated.

The moment properties of the rainfall intensity process depend only on the corresponding moments of the cell intensities and no particular distributional form need be assumed. Other properties, of extremes for example, will depend on the full distributional form. For model fitting, the continuous-time process is integrated (aggregated) to give rainfall depths over discrete disjoint time intervals.

The basic BL rainfall model can be used to represent rainfall intensities at time scales ranging from an hour to a day or so. However, as noted earlier, the assumed constant intensity of the rain cell means that the model cannot be used to give the subhourly structure that is needed for some hydrological applications, particularly in urban contexts. The basic model *could* be used to represent subhourly variation, if the time scales were reinterpreted, but this would be at the expense of larger-scale fluctuations. Alternatively, a jitter could be applied to the rectangular cells (e.g. see Rodriguez-Iturbe *et al.* 1987; Onof & Wheater 1994), but this approach, which is yet to be tested at fine time scales, is not germane to the overall Poisson cluster modelling strategy. Thus, in this paper, we further develop the basic model, by adding an additional level of structure within the rain cells, with the aim of extending the range of time scales over which it can be applied. In particular, we replace the constant cell intensity and assume, instead, that each rain cell origin, T_{ij} , initiates a sequence of rainfall pulses that occur in a Poisson process, $\{T_{ijk}\}$ say, of rate ξ . The combined effect of three Poisson processes should account for fluctuations over a wide range of scales and be sufficient for many practical applications (Koutsoyiannis 2002).

We do not assume that there is a pulse at the cell origin, and thus it is possible that no rainfall occurs during a rain cell. Similarly, and in contrast with the basic BL model, we shall not assume that there is a cell origin at the storm origin, so that storms may have no rainfall. These assumptions are for mathematical convenience and only affect the interpretation of the parameters. For example, in the pulse model, λ is the rate at which all storms occur, rather than just those that have non-zero amounts of rain. We will assume that the process of pulses terminates with the cell or storm lifetime, whichever is the sooner (i.e. at $\min(T_{ij} + L_{ij}, T_i + D_i)$), which has the effect of introducing dependence between cells within storms. For example, a short-duration storm will not only generate relatively few cells, but those cells will each tend to produce relatively few pulses owing to the common truncation effect of the storm lifetime.

Associated with each pulse is a random rainfall *depth*, X_{ijk} , generated by the pulse, so that the process $\{T_{ijk}, X_{ijk}\}$ is a marked point process (Cox & Isham 1980). In the algebraic derivations of model properties below, we allow pulses within a single cell to be dependent, although those in distinct cells will be assumed independent. For clarity, we will refer to this model as the Bartlett–Lewis pulse (BLP) model to distinguish it from the basic BL rainfall model. Note the differences from the latter model, where cells can continue after the corresponding storm had terminated, and where the variable X_{ij} is an *intensity*. As with the BL rainfall model, the continuous-time BLP process is integrated to give rainfall depths over discrete disjoint time intervals for model fitting and assessment. A schematic of the BLP model is shown in figure 1.

In the BLP model, the mean number of cells per storm is β/γ and the mean number of pulses per cell is $\xi/(\gamma + \eta)$, so that the mean number of pulses per storm is $\mu_p = \beta\xi\{\gamma(\gamma + \eta)\}^{-1}$. Thus, the expected total rainfall depth per storm

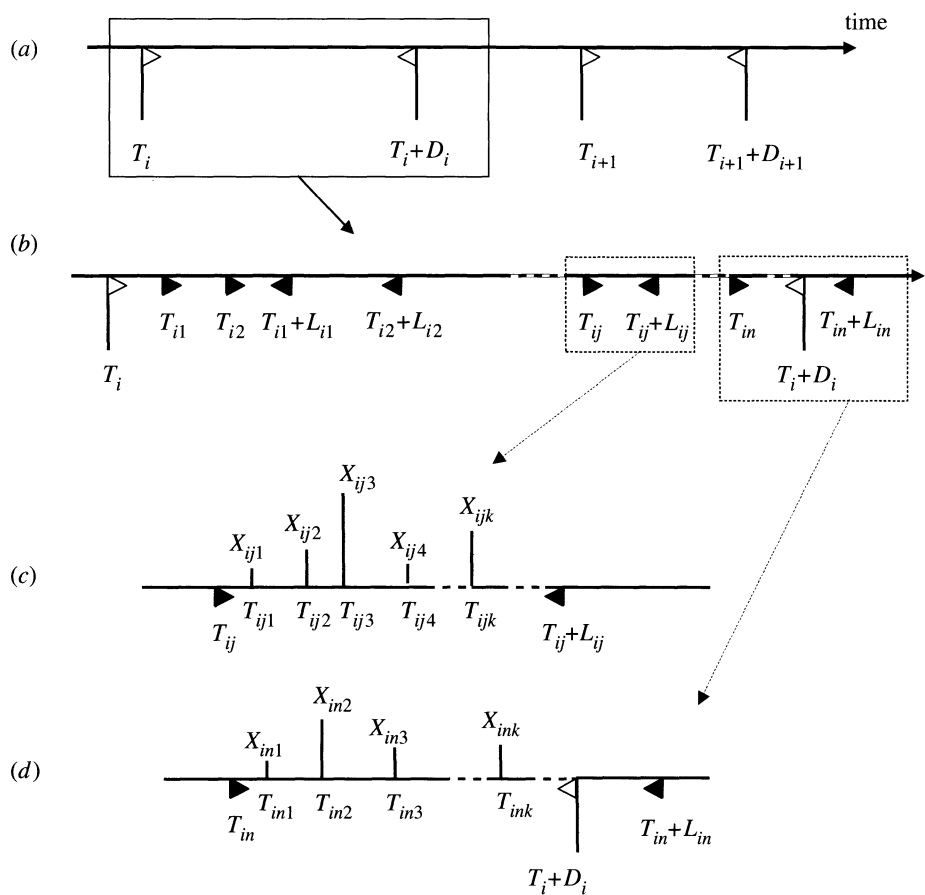


Figure 1. A schematic of the Bartlett–Lewis pulse model: (a) the storm arrival process, (b) the cell birth and death process in storm i , (c) the pulse processes in cell j , storm i for a cell that terminates before the end of the storm, and (d) the pulse processes in cell n , storm i for a cell terminating later than the end of the storm.

is $\mu_p \mu_X$, where $\mu_X = E(X_{ijk})$. In the fitting procedure (§6), we shall consider only the special case where the X_{ijk} are independent exponential random variables with parameter θ .

