

## Central Lancashire Online Knowledge (CLoK)

Title	Doubly stochastic Poisson pulse model for fine-scale rainfall
Type	Article
URL	<a href="https://clock.uclan.ac.uk/14645/">https://clock.uclan.ac.uk/14645/</a>
DOI	##doi##
Date	2016
Citation	Thayakaran, Rasiah orcid iconORCID: 0000-0002-4422-4095 and Ramesh, N. I (2016) Doubly stochastic Poisson pulse model for fine-scale rainfall. Stochastic Environmental Research and Risk Assessment, 31 (3). pp. 705-724. ISSN 1436-3240
Creators	Thayakaran, Rasiah and Ramesh, N. I

It is advisable to refer to the publisher's version if you intend to cite from the work. ##doi##

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

# Doubly stochastic Poisson pulse model for fine-scale rainfall

R Thayakaran

College of Health and Wellbeing, University of Central Lancashire

Preston, Lancashire, PR1 2HE, UK

Tel.: +44 (0)1772 89 3771

rthayakaran@uclan.ac.uk

and

N I Ramesh

Department of Mathematical Sciences, University of Greenwich, Maritime Greenwich Campus

Old Royal Naval College, Park Row, Greenwich

London SE10 9LS, UK

May 12, 2016

## Abstract

Stochastic rainfall models are widely used in hydrological studies because they provide a framework not only for deriving information about the characteristics of rainfall but also for generating precipitation inputs to simulation models whenever data are not available. A stochastic point process model based on a class of doubly stochastic Poisson processes is proposed to analyse fine-scale point rainfall time series. In this model, rain cells arrive according to a doubly stochastic Poisson process whose arrival rate is determined by a finite-state Markov chain. Each rain cell has a random lifetime. During the lifetime of each rain cell, instantaneous random depths of rainfall bursts (pulses) occur according to a Poisson process. The covariance structure of the point process of pulse occurrences is studied. Moment properties of the time series of accumulated rainfall in discrete time intervals are derived to model 5-minute rainfall data, over a period of 69 years, from Germany. Second-moment as well as third-moment properties of the rainfall are considered. The results show that the proposed model is capable of reproducing rainfall properties well at various sub-hourly resolutions. Incorporation of third-order moment properties in estimation showed a clear improvement in fitting. A good fit to the extremes is found at larger resolutions, both at 12-hour and 24-hour levels, despite underestimation at 5-minute aggregation. The proportion of dry intervals is studied by comparing the proportion of time intervals, from the observed and simulated data, with rainfall depth below small thresholds. A good agreement was found at 5-minute aggregation and for larger aggregation levels a closer fit was obtained when the threshold was increased. A simulation study is presented to assess the performance of the estimation method.

**Key Words:** Moment properties; Point process; Precipitation; Rainfall generator; Stochastic models; Sub-hourly rainfall

# 1 Introduction

An important challenge we face in environmental or ecological impact studies is to provide fast and realistic simulations of atmospheric variables such as rainfall at various temporal scales. Stochastic point process models provide a basis for generating synthetic rainfall input to hydrological models where the observed data at the required temporal scale are not available. They also enable us to assess the risk associated with hydrological systems. There has been a number of stochastic point process models developed by many authors over the years. Among these, the models based on Poisson cluster processes (Rodriguez-Iturbe et al. (1987), Cowpertwait (1994), Onof and Wheater (1994) and Wheater et al. (2005)) utilizing either the Neyman-Scott or Bartlett-Lewis processes have received much attention, since their model structure reflects well the climatological features of the rainfall generating mechanism. A good review of developments in modelling rainfall using Poisson cluster processes is provided by Onof et al. (2000). In addition, rainfall models based on Markov processes have also made a reasonable contribution to help tackle this challenge. See for example, Smith and Karr (1983), Bardossy and Plate (1991), Ramesh (1998), Onof et al. (2002) and Ramesh and Onof (2014) amongst others.

Much of the work on this topic, however, has concentrated on modelling rainfall data recorded at hourly or longer aggregation levels. Stochastic models for fine-scale rainfall are equally important, because in some hydrological applications there is a need to reproduce rainfall time series at fine temporal resolutions. For example, sub-hourly rainfall is required as input to urban drainage models and for small rural catchment studies. In addition, the study of climate change impacts on hydrology and water management initiatives requires the availability of data at fine temporal resolutions, which is usually not available from general circulation model (GCM) simulations. The best available approach to generating such rainfall currently lies in the combination of an hourly stochastic rainfall simulator, together with a disaggregator making use of downscaling techniques. There has been some recent work on modelling fine-scale rainfall using point process models. Rather than attempting to reproduce actual rainfall records at a fine-scale, using downscaling techniques or by other methods, these stochastic point process models aim to generate synthetic precipitation time series directly from the proposed stochastic model. One good example of this is provided by the work of Cowpertwait et al. (2007, 2011) who developed a Bartlett-Lewis pulse model to study fine-scale rainfall structure. Their model particularly enables the reproduction of the fine time-scale properties of rainfall. A class of doubly stochastic Poisson processes was employed by Ramesh et al. (2012, 2013) and Thayakaran and Ramesh (2013) to study fine-scale rainfall intensity using rainfall bucket tip times data. They utilised maximum likelihood methods to estimate parameters of their models.

Our main objective in this paper is to develop a simple stochastic point process model capable of reproducing fine-scale structure of the rainfall process. The other objective is to provide a fast and efficient way of generating synthetic fine-scale rainfall input to hydrological models directly from one stochastic model. In this regard, and to take this fine-scale rainfall modelling work further, we develop a simple point process model based on a doubly stochastic Poisson process, following the Poisson cluster pulse model approach of Cowpertwait et al. (2007). Our preliminary work on this (Ramesh and Thayakaran, 2012), analysing properties of rainfall time series at sub-hourly resolutions, produced encouraging initial results. In this paper, we extend this work further and accommodate third-order moments in estimation. Mathematical expressions for the moment properties of the accumulated rainfall in disjoint intervals are derived. The proposed stochastic model is fitted to 69 years of 5-minute rainfall time series from Germany. The results of the analysis show that the proposed model is capable of re-

producing rainfall properties well at various sub-hourly resolutions. Furthermore, the analysis incorporating third-order moments produced better results than the one that used only up to second-order moments. Unlike Cowpertwait et al. (2007), who used superposition of two Bartlett-Lewis pulse models to account for different storm types, we use one simple model to reproduce sub-hourly rainfall structure. The novel contribution of this study is the derivation of the third-order moment properties of the proposed model, as well as their incorporation in estimation, to reproduce fine-scale structure of rainfall more accurately. The proposed model provides a solid framework to generate synthetic fine-scale rainfall input to hydrological models directly from one stochastic point process model. In addition, the availability of a new stochastic model for the generation of fine-scale rainfall, at various sub-hourly scales, provides scientists with a useful tool for environmental or ecological impact studies.

The following section provides a background to this work, illustrates the model framework and then focuses on deriving moment properties of various component processes, such as the cell and pulse arrival processes. Properties of the aggregated rainfall sequence are studied and mathematical expressions for the third-order moment and the coefficient of skewness are derived. Section 3 presents the results of data analysis using 5-minute rainfall aggregations and compares the results of two different analyses that used second and third-order moments in estimation. Extremes of the historical data are compared with the simulated extremes at various resolutions. The proportion of dry intervals is also studied. A simulation study is carried out to evaluate the performance of the estimation method. Conclusions and possible further work are summarised in Section 4.

## 2 Model framework and moment properties

### 2.1 Background

Doubly stochastic Poisson processes (DSPP) provide a flexible set of point process models for fine-scale rainfall. Ramesh et al. (2012, 2013) utilised this class of processes and developed stochastic models, for a single-site and multiple sites respectively. These models were used to analyse data collected in the form of rainfall bucket tip-time series. One of the advantages of these models, over most other point process models for rainfall, is that it is possible to write down their likelihood function which allows us to estimate the model parameters using maximum likelihood methods. However, the rainfall bucket tip-time series is not usually available for a long period of time. Most of the longer series of rainfall data are available in accumulated form, hourly or sub-hourly, rather than in tip-time series format. Moreover, the above DSPP models cannot be fitted directly to data collected in aggregated form, such as hourly rainfall, using the maximum likelihood method. Therefore it is useful to look for alternative models, based on doubly stochastic processes, that can be used to model sub-hourly data collected in aggregated form. Motivated by the performance of the above class of doubly stochastic models, we seek to develop models with the same structure that can be used to analyse accumulated rainfall sequences at fine time scales.

### 2.2 Model formulation

The point process model we propose here is constructed from a special class of DSPP where the arrival rate of the point process is governed by a finite-state irreducible Markov chain. See

for example, [Ramesh \(1995, 1998\)](#) and [Davison and Ramesh \(1996\)](#). Suppose that the rain cells, at time  $T_i$  say, arrive according to a two-state DSPP where the arrival rate is switching between the high intensity ( $\phi_2$ ) and low intensity ( $\phi_1$ ) states at random times controlled by the underlying Markov chain. The transition rates of the Markov chain are  $\lambda$  (for  $1 \rightarrow 2$ ) and  $\mu$  (for  $2 \rightarrow 1$ ) respectively. Therefore, the parameters of the cell arrival process  $M(t)$  are specified by the arrival rate matrix  $\mathbf{L}$  of the cell occurrences and the infinitesimal generator  $\mathbf{Q}$  of the underlying Markov chain, where

$$\mathbf{L} = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix} \quad \text{and} \quad \mathbf{Q} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}.$$

Each rain cell generated by the process has a random lifetime of length  $L$  which is taken to be exponentially distributed with parameter  $\eta$  and independent of the lifetime of other cells. A cell originating at time  $T_i$  will be active for a period of  $L_i$  and terminates at time  $T_i + L_i$ . When the cell at  $T_i$  is active instantaneous random pulses of rainfall occur, during  $[T_i, T_i + L_i)$ , at times  $T_{ij}$  according to a Poisson process at rate  $\xi$ . This process of instantaneous pulse arrival terminates with the cell lifetime. Therefore, each cell of the point process generates a series of pulses during its lifetime, and associated with each pulse is a random rainfall depth,  $X_{ij}$ . As a result, the process  $\{T_{ij}, X_{ij}\}$  takes the form of a marked point process ([Cox and Isham, 1980](#)). In our derivation of model properties later on in Section 2.3, we treat the pulses in distinct cells as independent but allow those within a single cell to be dependent. We shall refer to this model as a doubly stochastic pulse (DSP) model. A diagram showing the structure of this process is given in Figure 1. It is clear from this that the formulation of our model is very similar to that of [Cowpertwait et al. \(2007\)](#), but the difference lies in the mechanism for the cell arrivals. Since this process operates in continuous time, we integrate the DSP process to obtain rainfall depths over discrete disjoint time intervals and use their moment properties for model fitting and assessment.

### 2.3 Moment properties of the pulse process

It follows from the structure of the point process that the moment properties of the DSP process are functions of those of the cell arrival process. Therefore, we shall first review the properties of the cell arrival process before moving onto derive the properties of the pulse process. The second-order moment properties of the cell arrival process  $M(t)$  can be obtained as functions of the model parameters and these are given below (see for example, [Ramesh \(1998\)](#)). The stationary distribution  $\boldsymbol{\pi} = (\pi_1, \pi_2)$  of the cell arrival process  $M(t)$  is obtained by solving  $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ , where  $\mathbf{0} = (0, 0)$ , and is given by  $\boldsymbol{\pi} = (\mu/(\lambda + \mu), \lambda/(\lambda + \mu))$ . Let  $\mathbf{1}$  be a column vector of ones, then the mean arrival rate of the cell process is given by

$$E(M(t)) = m = \boldsymbol{\pi}\mathbf{L}\mathbf{1} = \frac{\lambda\phi_2 + \mu\phi_1}{\lambda + \mu}.$$

The covariance density of the cell arrival process  $M(t)$  can be obtained as, for  $t > 0$ ,

$$c^M(t) = \frac{\lambda\mu}{(\lambda + \mu)^2}(\phi_1 - \phi_2)^2 e^{-(\lambda + \mu)t} = Ae^{-(\lambda + \mu)t} \quad (1)$$

which shows that its covariance decays exponentially with time.

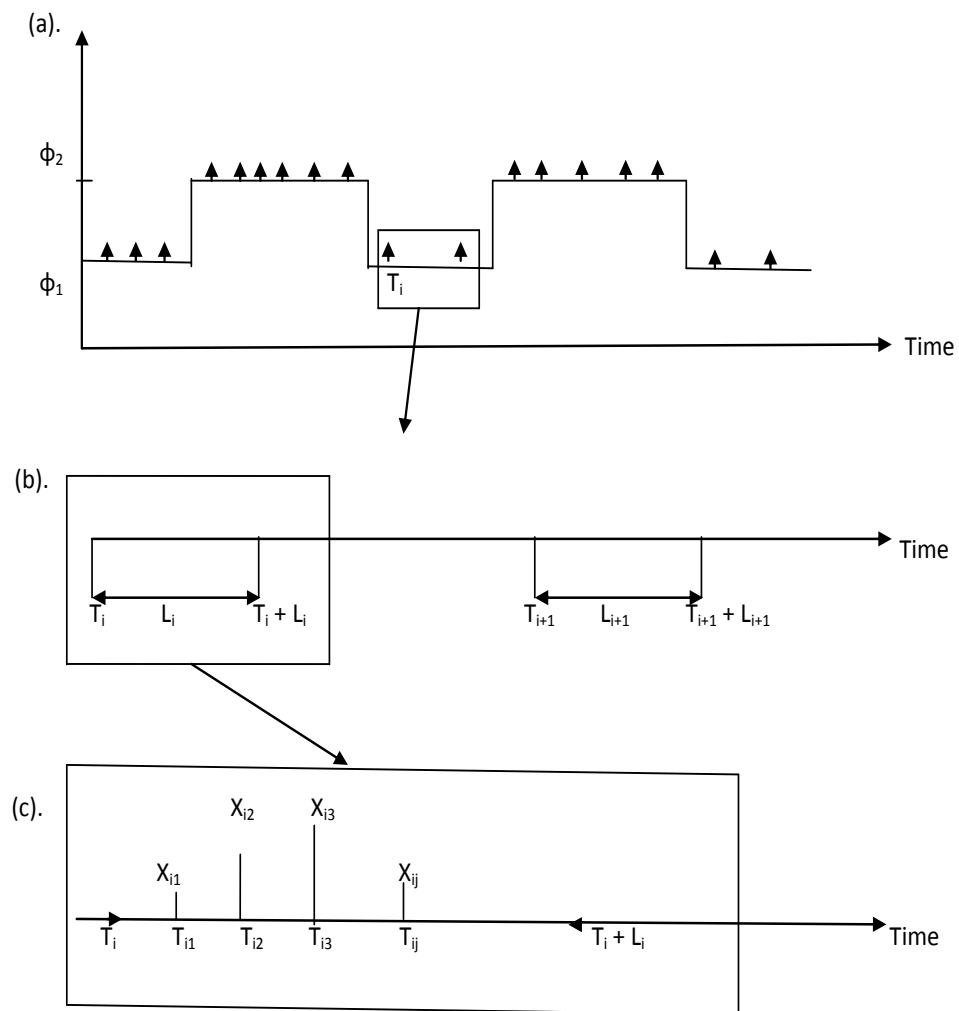


Figure 1: Schematic description of the DSP model: (a) the cell arrival process based on a two-state DSPP. (b) the cell lifetimes of the two cells at time  $T_i$  and  $T_{i+1}$ . (c) the pulse process in the cell that originates at time  $T_i$  and terminates at time  $T_i + L_i$ .

We shall now study the moment properties of the pulse arrival process and focus our attention on deriving an expression for its covariance density. These moment properties are required to derive the statistical properties of the aggregated rainfall process later in Section 2.4. The lifetimes  $L_i$  of the rain cells, under the DSP model framework, are assumed to be exponentially distributed with parameter  $\eta$  and hence we have  $E(L_i) = 1/\eta$ . Let us take  $N(t)$  as the counting process of pulse occurrences from all cells generated by the process  $M(t)$ . An active cell generates a series of instantaneous pulses at Poisson rate  $\xi$  during its lifetime and therefore the mean number of pulses per cell is  $\xi/\eta$ . As noted earlier, the arrival rate of the cell process is  $m$  and hence the mean arrival rate of the pulse process is given by  $E(N(t)) = m\xi/\eta$ .

It is well known that the covariance density of a point process can be expressed as a function of its product density. As shown by [Cox and Isham \(1980\)](#), the product density of the point process  $N(t)$  at distinct time points  $t_1, t_2, \dots, t_k$  for  $k=1, 2, 3, \dots$  can be written as

$$p_k(t_1, \dots, t_k) dt_1 \dots dt_k = P\{dN(t_1) = dN(t_2) = \dots = dN(t_k) = 1\}.$$

We consider two distinct pulses at time  $t$  and  $t + u$  ( $u > 0$ ) which may come from the same cell or different cells generated by the process  $M(t)$ . For pulses that come from the same cell, the contribution to the product density  $p_2(t, t + u)$  is given by  $p_2(t, t + u) = E(N(t))\xi e^{-\eta u} = (m\xi^2/\eta)e^{-\eta u}$ . For two pulses at time  $t$  and  $t + u$  that come from different cells, with their cell origins at  $t - v$  and  $t + u - w$  respectively, the contribution to the product density  $p_2(t, t + u)$  of the pulse process  $N(t)$  is given by

$$p_2(t, t + u) = \xi^2 \int_{v=0}^{\infty} e^{-\eta v} \int_{w=0}^u e^{-\eta w} p_2^M(u + v - w) dw dv. \quad (2)$$