3. Covariance structure of pulse arrival times

The covariance structure of the point process of pulse occurrences is a major component in the derivation (see §4) of a number of key mathematical properties of the aggregated rainfall process that will be used for model fitting and assessment. We will denote by $N(\cdot)$ the counting process corresponding to the process of pulse occurrences from all storms, use $dN(t)=1$ as a convenient shorthand to denote the event $N(t+dt) - N(t) = 1$, for small increments dt , and consider the product density $p_2(t, t+u)$ (Cox & Isham 1980, p. 37) given by

$$p_2(t, t+u)dt\,du = P(dN(t) = dN(t+u) = 1).$$

Two distinct pulses occurring at times t and $t+u$ ($u>0$) may come from (i) different storms; (ii) the same storm, but different cells; and (iii) the same cell. For case (i), the contribution to the product density is simply $(\lambda\mu_p)^2$. In case (iii), the corresponding contribution can be seen to be $\lambda\mu_p\xi \exp\{-(\gamma+\eta)u\}$, because after the first pulse, the cell must continue for an additional time u and the cell terminates at rate $\gamma+\eta$. For case (ii), we have a contribution

$$\lambda(\beta\xi)^2 \int_{x=0}^{\infty} e^{-\gamma(x+u)} \int_{v=0}^x e^{-\eta v} \int_{w=0}^{x+u} e^{-\eta w} dw dv dx,$$

where in the notation used here the storm origin is at $t-x$ and the cell origins are at $t-v$ and $t+u-w$.

The covariance density (Cox & Isham 1980, p. 33) can be expressed in terms of the product density as $c(u) = \lambda\mu_p\delta(u) + p_2(t, t+u) - (\lambda\mu_p)^2$, which after some algebra simplifies to

$$c(u) = \lambda\mu_p\delta(u) + A e^{-\gamma u} + (B_1 - B_2)e^{-(\gamma+\eta)u}, \quad (3.1)$$

for $u \geq 0$, where $A = \lambda\mu_p\beta\xi/\eta$, $B_1 = \lambda\mu_p\xi$, $B_2 = A\gamma/(\gamma+2\eta)$ and $\delta(\cdot)$ is the Dirac delta function. Although this expression could be further simplified by combining B_1 and B_2 , we keep them separate here because B_1 corresponds to the contribution from different pulses within the same cell, where the depths of these pulses may be dependent, while B_2 corresponds to different, and hence independent, cells.

4. Properties of the rainfall process

(a) Second-order properties

The continuous time rainfall process is generally observed as an aggregated process in discrete time. Thus, let $Y_i^{(h)}$ be the total rainfall in a time interval $((i-1)h, ih]$, which can be expressed as

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} X(t) dN(t), \quad (4.1)$$

where $X(t)$ is the depth of a pulse located at time t . It follows immediately that $E[Y_i^{(h)}] = \lambda\mu_p\mu_X h$.

The variance and autocovariance function for the aggregated process can be deduced using the covariance density of the point process given in (3.1) above, using Campbell's theorem (e.g. Daley & Vere-Jones 1988, p. 188). Thus, we can write

$$\begin{aligned} \text{var}[Y_i^{(h)}] &= \int_0^h \int_0^h \text{cov}[X(s)dN(s), X(t)dN(t)] \\ &= \int_0^h E[X^2]E[dN(t)] + \iint_{s \neq t} E[X(s)X(t)]\text{cov}[dN(s), dN(t)], \end{aligned}$$

where we need to distinguish whether the pulses at s and t come from the same cell or different ones. For contributions to $\text{cov}[dN(s), dN(t)]$ from pulses in distinct cells, the multiplier $E[X(s)X(t)]$ will be simply μ_X^2 , whereas for pulses in the same cell we will write $E[X(s)X(t)]$ as $E(X_{ijk}X_{ij\ell})$ to allow for within-cell depth dependence. It is

assumed that any two pulses within a cell have the same expected product of depths regardless of their locations within the cell. Location-specific dependence would require a separate treatment. Thus, we obtain

$$\text{var} \left[Y_i^{(h)} \right] = \lambda \mu_p E(X^2)h + 2A\mu_X^2 \phi(\gamma) + 2[B_1 E(X_{ijk} X_{ijl}) - B_2 \mu_X^2] \phi(\gamma + \eta), \quad (4.2)$$

where $\phi(\alpha) = (e^{-\alpha h} - 1 + \alpha h)/\alpha^2$, and A , B_1 and B_2 are as defined in §3.

The autocovariance function for lagged aggregated rainfall can be obtained similarly. Again contributions from pulses within the same cell must be distinguished. In particular, for $k \geq 1$,

$$\begin{aligned} \text{cov} \left[Y_i^{(h)}, Y_{i+k}^{(h)} \right] &= \int_{kh}^{(k+1)h} \int_0^h \text{cov}[X(s) dN(s), X(t) dN(t)] \\ &= \int_{kh}^{(k+1)h} \int_0^h E[X(s)X(t)] \text{cov}[dN(s), dN(t)] \\ &= A\mu_X^2 \psi(\gamma) + [B_1 E(X_{ijk} X_{ijl}) - B_2 \mu_X^2] \psi(\gamma + \eta), \end{aligned} \quad (4.3)$$

where $\psi(\alpha) = e^{-\alpha(k-1)h} (1 - e^{-\alpha h})^2 / \alpha^2$.

In the special case where all pulse depths are independent, we simply replace $E(X_{ijk} X_{ijl})$ by μ_X^2 in the above expressions.

(b) The third moment

The rainfall process is far from Gaussian, and second-order moments do not capture all aspects of its behaviour so that, for both model fitting and the assessment of model adequacy, further properties are required. One property that has been found useful in model fitting is skewness (Cowpertwait 1998). The method of derivation of the third moment, $E((Y^{(h)})^3)$, is similar to (although algebraically more complex) that of the second moment, so an outline of the method together with the final expression is deferred to appendix A.

(c) The probability that an interval is dry

A further property of the model that is useful for model fitting or for assessing model adequacy is the probability that there is no rainfall during a particular interval. Thus, we consider $P(Y_i^{(h)} = 0)$ for rainfall aggregated into time intervals of width h . The approach, which is relatively routine but requires the enumeration of a number of possible cases, can be outlined as follows. Without loss of generality, we take $i=1$ and investigate the chance that $(0, h]$ is dry.

Suppose $T > 0$, and consider storm origins occurring in intervals $[-T, 0]$ and $(0, h]$, the numbers of which are independent Poisson variables with means λT and λh , respectively. Conditionally on these numbers, the storm origins are independently and uniformly located over $[-T, 0]$ and $(0, h]$, respectively. The final result requires the limit as $T \rightarrow \infty$.

Now consider first a storm origin at $-x$ ($0 < x < T$). If the storm lifetime, D , is less than x then it will contribute no pulses to $(0, h]$. If $D = x + u$, where $0 < u < h$, then the storm contributes independent Poisson numbers of cells in $[-x, 0]$ and

$(0, u]$ with means βx and βu , respectively. Again conditionally upon their numbers, the locations of the cell origins are independent and uniformly distributed over the respective intervals. Given the cell origin, and the storm duration, the probability that the cell does not contribute any pulses to $(0, h]$ can be obtained by further conditioning on the cell lifetime. Combining these probabilities over all the cells involved gives the probability that the storm does not contribute any rain to $(0, h]$.

Similarly, but with rather fewer possibilities to consider, suppose that a storm has its origin at x ($0 < x < h$), and condition on the time of storm death, which may be before or after h ; the Poisson numbers of cells generated before h ; and the uniformly distributed locations of the cell origins. Again, the probability that a cell does not contribute any pulses to $(0, h]$ can be obtained by further conditioning on its lifetime, and the probabilities can be combined over all the cells to give the probability that the storm does not contribute any rain to $(0, h]$.