From equation (1), the product density of the cell arrival process  $M(t)$  becomes

$$p_2^M(u + v - w) = c^M(u + v - w) + m^2 = Ae^{-(\lambda + \mu)(u + v - w)} + m^2. \quad (3)$$

Substituting (3) in (2) and completing the integral shows that the contribution to the product density by two pulses that come from different cells is

$$p_2(t, t + u) = \left(\frac{m\xi}{\eta}\right)^2 (1 - e^{-\eta u}) + \xi^2 A \frac{[e^{-(\lambda + \mu)u} - e^{-\eta u}]}{\eta^2 - (\lambda + \mu)^2}. \quad (4)$$

The covariance density ([Cox and Isham, 1980](#)) of the pulse process  $N(t)$  for  $u \geq 0$  can be written in terms of its product density as

$$c(u) = \text{Cov}\{N(t), N(t + u)\} = \left(\frac{m\xi}{\eta}\right) \delta(u) + p_2(t, t + u) - \left(\frac{m\xi}{\eta}\right)^2,$$

where  $\delta(\cdot)$  is the Dirac delta function. Substitution from (4) and rearranging the terms in the above expression gives, after some algebra, the covariance density of this DSP process  $N(t)$  for  $u \geq 0$  as

$$c(u) = \left(\frac{m\xi}{\eta}\right) \delta(u) + A_1 e^{-(\lambda + \mu)u} + [B_2 - B_1] e^{-\eta u}, \quad (5)$$

where  $A_1 = \xi^2 A / (\eta^2 - (\lambda + \mu)^2)$ ,  $B_1 = (\xi m / \eta)^2 + \xi^2 A / (\eta^2 - (\lambda + \mu)^2)$  and  $B_2 = (\xi^2 m / \eta)$ . In the above expression,  $A_1$  and  $B_1$  correspond to the contribution from pulses generated by different cells whereas  $B_2$  corresponds to the contribution from different pulses within the same cell, where the depths of these pulses may be dependent.

## 2.4 Moment properties of the aggregated rainfall

In most applications, the rainfall data are usually available in aggregated form in equally spaced time intervals of fixed length. The DSP process we have developed, however, evolves in continuous time. We now, therefore, derive mathematical expressions for the moment properties of the aggregated rainfall arising from the DSP process. These expressions are useful to describe the properties of the accumulated rainfall and can be used for model fitting and assessment.

Let  $Y_i^{(h)}$  be the total amount of rainfall in disjoint time intervals of fixed length  $h$ , for  $i = 1, 2, \dots$ . We can express this as

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

where  $X(t)$  is the depth of a pulse at time  $t$ . Without assuming any distribution for the pulse depth, let  $E[X(t)] = \mu_X$  be the mean depth of the pulses. Then the mean of the aggregated rainfall in the intervals can be written as

$$E[Y_i^{(h)}] = \int_{(i-1)h}^{ih} E(X(t)) dN(t) = \left(\frac{m\xi}{\eta}\right) \mu_X h. \quad (6)$$

Using the well known Campbell's theorem from Daley and Vere-Jones (2007), and utilizing the covariance density of the pulse arrival process given in (5), we can now work out the variance and autocovariance function of the aggregated rainfall process. Following this, we have

$$\begin{aligned} \text{Var}[Y_i^{(h)}] &= \int_0^h \int_0^h \text{Cov}[X(s)dN(s), X(t)dN(t)] \\ &= \int_0^h E(X^2)E[dN(t)] + 2 \int_0^h \int_s^h E[X(s)X(t)] \text{Cov}[dN(s), dN(t)], \end{aligned}$$

where the double integral is separated into two parts to account for the cases  $s = t$  and  $s \neq t$ . In addition, we need to distinguish whether the pulses at times  $t$  and  $s$  belong to the same cell or come from different cells. For contributions to  $\text{Cov}[dN(s), dN(t)]$ , when pulses come from different cells, the multiplier  $E[X(s)X(t)]$  in the above expression becomes  $\mu_X^2$ . We shall write this multiplier as  $E[X_{ij}X_{ik}]$  when pulses come from the same cell. This will allow us to accommodate some within-cell depth dependence. However, it is assumed that any two pulses within a cell, regardless of their location within the cell, have the same expected product of depths. Under this setting, the variance function becomes

$$\begin{aligned} \text{Var}[Y_i^{(h)}] &= E(X^2) \left(\frac{m\xi}{\eta}\right) h \\ &\quad + 2\mu_X^2 A_1 \psi_1(\lambda + \mu) + 2[E[X_{ij}X_{ik}] B_2 - B_1 \mu_X^2] \psi_1(\eta), \end{aligned} \quad (7)$$

where  $\psi_1(\lambda + \mu) = [(\lambda + \mu)h - 1 + e^{-(\lambda + \mu)h}] / (\lambda + \mu)^2$  and  $\psi_1(\eta) = [\eta h - 1 + e^{-\eta h}] / \eta^2$ .

Following a similar approach we can derive the autocovariance function for the aggregated rainfall in two distinct intervals. Again by distinguishing the contributions from pulses within the same cell, we derived an expression given as, for  $k \geq 1$ ,

$$\begin{aligned} \text{Cov}[Y_i^{(h)}, Y_{i+k}^{(h)}] &= \int_{kh}^{(k+1)h} \int_0^h \text{Cov}[X(s)dN(s), X(t)dN(t)] \\ &= \int_{kh}^{(k+1)h} \int_0^h E[X(s)X(t)] \text{Cov}[dN(s), dN(t)] \\ &= \mu_X^2 A_1 \psi_2(\lambda + \mu) + [E[X_{ij}X_{ik}] B_2 - B_1 \mu_X^2] \psi_2(\eta), \end{aligned} \quad (8)$$



where  $\psi_2(\lambda + \mu) = e^{-(\lambda + \mu)(k-1)h} [1 - e^{-(\lambda + \mu)h}]^2 / (\lambda + \mu)^2$  and  $\psi_2(\eta) = e^{-\eta(k-1)h} [1 - e^{-\eta h}]^2 / \eta^2$ . When considering the special case where all pulse depths are independent  $E(X_{ij}X_{ik})$  can be replaced by  $\mu_X^2$  in equations (7) and (8).

Although the second-order properties capture the characteristics of the process well in most point process applications, higher moments may provide improved results in terms of reproduction of the properties of interest. To this end, following [Cowpertwait et al. \(2007\)](#), we shall incorporate the third-order moment in our analysis. The derivation of the third moment  $E[(Y^{(h)})^3]$  follows a similar approach to that of the second moment, but becomes more complex algebraically. Therefore only an outline of the derivation is given along with the final expression in Appendix A. The third moment about the mean  $\mu_3^{(h)}$  is

$$\begin{aligned} \mu_3^{(h)} &= E \left[ \left( Y_i^{(h)} - E \left[ Y_i^{(h)} \right] \right)^3 \right] \\ &= E \left[ \left[ Y_i^{(h)} \right]^3 \right] - 3E \left[ Y_i^{(h)} \right] \text{Var} \left[ Y_i^{(h)} \right] - \left[ E \left[ Y_i^{(h)} \right] \right]^3. \end{aligned}$$

From this the coefficient of skewness of the aggregated rainfall process is shown to be

$$\kappa^{(h)} = \frac{E \left[ \left( Y_i^{(h)} - E \left[ Y_i^{(h)} \right] \right)^3 \right]}{E \left[ \left( Y_i^{(h)} - E \left[ Y_i^{(h)} \right] \right)^2 \right]^{3/2}} = \frac{\mu_3^{(h)}}{\left[ \text{Var} \left[ Y_i^{(h)} \right] \right]^{3/2}}. \quad (9)$$

### 3 Model fitting and assessment

We shall explore the application of the proposed DSP model in the analysis of fine-scale rainfall data and assess how well it reproduces the properties of the rainfall over a range of sub-hourly resolutions. We aim to do this using 69 years (1931-1999) of 5-minute rainfall data from Bochum in Germany.

In this analysis, we consider the special case where the pulse depths  $X_{ij}$  are independent random variables that follow an exponential distribution with parameter  $\theta$ . Therefore, we have  $\mu_X = E[X(t)] = 1/\theta$ . The data, recorded at the 5-minute aggregation level, allow us to fit the model over a range of sub-hourly accumulations. We shall make use of the mathematical expressions for the moment properties of the accumulated rainfall in our model fitting. There are 7 parameters in our model and we estimate 6 of them by the method of moments approach using the observed and theoretical values of these properties. The parameter  $\mu_X = 1/\theta$  is estimated separately for each month from the sample mean of rainfall depth by using the equation, which follows from equation (6),

$$\mu_X = \left( \frac{\eta}{m\xi} \right) \bar{x}, \quad (10)$$

where  $\bar{x}$  is the estimated average of hourly rainfall for each month.

Although we employed the method of moments to estimate the other parameters, which essentially equates the sample moments to theoretical moments from the model, the estimation can be done in different ways. In this application, we constructed an objective function as the sum of squares of differences between the sample and theoretical values of the moment properties

at different aggregation levels and then minimized it using standard optimisation routines. This is carried out separately for each month by considering the data for a month as realisations of a stationary point process. Essentially the process of model fitting involves calculating the empirical mean, variance, correlation, coefficient of variation and skewness from the observed data at each aggregation level and matching these with the corresponding theoretical values, calculated using equations (6) to (9), for a given set of parameter values. The role of the objective function and the optimiser employed was to find the best possible match using a minimum error criterion. We used the statistical software environment R for the optimisation and for the simulation of the process (R Development Core Team, 2011). A number of options were available for the objective function depending on the application, but we used a form of weighted sum of squares. We employed the routines “optim” and “nlminb” in R for parameter estimation in our analysis. The following subsections describe two different methods used to estimate the first 6 parameters of the model and discuss the results produced. Once these parameter estimates were determined, equation (10) was used to estimate the final parameter  $\mu_X$ .

### 3.1 Estimation using second-order moments

The first 6 parameters of the model  $\lambda$ ,  $\mu$ ,  $\phi_1$ ,  $\phi_2$ ,  $\eta$  and  $\xi$  were estimated using the following dimensionless functions; the coefficient of variation  $\nu(h)$  and the autocorrelation  $\rho(h)$  at lag 1 of the aggregated rainfall process. Explicitly, these are

$$\nu(h) = \frac{\text{E} \left[ \left( Y_i^{(h)} - \text{E} \left[ Y_i^{(h)} \right] \right)^2 \right]^{1/2}}{\text{E}(Y_i^{(h)})}, \quad \rho(h) = \text{Corr} \left[ Y_i^{(h)}, Y_{i+1}^{(h)} \right]. \quad (11)$$

We need at least 6 sample properties of the aggregated process to fit the model and we employed 8 properties in our estimation. They are  $\nu(h)$  and  $\rho(h)$  at  $h=5, 20, 30$  and 60 minutes aggregation levels. The estimates of the functions from the empirical data, denoted by  $\hat{\nu}(h)$  and  $\hat{\rho}(h)$  (for  $h=1/12, 1/3, 1/2$  and 1 hours), were calculated for each month using 69 years of 5-minute rainfall series accumulated at appropriate time scales. Parameter estimates can be obtained by using an objective function constructed from the sum of squares of differences between the sample values and their corresponding theoretical values of the proposed model. The estimated values of parameters  $\{\hat{\lambda}, \hat{\mu}, \hat{\phi}_1, \hat{\phi}_2, \hat{\eta}$  and  $\hat{\xi}\}$  for each month were obtained by minimising the weighted sum of squares of dimensionless functions, as given below in equation (12), using standard optimisation routines:

$$\sum_{h=\frac{1}{12}, \frac{1}{3}, \frac{1}{2}, 1} \left[ \frac{1}{\text{Var}(\hat{\nu}(h))} (\hat{\nu}(h) - \nu(h))^2 + \frac{1}{\text{Var}(\hat{\rho}(h))} (\hat{\rho}(h) - \rho(h))^2 \right]. \quad (12)$$

In the above expression,  $\nu(h)$  and  $\rho(h)$  are theoretical values as given by equations (6)-(11) whereas  $\hat{\nu}(h)$  and  $\hat{\rho}(h)$  are calculated for each month from the 69 years of data. In addition, following Jesus and Chandler (2011), the weights in the objective function were taken as the reciprocal of the variance of the empirical values of the functions calculated separately for each of the 69 years. We also experimented with other objective functions but this was found to give better results. The above function was minimised separately for each month to obtain estimates of the model parameters. We used the simplex algorithm of Nelder and Mead (1965) for the optimisation, since it does not require the calculation of derivatives.

The estimated parameters were then used to calculate the fitted values of the various theoretical properties. The observed and fitted values of the mean, standard deviation, coefficient of variation and lag 1 autocorrelation of the aggregated rainfall are displayed in Figures 2-5. In almost all cases near perfect fits, and in some cases exact, were obtained for all properties with the exception of the lag 1 autocorrelation. The mean gave an excellent fit at all time scales. The standard deviation and coefficient of variation showed near perfect fits for most months, at all time scales, with small deviations from the perfect fit during the summer months. The autocorrelation was reproduced well at smaller aggregation levels, however there appeared to be a slight underestimation at larger aggregations for all months. One point to note here is that  $h=10$  minutes aggregation was not used in the fitting but the model has certainly reproduced all the properties well for this time-scale. This reveals that the model is capable of producing estimates of quantities not used in the fitting which adds strength to this DSP modelling framework.

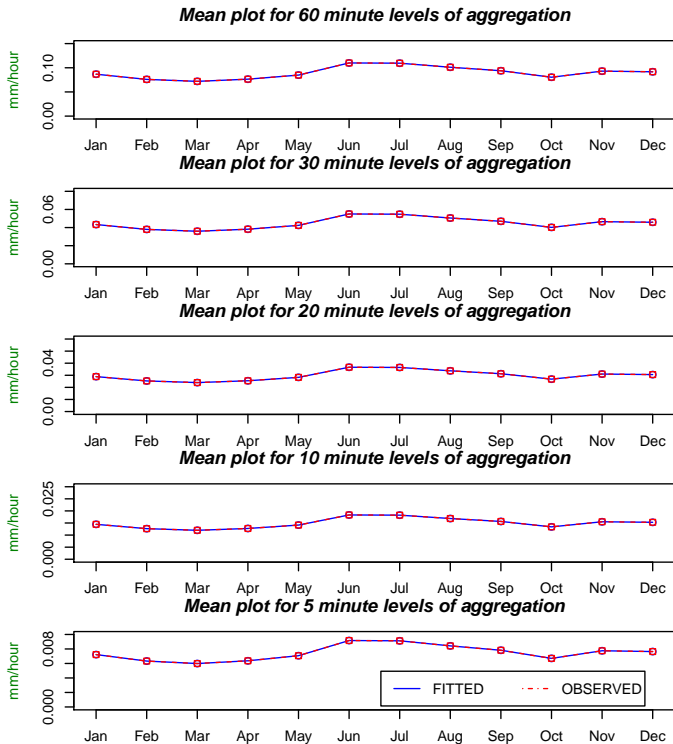


Figure 2: Observed and fitted values of the mean of the aggregated rainfall for the DSP model at  $h=5, 10, 20, 30, 60$  minutes aggregations, using second-order moments in estimation.

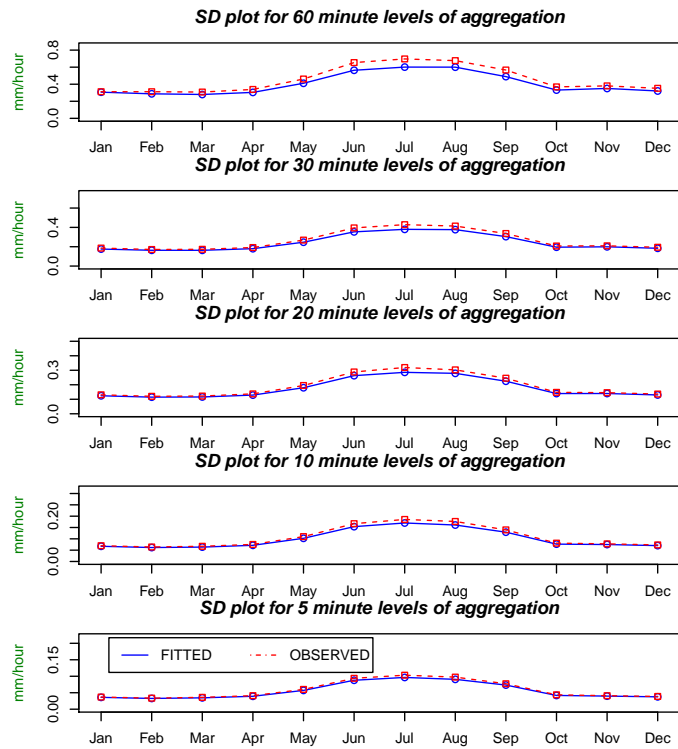


Figure 3: Observed and fitted values of the standard deviation of the aggregated rainfall for the DSP model at  $h=5, 10, 20, 30, 60$  minutes aggregations, using second-order moments in estimation.