Finally, the contributions from all storms can be pooled and the limit as $T \rightarrow \infty$ taken to obtain $P(Y_i^{(h)} = 0)$, which can be expressed in the following fairly simple form:

$$-\lambda^{-1} \ln P(Y_i^{(h)} = 0) = \int_0^\infty dx [e^{-\gamma x} - f(h, x) - \gamma \int_0^h du f(u, x)] \\ + \int_0^h dx \left[1 - g(x) - \gamma \int_0^x du g(u) \right], \quad (4.4)$$

where

$$-\ln f(u, x) = \gamma(x + u) + \frac{\beta\xi}{\eta(\eta + \xi)} (1 - e^{-\eta x})(1 - e^{-(\eta + \xi)u}) + \frac{\beta u \xi}{\eta + \xi} \\ - \frac{\beta\xi}{(\eta + \xi)^2} (1 - e^{-(\eta + \xi)u}) \quad (4.5)$$

and

$$-\ln g(u) = \gamma u + \frac{\beta u \xi}{\eta + \xi} - \frac{\beta\xi}{(\eta + \xi)^2} (1 - e^{-(\eta + \xi)u}). \quad (4.6)$$

In equation (4.4) above, the first term on the r.h.s. gives the combined contribution from all storms starting before 0, while the second term corresponds to storms starting during $(0, h]$.

5. Superposition of BLP models

In a previous work, it has been noted that distinct types of precipitation occur, for example convective and stratiform rain, which can be modelled using point processes that contain different cell types (Cowpertwait 1994). Alternatively, the cell duration parameter (η) can be made random between storms, so that different distributions of cell duration occur for different storms (Rodriguez-Iturbe *et al.* 1988). Another approach uses superposed independent processes of different storm types (Cowpertwait 2004).

An advantage of the last approach is that properties from independent processes can be aggregated to give the properties of the superposed process, thus making use of the derived properties of the independent processes. To illustrate, suppose there are n independent BLP processes, denoted BLP_i , with parameters $\lambda_i, \beta_i, \xi_i, \gamma_i, \eta_i$ and θ_i ($i=1, \dots, n$), where $\theta=1/\mu_X$. If $\{Y_{ij}^{(h)} : j=1, 2, \dots\}$ is the aggregated rainfall series sampled over an interval of width h due to a BLP_i process, then the total rainfall in the j th time interval, due to the superposition of the n independent BLP processes, is the sum $Y_j^{(h)} = \sum_{i=1}^n Y_{ij}^{(h)}$. Properties (up to third order) of the superposed process are obtained as the sum of the equivalent properties from each independent BLP_i process, i.e. $E[\{Y_j^{(h)} - \mu_j^{(h)}\}^k] = \sum_{i=1}^n E[\{Y_{ij}^{(h)} - \mu_{ij}^{(h)}\}^k]$ ($k \leq 3$), where $\mu_j^{(h)} = E[Y_j^{(h)}]$ and $\mu_{ij}^{(h)} = E[Y_{ij}^{(h)}]$.

Superposing processes can provide considerable flexibility. However, this can be at the expense of introducing large numbers of parameters into the model. Consequently, the number of superposed processes is usually restricted *a priori*. In the fitting procedure that follows we take $n=2$, which might provide a suitable approximation for the occurrence of convective and stratiform storms.

6. Fitted model

One motivation for developing the BLP model is to achieve a good fit to series of rainfall depths over a range of time scales, from fine resolutions, e.g. 5 min, to higher aggregation levels, such as daily. In this section, we aim to verify that the BLP model does satisfy this requirement for properties, including moments up to third order and extremes, that are important for hydrological applications. We shall consider only the special case of the model in which the X_{ijk} are independent exponentially distributed random variables, with parameter θ .

The New Zealand National Institute of Water and Atmospheric Research provided us with a 60 year record (1945–2004) of rainfall data recorded at a site in Kelburn (near Wellington, New Zealand). The data were based on a digitized pluviograph from a Dine's tilting siphon rain gauge (Sansom 1992), and were aggregated over 5 min time intervals. The data were therefore suitable for assessing the fit of the BLP model at fine resolutions.

A superposed model consisting of two independent BLP processes has a total of 12 parameters: $\lambda_i, \beta_i, \xi_i, \gamma_i, \eta_i$ and θ_i ($i=1, 2$). Some of these parameters can be treated as equal for different storm types, thus reducing the total number. A convenient choice for estimation purposes, which is made to obtain the results reported in this paper, is to assume that the mean pulse depth is the same for both storm types (i.e. that $\theta_1 = \theta_2 = \theta$) so that this quantity can then be estimated directly from the mean rainfall (see (6.5) below). This assumption implies that different storm types are distinguished by their underlying point processes rather than by the associated pulse depths, and would be plausible if, for example, pulses represent something like a minimum observable depth. In that case, perhaps θ might be further constrained to be the same for each month, but we do not pursue this idea here. With the assumption that $\theta_1 = \theta_2$, we have an 11

parameter model in which all the parameters, except θ , can be estimated from the dimensionless functions

$$\text{coefficient of variation, } \nu(h) = E \left[\left\{ Y_j^{(h)} - E(Y_j^{(h)}) \right\}^2 \right]^{1/2} / E(Y_j^{(h)}) \quad (6.1)$$

$$\text{autocorrelation (lag 1), } \rho(h) = \text{corr} \left(Y_j^{(h)}, Y_{j+1}^{(h)} \right) \quad (6.2)$$

and

$$\text{coefficient of skewness, } \kappa(h) = E \left[\left\{ Y_j^{(h)} - E(Y_j^{(h)}) \right\}^3 \right] / E \left[\left\{ Y_j^{(h)} - E(Y_j^{(h)}) \right\}^2 \right]^{3/2} \quad (6.3)$$

As the above functions contain 10 parameters, at least 10 sampled properties are needed to fit the model to data. We chose the following 12 properties: the coefficient of variation, lag 1 autocorrelation and coefficient of skewness, each taken at aggregation levels 5 min, 1, 6 and 24 hours, to cover a range of scales over which the model parameters might be identified. The sample estimates of these properties, denoted $\hat{\nu}(h)$, $\hat{\rho}(h)$ and $\hat{\kappa}(h)$ (for $h = 1/12, 1, 6$, and 24 hours), were calculated for each calendar month in the series, by pooling all available data for the month over the 60 year record for Kelburn. For each month, the estimates $\hat{\lambda}_i$, $\hat{\beta}_i$, $\hat{\xi}_i$, $\hat{\gamma}_i$, $\hat{\eta}_i$ ($i = 1, 2$) were obtained by minimizing the sum of squares

$$\sum_{h=1/12, 1, 6, 24} \left\{ \left(1 - \frac{\hat{\nu}(h)}{\nu(h)} \right)^2 + \left(1 - \frac{\nu(h)}{\hat{\nu}(h)} \right)^2 + \left(1 - \frac{\hat{\rho}(h)}{\rho(h)} \right)^2 + \left(1 - \frac{\rho(h)}{\hat{\rho}(h)} \right)^2 + \left(1 - \frac{\hat{\kappa}(h)}{\kappa(h)} \right)^2 + \left(1 - \frac{\kappa(h)}{\hat{\kappa}(h)} \right)^2 \right\}. \quad (6.4)$$

In (6.4), the model function is divided by the sample estimate to ensure that large sample values do not dominate the minimization procedure, so that each sample value receives an equal weight. In addition, the sum contains two terms per model function, one of which contains the model function divided by the sample value and the other contains the reciprocal term. This helps to ensure that the optimal solution is not biased from above or below the sample values, when an exact fit to the sample value is not obtained. However, the moment estimates themselves may be slightly biased owing to serial correlation, especially at the 5 min level (see §7a).

The sum of squared terms (6.4) was minimized using the simplex method (Nelder & Mead 1965); the resulting parameter estimates are given in table 1. In table 1, the scale parameter θ has been estimated for each month using the sample mean

$$\hat{\theta} = \bar{x}^{-1} \left[\frac{\hat{\lambda}_1 \hat{\beta}_1 \hat{\xi}_1}{\hat{\gamma}_1 (\hat{\gamma}_1 + \hat{\eta}_1)} + \frac{\hat{\lambda}_2 \hat{\beta}_2 \hat{\xi}_2}{\hat{\gamma}_2 (\hat{\gamma}_2 + \hat{\eta}_2)} \right], \quad (6.5)$$

where \bar{x} is the estimated mean 1 hour rainfall for the month. The sample and fitted values are shown in figures 2–5.

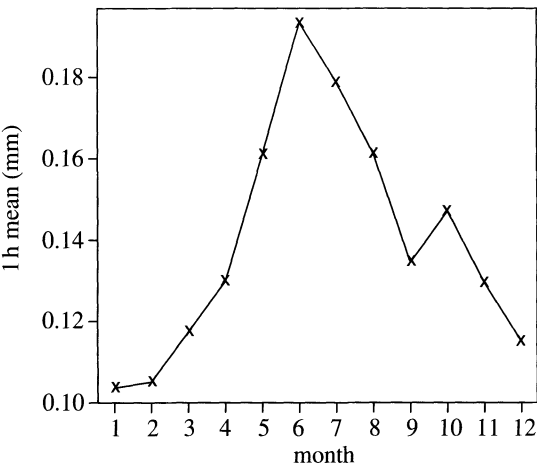


Figure 2. The mean 1 hour rainfall against month. The sample estimates are shown as crosses (cross) and the fitted values as lines.

Table 1. Fitted model: parameter estimates for Kelburn series. (The unit is h^{-1} for all estimates except that θ has the unit mm^{-1} .)

month	$\hat{\lambda}_1$	$\hat{\beta}_1$	$\hat{\xi}_1$	$\hat{\gamma}_2$	$\hat{\eta}_1$	$\hat{\lambda}_2$	$\hat{\beta}_2$	$\hat{\xi}_2$	$\hat{\gamma}_2$	$\hat{\eta}_2$	$\hat{\theta}^{-1}$
1	0.00641	0.133	509	0.0401	0.733	0.000273	1.43	4620	0.178	7.96	0.00676
2	0.00791	0.548	347	0.0988	1.23	0.000721	3.92	1810	0.269	7.73	0.0076
3	0.0116	0.359	446	0.107	0.822	0.00104	3.02	3250	0.401	7.12	0.00531
4	0.0109	0.137	684	0.0491	0.641	0.00123	1.53	4980	0.172	5.98	0.00333
5	0.0149	0.0654	320	0.0346	0.347	0.0015	0.773	2840	0.132	3.81	0.00537
6	0.0168	0.051	116	0.0299	0.228	0.00161	0.488	1510	0.179	2.8	0.0127
7	0.0101	0.0831	814	0.0245	0.294	0.00295	2.39	6210	0.197	5.94	0.00144
8	0.0166	0.909	661	0.103	1.38	0.00153	3.60	4350	0.271	6.53	0.00206
9	0.0153	0.398	211	0.0911	0.907	0.00146	2.85	1200	0.388	5.58	0.00824
10	0.0151	0.38	200	0.0923	0.897	0.000305	2.11	1620	0.236	6.17	0.0111
11	0.0122	0.259	389	0.0676	0.823	0.00103	3.74	3890	0.482	8.44	0.00547
12	0.0145	0.744	370	0.169	1.43	0.000218	2.46	2880	0.158	6.32	0.00705

An almost exact fit is obtained over all time intervals to properties up to third order (figures 2–5). A possible exception is the fit to the 5 min lag 1 autocorrelation shown in figure 5*a*. However, the actual differences in the correlations are of the order 0.01, which is unlikely to be of any practical consequence.

Comparing the parameter estimates for the different superposed processes, type 1 storms tend to be longer than type 2 storms ($\gamma_1 < \gamma_2$) and contain longer less intense cells ($\eta_1 < \eta_2$; $\xi_1 < \xi_2$; table 1). Broadly speaking, type 2 storms may be interpreted as rarer but rather heavier storms, while type 1 storms are more common and longer lasting. The estimates for the pulse depths are very small, generally $\hat{\theta} < 0.01$ mm, which suggests that the pulses may possibly correspond to small-scale droplets of rain.

Some evidence of seasonal variation is present in the parameter estimates (table 1). For example, over the summer months of January and February, $\hat{\lambda}$ takes a lower value when compared with the other months reflecting the lower

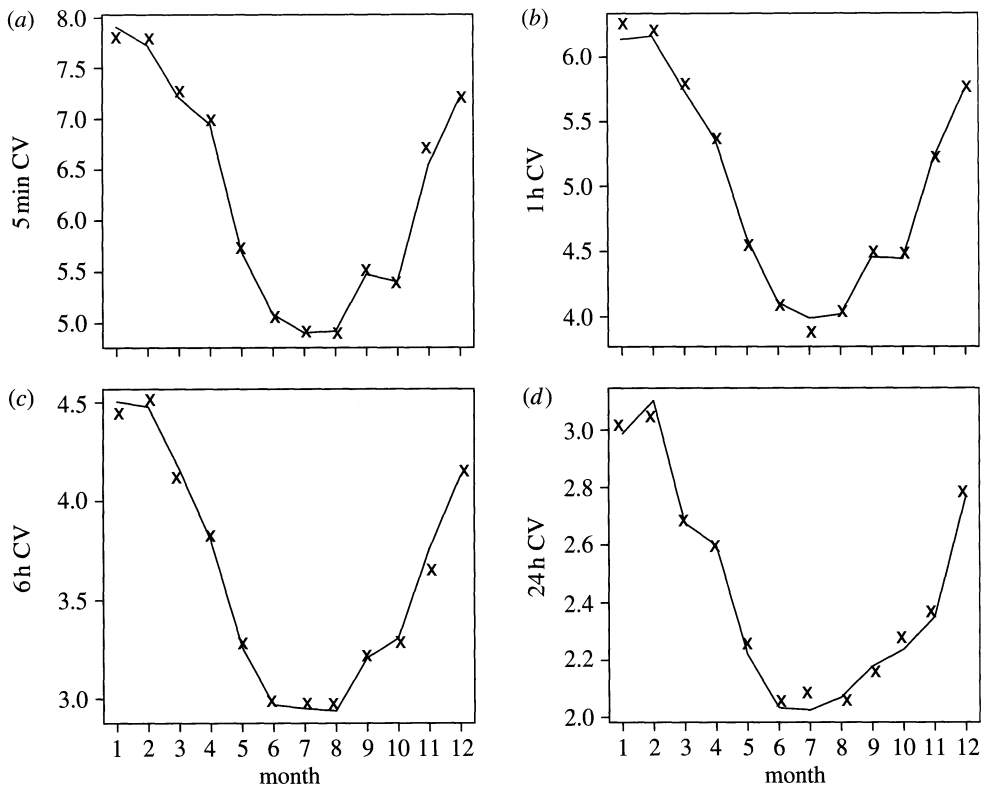


Figure 3. Coefficient of variation (CV) against month for aggregation levels: (a) 5 min; (b) 1 hour; (c) 6 hours; and (d) 24 hours. The sample estimates are shown as crosses (cross) and the fitted values as lines.

occurrence rate of summer storms. The other estimates show less seasonal variation, which could be due to irregular changes in the sample estimates between adjacent months, especially in the autocorrelation (figure 5).