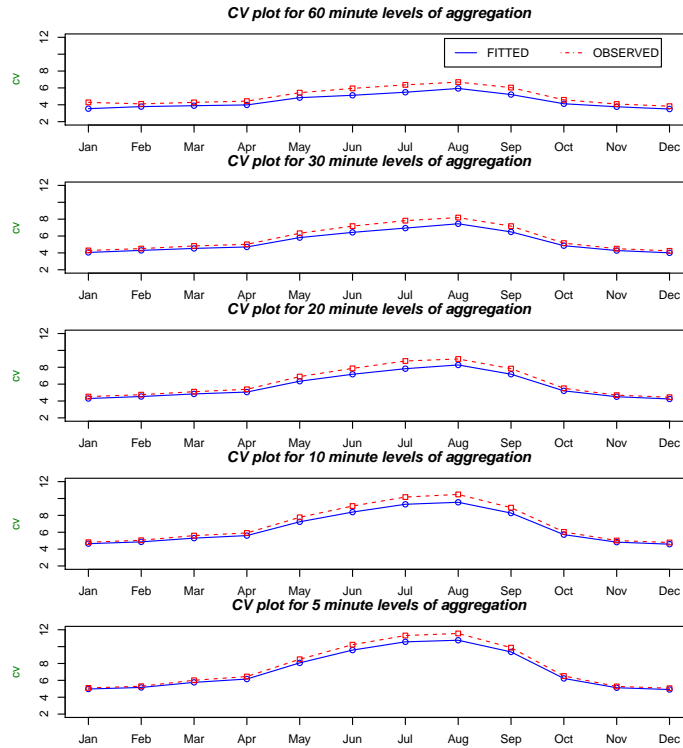


Figure 4: Observed and fitted values of the coefficient of variation of the aggregated rainfall for the DSP model at  $h=5, 10, 20, 30, 60$  minutes aggregations, using second-order moments in estimation.

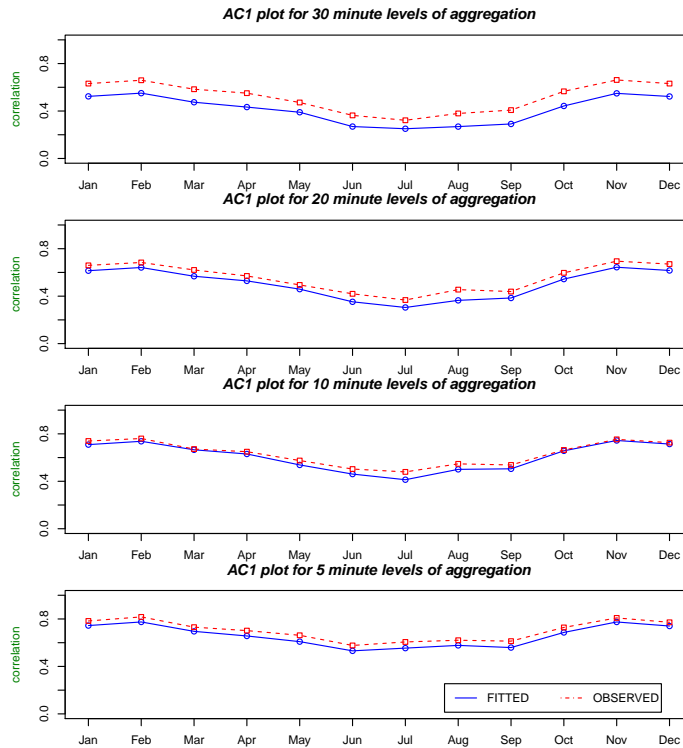


Figure 5: Observed and fitted values of the autocorrelation (lag one) of the aggregated rainfall for the DSP model at  $h=5, 10, 20, 30$  minutes aggregations, using second-order moments in estimation.

### 3.2 Estimation incorporating third-order moments

We now analyse results which incorporated third-order moments in the fitting, since this has been found useful in modelling rainfall (Cowpertwait et al., 2007). With the coefficient of skewness  $\kappa^{(h)}$  given in equation (9), we now have three model functions to utilise in estimation. Under the model framework, we have six parameters to estimate using the moment method and we consider nine properties to include the coefficient of skewness. These are the coefficient of variation  $\nu(h)$ , the autocorrelation  $\rho(h)$  at lag 1 and the coefficient of skewness  $\kappa(h)$ , all at three different aggregation levels of  $h = 5, 10$  and  $20$  minutes. The following objective function, which incorporates third-order moments in parameter estimation, was used in our analysis:

$$\sum_{h=\frac{1}{12}, \frac{1}{6}, \frac{1}{3}} \left[ \frac{1}{\text{Var}(\hat{\nu}(h))} (\hat{\nu}(h) - \nu(h))^2 + \frac{1}{\text{Var}(\hat{\rho}(h))} (\hat{\rho}(h) - \rho(h))^2 + \frac{1}{\text{Var}(\hat{\kappa}(h))} (\hat{\kappa}(h) - \kappa(h))^2 \right].$$

This objective function was minimised, separately for each month, to obtain estimates of the model parameters. These are given in Table 1. Estimated parameters showed some variation across months. Values of  $\hat{\mu}$  were larger for summer months showing smaller mean sojourn times ( $1/\mu$ ) in the higher rainfall intensity state. The estimates  $\hat{\phi}_2$  were also higher, in general, for the summer months and showed that the cell arrival rates in state 2 vary from about 2.4 to 7.8 cells per hour. The pulse arrival rate  $\hat{\xi}$  also showed variation across months from about 223 to 290 pulses per hour throughout the year. The mean duration of cell lifetime ( $1/\eta$ ) fell between 8 to 24 minutes. It was noticeable that the cell durations were shorter in summer months (8-10 minutes) when compared with other months. This is consistent with the nature of the summer rainfall, as they are mostly generated from thunder storms of high intensity and shorter duration. The estimates also showed that the mean depth of the pulses ( $\mu_x$ ) tend to be larger in summer months with the highest value in July.

Table 1: Parameter estimates for the DSP model incorporating third-order moments in estimation.

Month	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\eta}$	$\hat{\xi}$	$\hat{\mu}_x$
JAN	0.0306	2.6900	0.0998	3.4795	3.5699	263.5150	0.0085
FEB	0.0161	1.6124	0.0819	2.9469	3.0257	264.8286	0.0079
MAR	0.0090	5.0823	0.1282	5.6073	5.7007	289.9880	0.0102
APR	0.0245	4.8790	0.0764	5.0201	5.1021	246.6266	0.0156
MAY	0.0529	7.0143	0.0400	7.1321	7.1555	239.9341	0.0272
JUN	0.0411	7.7027	0.0331	7.8086	7.8255	237.8311	0.0486
JUL	0.0199	6.7546	0.0318	6.8875	6.8837	245.2713	0.0591
AUG	0.0197	6.1291	0.0276	6.3391	6.3118	265.9815	0.0502
SEP	0.0491	6.9914	0.0205	7.1563	7.1348	246.2618	0.0387
OCT	0.0147	1.9679	0.0362	2.4563	2.4956	223.6488	0.0166
NOV	0.1154	3.7691	0.0430	4.1023	4.1339	269.6375	0.0087
DEC	0.0234	1.8008	0.0860	2.9616	3.0464	227.5842	0.0099

Figures 6 to 9 show the corresponding results when the third-order moments are incorporated into the parameter estimation process. These clearly show an improvement over the earlier results of the method that used the moments up to the second-order. Although the same model is used here, the parameter estimates are obviously different from the earlier values due

to the fact that the coefficient of skewness was used in estimation for this second method. In most cases a near perfect fit was obtained and in some cases an exact fit was obtained when the third-order moments were incorporated. The mean, standard deviation and coefficient of variation have all been reproduced remarkably well at all aggregation levels, including those that were not used in fitting ( $h=30, h=60$  minutes). Comparing these with earlier results reveals that incorporating third-order moments in estimation has certainly produced much better results for most of the properties.

The mean appears to be estimated equally well by both methods. The standard deviation and coefficient of variation were clearly in better agreement with their empirical counterparts at all aggregations when third-order moments are used. This is visible in the plots, particularly for the summer months where the first method consistently showed a slight underestimation. The autocorrelation at lag one, however, was an exception. Although it was estimated well at small aggregation levels, the fit got worse at higher levels of aggregation, especially at those that were not used in fitting, showing consistent underestimation. Nevertheless, the observed and fitted values of the coefficient of skewness were in very good agreement in Figure 10. Here again the results showed a near perfect fit at all aggregation levels. In addition to this, as we will see in Section 3.3, incorporation of third-order moments in estimation reproduced the extremes better than the earlier method.

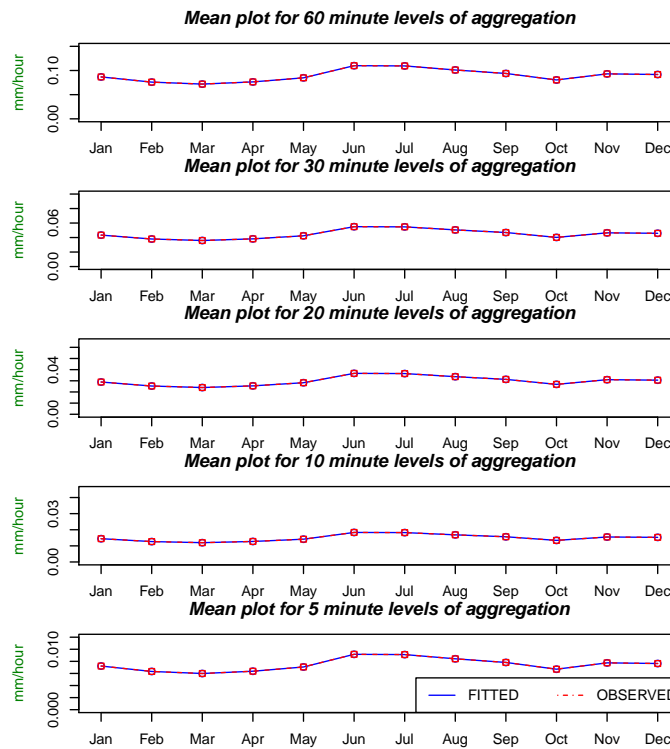


Figure 6: Observed and fitted values of the mean of the aggregated rainfall for the DSP model at  $h = 5, 10, 20, 30, 60$  minutes aggregations, using third-order moments in estimation.