7. Comparison with simulated series

(a) *Extreme values*

Before the design of a hydrological system is implemented, it is usually essential to quantify the performance of the system under extreme conditions. For certain systems, e.g. urban sewerage networks, this can be achieved by simulating rainfall series using a stochastic model, such as the model presented in this paper, and then feeding the simulated series into a hydraulic flow simulation model of the system. It is therefore important to assess the performance of a stochastic model of rainfall with respect to extreme values. Furthermore, as extreme values are unusual they form a stringent test when assessing model performance, especially when they are not directly used in the fitting procedure. In this section, extreme values generated by the model are compared to those in the 60 year historical record from Kelburn.

Using the fitted model (table 1), 20 samples of 5 min series, each of 60 year duration, were simulated. For each sample, the annual maxima over a 5 min

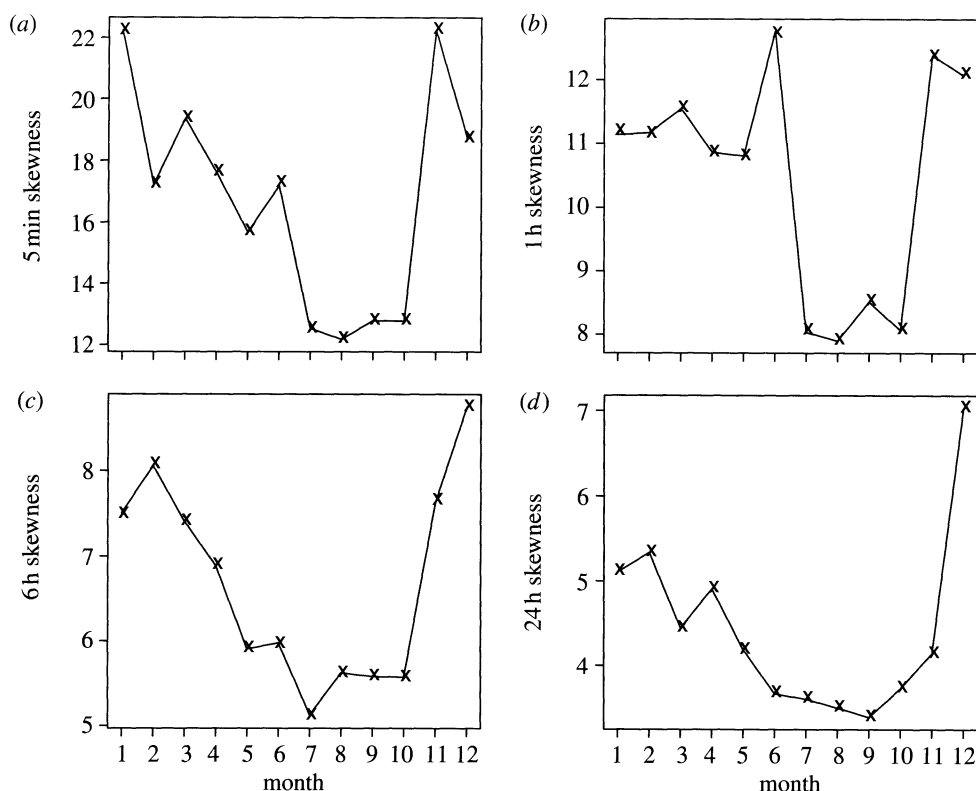


Figure 4. Coefficient of skewness against month for aggregation levels: (a) 5 min; (b) 1 hour; (c) 6 hours; and (d) 24 hours. The sample estimates are shown as crosses (cross) and the fitted values as lines.

interval were extracted, ordered and plotted against the reduced Gumbel variate. The results, with the equivalent value from the 60 year historical record, are shown in figure 6. In addition, each generated sample of 5 min data was aggregated to the 1 and 24 hours levels, and the annual maxima were extracted for the aggregated series (figures 7 and 8).

At the higher levels of aggregation, there is good agreement; the historical values generally fall within the range of simulated sample values (figures 7 and 8). However, for the 5 min aggregation level, there is evidence of underestimation for return periods in excess of about 3 years (figure 6). There are many possible reasons for this, one of which is that there is probably some bias in the sample estimate of skewness, which is likely to be more significant at smaller aggregation levels, such as 5 min, due to high autocorrelation (e.g. compare figure 5a and 5d). This requires further detailed investigation, the results of which will be reported in due course, along with a more extensive empirical analysis.

(b) *Proportion below a threshold*

Another property important in many applications is the proportion of dry intervals. For example, in urban drainage applications, a heavy storm following a long dry spell may cause a high concentration of pollution in the receiving water course (Cowpertwait *et al.* 1996a,b).

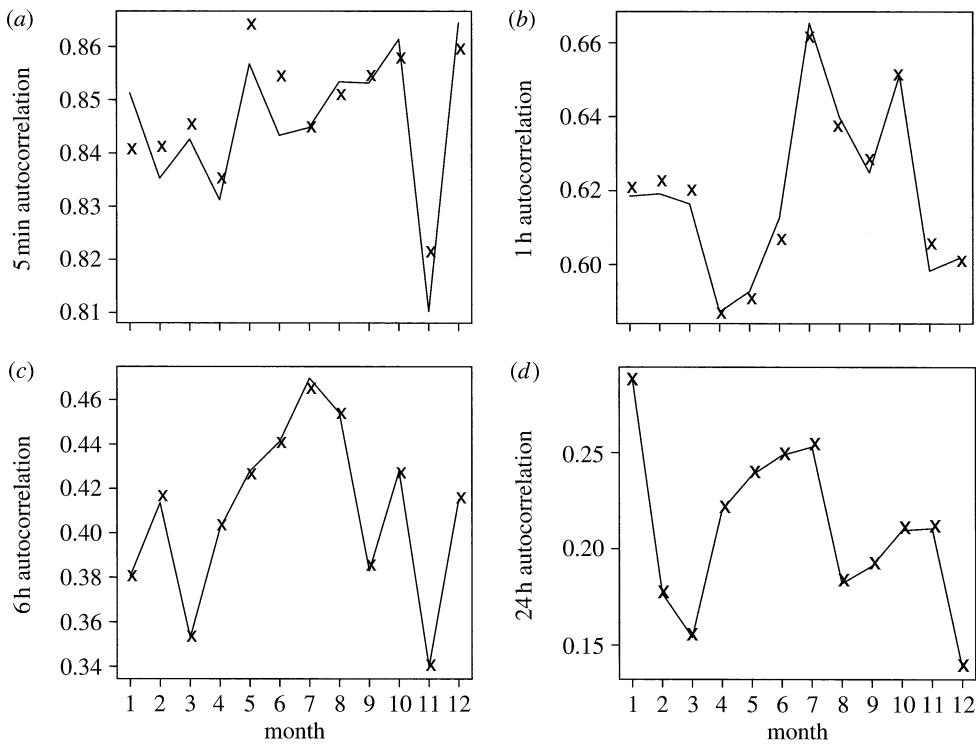


Figure 5. Lag 1 autocorrelation against month for aggregation levels: (a) 5 min; (b) 1 hour; (c) 6 hours; and (d) 24 hours. The sample estimates are shown as crosses (cross) and the fitted values as lines.

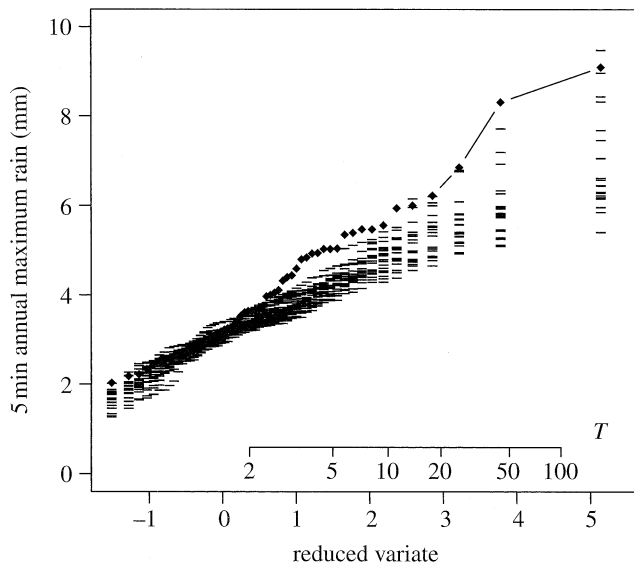


Figure 6. Ordered annual maxima taken over 5 min time intervals plotted against the reduced Gumbel variate. The values from the historical series are shown as (filled diamonds), and are connected with lines. The simulated values are shown as bars (dash). For the simulated series there are 20 samples, each of 60 year duration.

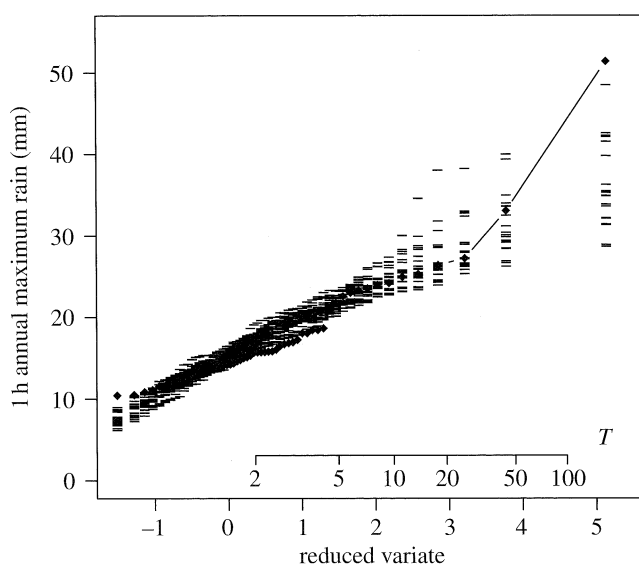


Figure 7. Ordered annual maxima taken over 1 hour time intervals plotted against the reduced Gumbel variate. The values from the historical series are shown as (filled diamonds) and are connected with lines. The simulated values are shown as bars (dash). For the simulated series there are 20 samples, each of 60 year duration.

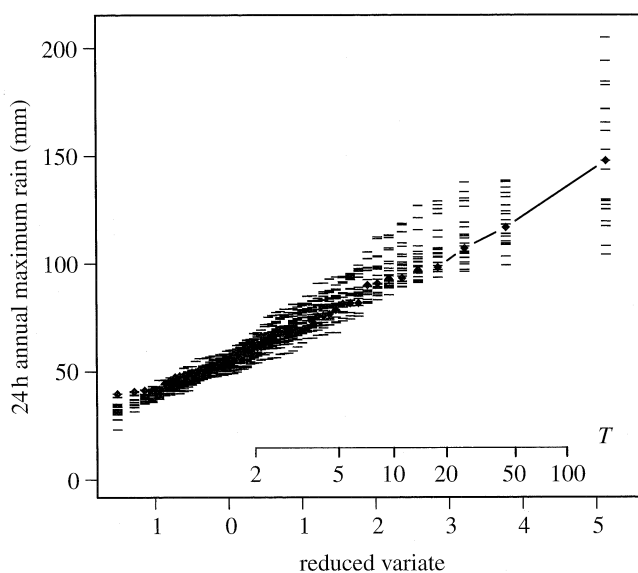


Figure 8. Ordered annual maxima taken over daily time intervals plotted against the reduced Gumbel variate. The values from the historical series are shown as (filled diamonds) and are connected with lines. The simulated values are shown as bars (dash). For the simulated series there are 20 samples, each of 60 year duration.

Rainfall series are often stored in a rounded form (e.g. to the nearest 0.1 mm), so that we prefer to calculate the proportion of values less than a small threshold ($\hat{p}\{Y_i^{(h)} < \delta\}$ for $\delta > 0$), rather than the actual proportion of

Table 2. Proportion below a threshold: $\hat{p}\{Y_i^{(h)} < \delta\}$. (\hat{p}_O and \hat{p}_S represent the estimates for the observed and simulated series.)

month	\hat{p}_O	\hat{p}_S	\hat{p}_O	\hat{p}_S	\hat{p}_O	\hat{p}_S	\hat{p}_O	\hat{p}_S
1	0.963	0.971	0.924	0.953	0.746	0.785	0.817	0.814
2	0.962	0.972	0.922	0.954	0.741	0.787	0.814	0.817
3	0.957	0.967	0.907	0.946	0.697	0.765	0.780	0.799
4	0.951	0.945	0.896	0.921	0.646	0.709	0.748	0.75
5	0.935	0.926	0.864	0.903	0.576	0.647	0.690	0.695
6	0.920	0.920	0.831	0.888	0.517	0.600	0.629	0.661
7	0.922	0.915	0.836	0.883	0.521	0.581	0.650	0.648
8	0.927	0.924	0.843	0.890	0.525	0.596	0.639	0.665
9	0.939	0.939	0.873	0.904	0.604	0.634	0.709	0.699
10	0.938	0.954	0.874	0.924	0.592	0.668	0.709	0.725
11	0.947	0.957	0.890	0.931	0.630	0.701	0.734	0.748
12	0.956	0.967	0.907	0.946	0.668	0.752	0.772	0.787
threshold	0.05	0.05	0.05	0.05	0.5	0.5	2	2
δ (mm)								
scale (h)	1/12	1/12	1	1	24	24	24	24
s.d.	—	0.003	—	0.03	—	0.07	—	0.06

dry intervals ($\delta=0$). To approximate the proportion of dry intervals, a small value of δ can be chosen.