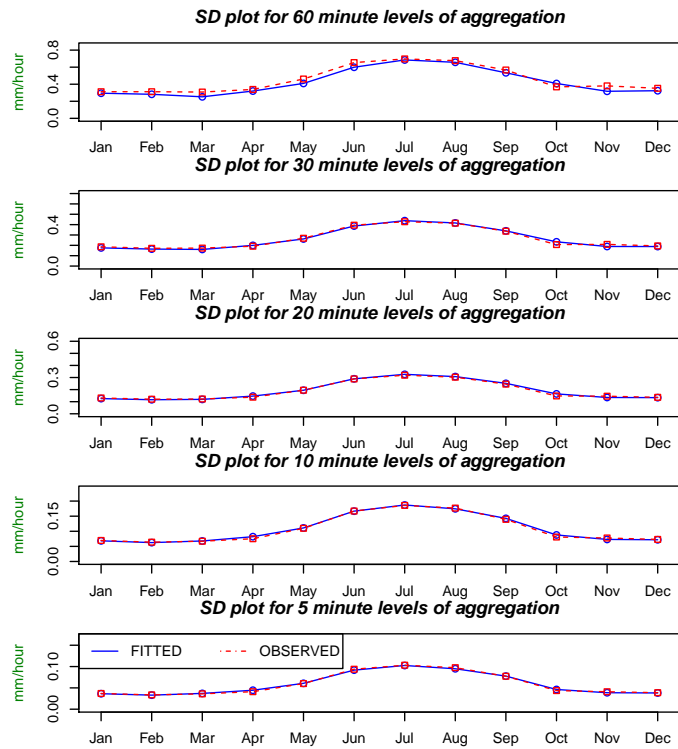


Figure 7: Observed and fitted values of the standard deviation of the aggregated rainfall for the DSP model at  $h = 5, 10, 20, 30, 60$  minutes aggregations, using third-order moments in estimation.



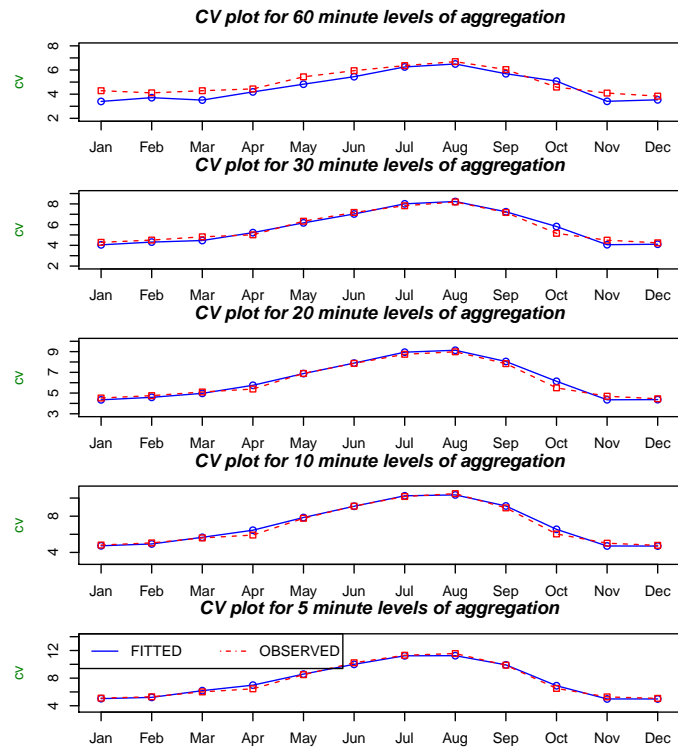


Figure 8: Observed and fitted values of the coefficient of variation of the aggregated rainfall for the DSP model at  $h = 5, 10, 20, 30, 60$  minutes aggregations, using third-order moments in estimation.

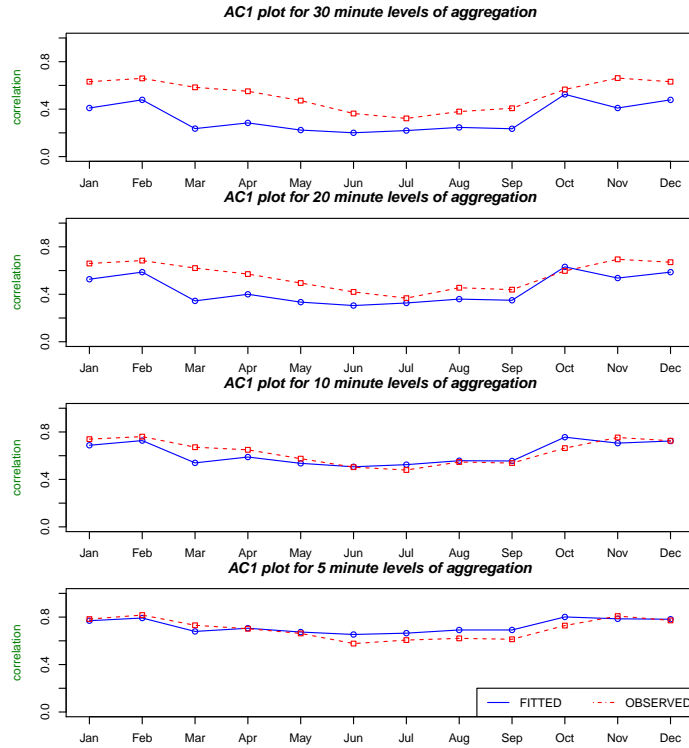


Figure 9: Observed and fitted values of the autocorrelation (lag one) of the aggregated rainfall for the DSP model at  $h = 5, 10, 20, 30$  minutes aggregations, using third-order moments in estimation.

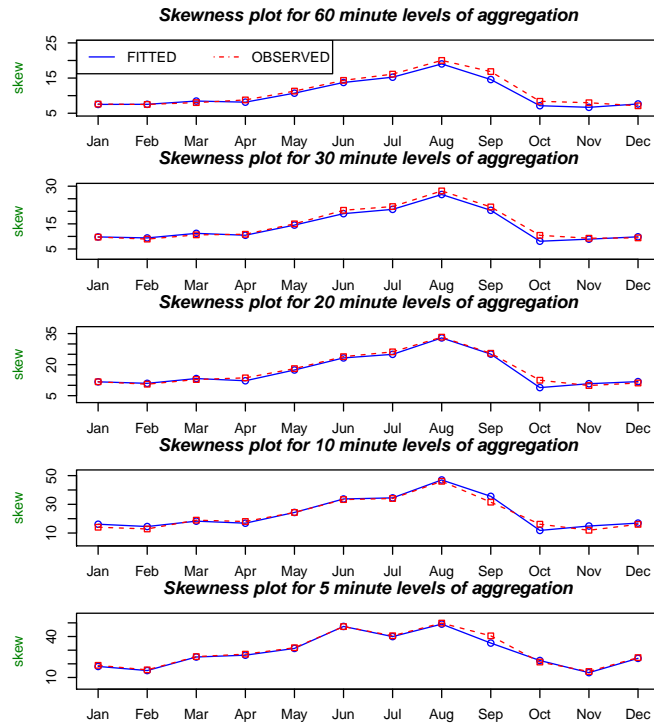


Figure 10: Observed and fitted values of the coefficient of skewness of the aggregated rainfall for the DSP model at  $h = 5, 10, 20, 30, 60$  minutes aggregations, using third-order moments in estimation.

### 3.3 Extremes and proportion of dry periods

In many hydrological applications, more emphasis is placed upon a stochastic model's ability to reproduce the properties of extreme rainfall rather than the usual moment properties. One good example arises in urban drainage modelling within the context of flood estimation. One can find many other examples of this in analyses involving environmental data, see for example [Ramesh and Davison \(2002\)](#), [Davison and Ramesh \(2000\)](#) and [Leiva et al. \(2016\)](#). In addition, knowledge of the extremes enables us to assess the risk associated with hydrological systems. In view of this, we shall evaluate the performance of our proposed model in capturing the properties of extreme rainfall. In this regard, we compare the extreme values of the 69 years of observed rainfall data with those generated by the proposed DSP model.

The annual maxima of the empirical data from the observed 69 year long historical record were extracted, ordered and plotted against the corresponding Gumbel reduced variates at each aggregation level. Fifty copies of the 5-minute rainfall series, each 69 years long, were then simulated from the fitted model. Each copy of the simulated data was subsequently aggregated to generate 1, 12 and 24 hour rainfall series. The annual maxima of each of the 50 simulated series, at each aggregation level, were extracted and ordered to make up the interval plots against the corresponding Gumbel reduced variates. These were superimposed on the corresponding Gumbel reduced variate plots of the empirical data for comparison.