Using the 20 samples of 60 years of simulated 5 min data, the proportion below a threshold was calculated for each month for the 5 min, 1 hour and daily aggregation levels, and the mean and standard deviations of these 20 values were found for each calendar month (table 2). To get an estimate of the proportion of dry periods, a small threshold of $\delta=0.05$ mm was used for the 5 min ($h=1/12$) and 1 hour ($h=1$) time scales. For the daily time scale, a higher threshold of $\delta=0.5$ mm was used, as it may be erroneous to assume that values less than 0.5 mm measured over a day are caused by a precipitation event, when they could be caused by a build-up of condensation in the gauge. For the daily level, a higher value of $\delta=2$ mm was also used to assess whether very light rain may cause discrepancies between the simulated and observed proportions (table 2).

In table 2, it is evident that the model generally reproduces the observed proportions well at the 5 min level, with differences of the order of 0.01, with a tendency slightly to over-estimate the proportions in the summer months, with the winter proportions being nearly exact in two cases (June, September) and slightly underestimated in the others. At the 1 hour level, the model slightly overestimates the observed proportions, although the simulated values tend to be within one or two standard deviations of the observed values, so the differences are not of much significance. For the daily level, all the values for the simulated series overestimate the observed values when the threshold is 0.5 mm. However, when the threshold is increased to 2 mm, the proportions for the observed and simulated series become close, which suggests that the

frequency of very small rainfall events, over the period of a day, is greater in the historical record than in the simulated record.

8. Summary and conclusions

A stochastic process of pulses has been introduced into the cell structure of the BL rainfall model (Rodriguez-Iturbe *et al.* 1987). Moments up to third order have been derived and used to fit the model to 5 min data. The fitted properties of the model generally agree well with observed values, indicating that the model is able to fit data over a range of time scales from 5 min upwards. This suggests that the model has potential application in a number of areas, for example urban drainage catchment studies, which usually require 5 min rainfall series.

The analysis of the proportions of intervals with rainfall depths below thresholds indicates that the model generally reproduces these values well, especially as the threshold increases, which suggests that the model may be of particular practical value in some urban hydrological applications. Some lack-of-fit at small thresholds may be due to very small events in the observed data, which are not reproduced at the correct frequency in the simulations. In principle, this could be addressed by introducing further parameters into the model, but this is not a priority in model development as the practical consequence will be small.

The simulated extreme values at the 1 and 24 hour levels of aggregation show good agreement with the equivalent values taken from the historical record. The 5 min extremes from the historical series are underestimated for return periods greater than about 3 years. This could be due to bias in the estimation of skewness, which may be more significant at the 5 min aggregation level, as compared with higher levels, owing to the larger autocorrelation values. A modification of the model, for example to allow a longer-tailed distribution for pulse depth, might also improve the fit of simulated extremes at the 5 min level, while retaining the existing good fit at the higher levels. Some further research is required to assess these ideas.

The analysis of extremes based on annual maxima is a stringent test of model performance, and it is likely that the existing model, without further modification, will be sufficient for many practical applications. Further testing is required, appropriate and specific to any intended application, before practical implementation. A comparison with disaggregation models currently used in practical applications (e.g. Cowpertwait *et al.* 1996*a,b*; Onof *et al.* 2005) will be the subject of a further study.

Many variants of the BLP model can be considered for development. For example, a NS process could replace the Barlett–Lewis cell arrival process and/or the pulse occurrence process. In addition, if pulse arrivals are defined relative to cell arrivals (e.g. via a NS process), then a further option is to allow pulse arrivals to precede their respective cell origins, which would enable a non-uniform distribution of pulse arrivals to occur during cell lifetimes. The impact of such variations in model performance is hard to predict owing to the levels of clustering involved. Mathematical tractability will also be an important criterion in any such further developments. A range

of algebraically tractable options is currently being investigated, the results of which will be reported in due course.

The National Institute of Water and Atmospheric Research is gratefully acknowledged for providing the data.

Appendix A

The third moment $E((Y^{(h)})^3)$ is given by

$$E[(Y_i^{(h)})^3] = \int_0^h \int_0^h \int_0^h E[X(x)X(y)X(z)]E[dN(x)dN(y)dN(z)].$$

In the region of integration, for $u>0, v>0$, there are six orderings of the form $x < y = x + u < z = x + u + v$ and permutations, three orderings of the form $x < y = z = x + u$, a further three of the form $x = y < z = x + u$ (and, given the independencies assumed, the contributions are the same whether the repeated depth is at the earlier or later time) and, finally, the possibility $x = y = z$ to consider. Thus, we can write

$$\begin{aligned} E[(Y_i^{(h)})^3] &= 6 \int_{x=0}^h \int_{u=0}^{h-x} \int_{v=0}^{h-x-u} E[X(x)X(x+u)X(x+u+v)dN(x)dN(x+u)dN(x+u+v)] \\ &\quad + 6 \int_{x=0}^h \int_{u=0}^{h-x} E[X(x)(X(x+u))^2dN(x)dN(x+u)] \\ &\quad + \int_{x=0}^h E[(X(x))^3dN(x)]. \end{aligned} \tag{A 1}$$

As before, in obtaining the product densities, we need to retain information about whether or not the pulses at x , y and z belong to the same cell. For the second of the three integrals above, where there are only two time points x and $x+u$, the breakdown is exactly as that used to determine the second moments. For the first integral, a rather more complicated breakdown of possible cases is involved, first depending on whether x , y and z belong to the same or different storms and then, when the storms are the same, whether the cells are the same or not. The third integral is easily seen to reduce to $\lambda\mu_p E(X^3)h$.

After enumeration of the various possibilities, we obtain the following result, which has been used for model fitting. Here, $E(X_{ijk}, X_{ijl}, X_{ijn})$ denotes the product moment of the depths of three pulses, all generated within the same cell. The terms can be grouped together in a variety of alternative ways, but none are

especially simple or transparent.

$$\begin{aligned}
& \frac{E((Y^{(h)})^3)}{\lambda\beta\xi^3} \\
&= \frac{6}{(\eta+\gamma)^2} \left[\frac{E(X_{ijk}X_{ij\ell}X_{ijn})}{\gamma(\eta+\gamma)} + \frac{E(X_{ijk}X_{ij\ell})\mu_X\beta}{\gamma(\eta+\gamma)} \left(\frac{1}{\eta} + \frac{\lambda}{\gamma(\eta+\gamma)} - \frac{\gamma}{\eta(2\eta+\gamma)} \right) \right. \\
&\quad \left. - \frac{\mu_X^3\beta^2}{\eta(\eta+\gamma)(2\eta+\gamma)} \left(\frac{1}{\eta} + \frac{\lambda}{\gamma(\eta+\gamma)} \right) \right] \left[h - \frac{2}{(\eta+\gamma)} + \frac{2e^{-(\eta+\gamma)h}}{(\eta+\gamma)} + he^{-(\eta+\gamma)h} \right] \\
&+ \frac{6}{(\eta+\gamma)(2\eta+\gamma)} \left[-\frac{2E(X_{ijk}X_{ij\ell})\mu_X\beta}{\eta(\eta+\gamma)(2\eta+\gamma)} + \frac{\mu_X^3\beta^2}{\eta^2(2\eta+\gamma)(3\eta+\gamma)} \right] \\
&\quad \times \left[h - \frac{3\eta+2\gamma}{(\eta+\gamma)(2\eta+\gamma)} + \frac{(2\eta+\gamma)e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)} - \frac{(\eta+\gamma)e^{-(2\eta+\gamma)h}}{\eta(2\eta+\gamma)} \right] \\
&+ \frac{6\mu_X^3\beta^2}{\eta\gamma^3(\eta+\gamma)} \left(\frac{1}{\eta} + \frac{\lambda}{\gamma(\eta+\gamma)} \right) \left[h - \frac{2}{\gamma} + \frac{2e^{-\gamma h}}{\gamma} + he^{-\gamma h} \right] \\
&+ \frac{6}{\gamma(\eta+\gamma)} \left[\frac{2E(X_{ijk}X_{ij\ell})\mu_X\beta}{\eta\gamma(\eta+\gamma)} - \frac{\mu_X^3\beta^2}{\eta^2(\eta+\gamma)(2\eta+\gamma)} \right] \\
&\quad \times \left[h - \frac{\eta+2\gamma}{\gamma(\eta+\gamma)} + \frac{(\eta+\gamma)e^{-\gamma h}}{\eta\gamma} - \frac{\gamma e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)} \right] \\
&+ \frac{12}{\eta+\gamma} \left[\frac{E(X_{ijk}X_{ij\ell})\mu_X\lambda\beta}{\gamma^2(\eta+\gamma)^2} - \frac{\mu_X^3\lambda\beta^2}{\eta\gamma(\eta+\gamma)^2(2\eta+\gamma)} \right] \\
&\quad \times \left[\frac{1}{2}h^2 - \frac{h}{\eta+\gamma} + \frac{1}{(\eta+\gamma)^2} - \frac{e^{-(\eta+\gamma)h}}{(\eta+\gamma)^2} \right] \\
&+ \frac{12\mu_X^3\lambda\beta^2}{\eta\gamma^3(\eta+\gamma)^2} \left[\frac{1}{2}h^2 - \frac{h}{\gamma} + \frac{1}{\gamma^2} - \frac{e^{-\gamma h}}{\gamma^2} \right] + \frac{\mu_X^3\lambda^2\beta^2h^3}{\gamma^3(\eta+\gamma)^3} \\
&+ \frac{6E(X_{ijk}^2X_{ij\ell})}{\xi\gamma(\eta+\gamma)^2} \left[h - \frac{1}{\eta+\gamma} + \frac{e^{-(\eta+\gamma)h}}{\eta+\gamma} \right] \\
&+ \frac{6E(X^2)\mu_X\beta}{\xi\eta\gamma^2(\eta+\gamma)} \left[h - \frac{1}{\gamma} + \frac{e^{-\gamma h}}{\gamma} - \frac{\gamma^2}{(\eta+\gamma)(2\eta+\gamma)} \left(h - \frac{1}{\eta+\gamma} + \frac{e^{-(\eta+\gamma)h}}{\eta+\gamma} \right) \right] \\
&+ \frac{3E(X^2)\mu_X\lambda\beta h^2}{\xi\gamma^2(\eta+\gamma)^2} + \frac{E(X^3)h}{\xi^2\gamma(\eta+\gamma)}.
\end{aligned}$$

References

- Cowpertwait, P. S. P. 1994 A generalized point process model for rainfall. *Proc. R. Soc. A* **447**, 23–37. (doi:10.1098/rspa.1994.0126)
- Cowpertwait, P. 1998 A Poisson-cluster model of rainfall: high order moments and extreme values. *Proc. R. Soc. A* **454**, 885–898. (doi:10.1098/rspa.1998.0191)

Proc. R. Soc. A (2007)

- Cowpertwait, P. 2004 Mixed rectangular pulses models of rainfall. *Hydrol. Earth Syst. Sci.* **8**, 993–1000.
- Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V. & Mawdsley, J. A. 1996*a* Stochastic point process modelling of rainfall. I. Single-site fitting and validation. *J. Hydrol.* **175**, 17–46. (doi:10.1016/S0022-1694(96)80004-7)
- Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V. & Mawdsley, J. A. 1996*b* Stochastic point process modelling of rainfall. II. Regionalisation and disaggregation. *J. Hydrol.* **175**, 47–65. (doi:10.1016/S0022-1694(96)80005-9)
- Cox, D. & Isham, V. 1980 *Point processes*. London, UK: Chapman and Hall.
- Daley, D. & Vere-Jones, D. 1988 *An introduction to the theory of point processes*. New York, NY: Springer.
- Koutsoyiannis, D. 2002 The Hurst phenomenon and fractional gaussian noise made easy. *Hydrol. Sci. J.* **47**, 573–595.
- Nelder, J. & Mead, R. 1965 A simplex algorithm for function minimization. *Comput. J.* **7**, 308–313.
- Onof, C. & Wheeler, H. 1994 Improvements to the modelling of British rainfall using a modified random parameter Bartlett–Lewis rectangular pulse model. *J. Hydrol.* **157**, 177–195. (doi:10.1016/0022-1694(94)90104-X)
- Onof, C., Chandler, R., Kakou, A., Northrop, P., Wheeler, H. & Isham, V. 2000 Rainfall modelling using Poisson-cluster processes. *Stoch. Environ. Res. Risk Assess.* **14**, 384–411. (doi:10.1007/s004770000043)
- Onof, C., Townend, J. & Kee, R. 2005 Comparison of two hourly to 5-min rainfall disaggregators. *Atmos. Res.* **77**, 176–187. (doi:10.1016/j.atmosres.2004.10.022)
- Rodriguez-Iturbe, I., Cox, D. & Isham, V. 1987 Some models for rainfall based on stochastic point processes. *Proc. R. Soc. A* **410**, 269–288. (doi:10.1098/rspa.1987.0039)
- Rodriguez-Iturbe, I., Cox, D. & Isham, V. 1988 A point process model for rainfall: further developments. *Proc. R. Soc. A* **417**, 283–298. (doi:10.1098/rspa.1988.0061)
- Sansom, J. 1992 Breakpoint representation of rainfall. *J. Appl. Meteorol.* **31**, 1514–1519. (doi:10.1175/1520-0450(1992)031<1514:BROR>2.0.CO;2)
- Verhoest, N., Troch, P. & Troch, F. D. 1997 On the applicability of Bartlett–Lewis rectangular pulses models in the modelling of design storms at a point. *J. Hydrol.* **202**, 108–120. (doi:10.1016/S0022-1694(97)00060-7)
- Wheeler, H., Chandler, R., Onof, C., Isham, V., Bellone, E., Yang, C., Lekkas, D., Lourmas, G. & Segond, M.-L. 2005 Spatial-temporal rainfall modelling for flood risk estimation. *Stoch. Environ. Res. Risk Assess.* **19**, 403–416. (doi:10.1007/s00477-005-0011-8)