Figure 11 shows the results for  $h=5$  minutes (top panel) and  $h=60$  minutes (bottom panel) aggregations whereas the results of  $h=12$ , 24 hours are displayed in Figure 12. The red solid circles show the mean of the maxima from the 50 simulations and the blue squares connected by the solid line show the empirical annual maximum values. At  $h=5$  minutes aggregation level the model vastly underestimates the extremes. When  $h=60$  minutes there was evidence of over-estimation at the lower end and underestimation at the upper end of the reduced variates. Nevertheless, there was evidence of substantial improvement at larger values of  $h$ . Clearly there was very good agreement at  $h=12$ , 24 hour aggregations, as all of the empirical values fell within the range of the simulated values. Furthermore, the mean of the simulated annual maxima fell reasonably close to the empirical annual maxima, at  $h=12$  and 24, for much of the period. It is encouraging to note that this was the case for the values corresponding to the return periods from about 5 years to 100 years. Hence, the proposed model, although underestimating the extremes at sub-hourly aggregation, appears to capture the extremes well at larger aggregations. The estimation of extreme values at smaller time-scales is a common problem for most stochastic models for rainfall, and our results concur with the findings of previous published studies, see for example [Cowpertwait et al. \(2007\)](#) or [Verhoest et al. \(1997\)](#).

In order to assess the performance of the two estimation methods in reproducing extremes, we compared the ordered empirical annual maxima with those generated by the parameter estimates of the two methods. Table 2 shows the means of the ordered annual maxima from 50 simulations for the two methods, together with the ordered empirical annual maxima, at 12-hour and 24-hour aggregation levels. These are compared for a range of reduced Gumbel variates covering the values that correspond to the return period from 5 years to 100 years. Note that the interval plots based on the simulations, using the method which incorporates third-order moments in estimation (Method 2), are shown in Figure 12 for the whole period. Results presented in Table 2 show that Method 2 performed much better than the method that used second-order moments in estimation (Method 1), at both aggregation levels. The above result and Figure 12 clearly show that Method 2 outperformed Method 1 and reproduced extremes well at higher aggregations.

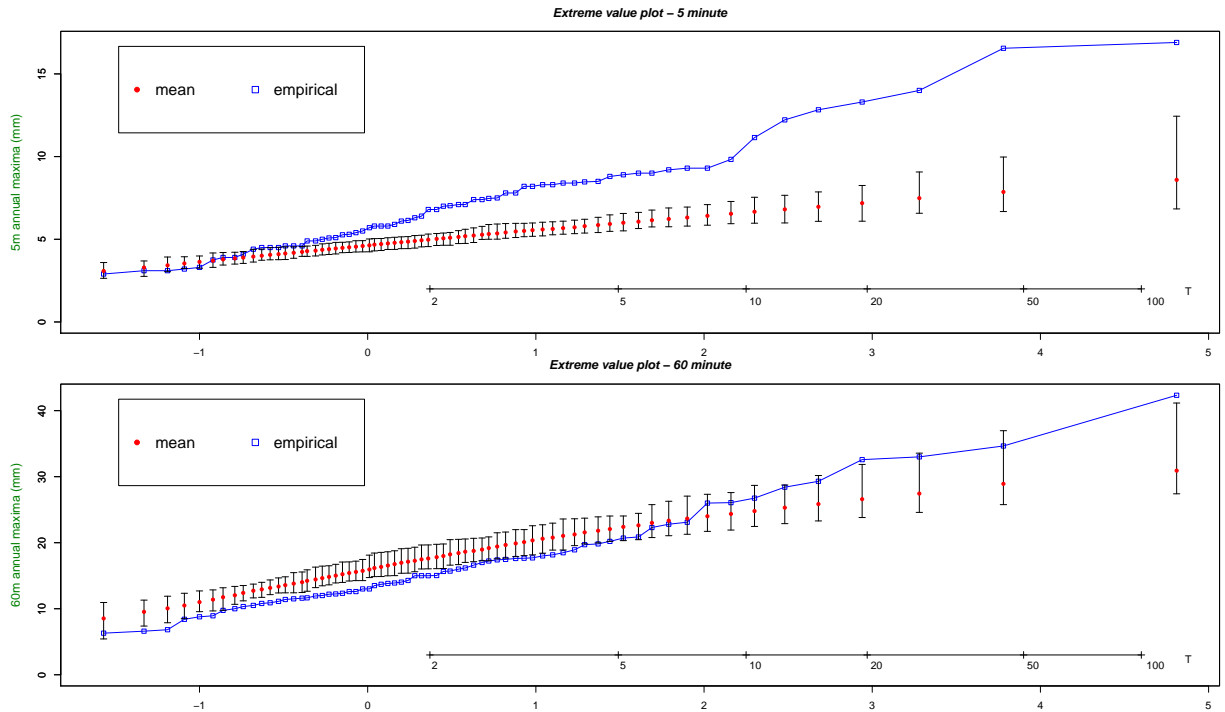


Figure 11: Ordered annual maxima of the 5-minute (top panel) and one-hour (bottom panel) aggregated rainfall plotted against the reduced Gumbel variate. Empirical annual maximum values are shown as blue squares connected by a solid line. Interval plots based on annual maxima of 50 simulations, each 69 years long, are also shown. The red solid circles are the mean of the 50 simulated maxima.

Another property of interest to hydrologists is the proportion of intervals with little or no rain. This will help to quantify the proportion of dry periods. Very often the gauge rainfall data are recorded in a rounded form (to the nearest 0.1 mm) and hence, following [Cowpertwait et al. \(2007\)](#), we calculate the proportion of rainfall below a small threshold ( $\hat{p}\{Y_i^h < \delta\}$  for  $\delta > 0$ ) instead of the actual proportion of dry intervals ( $\delta = 0$ ). Therefore by choosing smaller values of  $\delta$  we can provide approximate estimates of the proportion of dry intervals.

The proportion of intervals below a given threshold were calculated for each month, using 50 samples of 69 years of simulated 5-minute data, for each of the 5-minute, 1 hour and 24 hour aggregations ( $h = 1/12, 1, 24$ ). The mean of the 50 values was calculated, at each aggregation, for each calendar month. The observed proportions for the historical data and the average of the simulated values, from the fitted model, for different thresholds are given in Table 3. In order to find a good estimate of the proportion of dry periods at  $h = 1/12$  aggregation level, a small threshold of  $\delta = 0.05$  mm was used. For the hourly rainfall, two threshold values of  $\delta = 0.05, 0.1$  mm were used. Higher threshold values ( $\delta = 0.5, 2$  mm) were used for the daily rainfall, as occasional light rain during the day may cause discrepancies between the observed and simulated proportions.

The numerical results presented in Table 3 show that, in general, the model reproduces the observed proportions well at 5-minute aggregation, although it tends to over-estimate slightly. However, the differences we observed are between 0.01 for February and 0.022 for June. At the hourly aggregation level, the model tended to consistently underestimate the proportions each month when  $\delta = 0.05$  mm. However, when the threshold was increased to 0.1 mm the observed

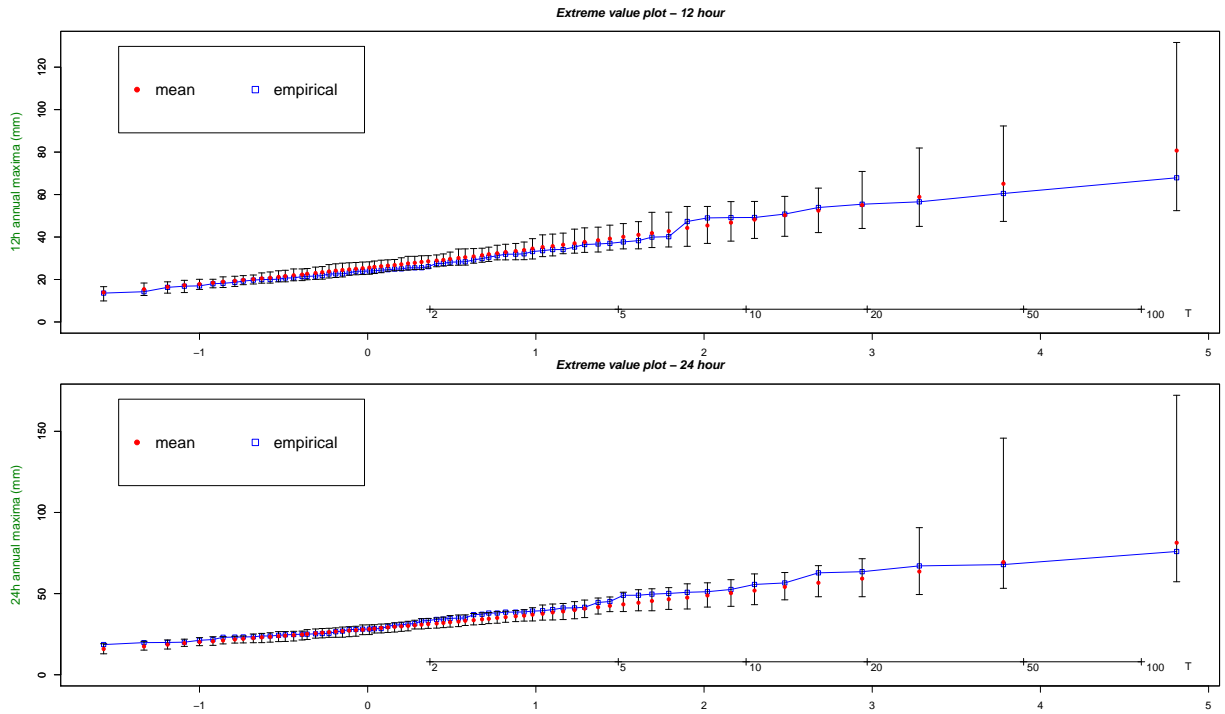


Figure 12: Ordered annual maxima of the 12-hour (top panel) and 24-hour (bottom panel) aggregated rainfall plotted against the reduced Gumbel variate. Empirical annual maximum values are shown as blue squares connected by a solid line. Interval plots based on annual maxima of 50 simulations, each 69 years long, are also shown. The red solid circles are the mean of the 50 simulated maxima.

and simulated proportions were in good agreement, except for October, with the summer month June again showing the largest difference of 0.027. Finally, at a daily level, there was consistent over-estimation when the threshold was set at 0.5 mm. When the threshold increased to 2 mm the observed and simulated proportions became very close. This indicates that the frequency of very light rain events, over a period of a day, was greater in the observed data. The differences tended to be larger in the summer months and smaller in the winter months, with June once again showing the highest difference of 0.026 and February showing the smallest difference of 0.012.

Table 2: Ordered empirical annual maxima for the 12-hour and 24-hour aggregated rainfall and the corresponding ordered reduced Gumbel variates. The average annual maxima, based on 50 simulations of 69-year long series, for the two estimation methods are also shown.

Gumbel Reduced variate	12-hour aggregation			24-hour aggregation		
	Empirical maxima	Simulated maxima		Empirical maxima	Simulated maxima	
		Method 1	Method 2		Method 1	Method 2
1.44	37.05	22.53	39.25	45.12	24.55	42.50
1.52	37.70	22.86	40.10	48.96	24.91	43.37
1.61	38.30	23.27	41.04	49.00	25.39	44.37
1.70	39.95	23.65	41.86	49.68	25.81	45.46
1.80	40.16	24.10	42.78	50.1	26.27	46.52
1.90	47.32	24.65	44.29	50.81	26.87	47.54
2.02	49.00	25.19	45.43	51.17	27.41	48.86
2.16	49.15	25.77	46.80	52.55	27.93	50.30
2.31	49.20	26.39	48.20	55.65	28.60	51.86
2.48	50.81	27.07	50.24	56.58	29.40	54.14
2.68	53.90	28.09	52.44	62.80	30.47	56.66
2.94	55.45	29.44	55.02	63.50	31.96	59.26
3.28	56.58	31.04	58.89	67.08	33.39	63.60
3.78	60.50	33.47	65.08	67.90	35.68	69.42
4.81	67.90	37.99	80.71	75.95	40.40	81.35

Table 3: Proportion of rainfall below defined thresholds at different time scales. Here  $\hat{p}_O$  and  $\hat{p}_S$  represent the estimates of the proportions for the observed and simulated series.

Month	$p_o$	$p_s$	$p_o$	$p_s$	$p_o$	$p_s$	$p_o$	$p_s$	$p_o$	$p_s$
JAN	0.951	0.964	0.848	0.757	0.866	0.865	0.903	0.975	0.976	0.991
FEB	0.958	0.968	0.867	0.769	0.883	0.872	0.915	0.979	0.979	0.991
MAR	0.961	0.972	0.867	0.841	0.892	0.907	0.919	0.971	0.980	0.995
APR	0.960	0.972	0.878	0.821	0.893	0.877	0.917	0.974	0.979	0.991
MAY	0.964	0.980	0.898	0.867	0.908	0.891	0.924	0.969	0.974	0.991
JUN	0.962	0.984	0.892	0.856	0.902	0.875	0.916	0.958	0.968	0.994
JUL	0.965	0.987	0.895	0.892	0.907	0.902	0.919	0.950	0.969	0.992
AUG	0.968	0.986	0.906	0.886	0.916	0.898	0.927	0.957	0.969	0.993
SEP	0.965	0.984	0.899	0.873	0.909	0.891	0.923	0.964	0.972	0.993
OCT	0.959	0.971	0.887	0.793	0.899	0.859	0.921	0.978	0.975	0.991
NOV	0.952	0.964	0.852	0.730	0.872	0.848	0.905	0.978	0.973	0.989
DEC	0.949	0.966	0.849	0.748	0.869	0.875	0.902	0.975	0.972	0.992
threshold $\delta$ (mm)	0.05	0.05	0.05	0.05	0.1	0.1	0.5	0.5	2	2
scale(h)	1/12	1/12	1	1	1	1	24	24	24	24

### 3.4 Simulation study

Since the likelihood function of the model we proposed for the accumulated rainfall data was not available, we employed the moment method to estimate the parameters in our analysis. The method of moments simply equates sample moments from the observed data to the theoretical moments of the model being fitted to obtain estimates of the parameters. It is important to note that moment estimators, unlike maximum likelihood estimators, do not necessarily have the usual large sample properties leading to asymptotic results. Nevertheless, we carried out a simulation study to evaluate the statistical performance of the estimation method we employed in our analysis.

The observed data was a 69-year long accumulated rainfall series in 5-minute intervals and our stochastic point process model was fitted separately for each month. A month with 31 days had 616032 observations, although most of them were zero, especially for a summer month. The simulation study was carried out for a typical summer month, August, with simulated rainfall data over a period of length  $n_l = 20, 40, 60, 69$  years, as described in the algorithm given below.

- Simulate one hundred ( $n = 100$ ) sample series from the fitted values of the parameters ( $\theta$ ) for the month.
- Aggregate the simulated data in  $h = 5, 10, 20$  minutes intervals.
- Calculate their sample statistics: coefficient of variation  $\nu(h)$ , lag 1 autocorrelation  $\rho(h)$  and coefficient of skewness  $\kappa(h)$  for each value of  $h$ .
- Compute the moment estimates ( $\hat{\theta}$ ) for the simulated samples, using the objective function used in Section 3.2.
- Compute the mean, bias and mean squared error of the moment estimates, separately for each of the parameters, using the expressions

$$\text{Mean} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}, \quad \text{Bias} = \frac{1}{n} \sum_{i=1}^n (\hat{\theta} - \theta), \quad \text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{\theta} - \theta)^2,$$

where  $n=100$  and  $\theta = \lambda, \mu, \phi_1, \phi_2, \eta$  and  $\xi$  are the fitted parameter values of the empirical rainfall data for the month under study and  $\hat{\theta}$  is the corresponding estimate for the simulated data.

The above steps were repeated with  $n_l = 20, 40, 60, 69$  years of data to produce a table of average, bias, root mean squared error ( $\sqrt{\text{MSE}}$ ) for each of the 6 parameters. The results are presented in Table 4 where the fitted parameter values for August are noted as the true values used in the simulation.

The means of the simulated estimates showed that these are generally close to the corresponding true values. The means of  $\hat{\mu}$  and  $\hat{\xi}$  were converging towards their true values as  $n_l$  became large, whereas the means of  $\hat{\phi}_2$  and  $\hat{\eta}$  did not seem to exhibit this property. It is worth pointing out that the parameters  $\phi_2$  and  $\eta$  are positively correlated and when  $\phi_2$  increased  $\eta$  also increased. The reason for this is that much of the rain cells come from the high intensity state 2 and, as we are equating the rainfall moments in our estimation, increased arrival rate  $\phi_2$  of rain cells results in smaller cell lifetimes of  $1/\eta$ , and vice-versa. Means of  $\hat{\lambda}$  and  $\hat{\phi}_1$  do not seem to show much variation.

Table 4: Simulation study results: values of the mean, bias and root mean squared error ( $\sqrt{\text{MSE}}$ ) for the moment estimates of the parameters based on 100 simulated series of length  $n_l$  years for the month of August. True values are the fitted parameter values for the observed rainfall data for August.

Mean						
$n_l$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\eta}$	$\hat{\xi}$
20	0.01964	6.21585	0.02008	6.21159	6.22667	270.2398
40	0.01966	6.20756	0.02046	6.19146	6.20708	268.9749
60	0.01963	6.17081	0.01971	6.16408	6.17867	269.7928
69	0.01963	6.14516	0.01983	6.14164	6.15634	268.2950
True values	0.01970	6.12910	0.02760	6.33910	6.31180	265.9815

Bias						
$n_l$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\eta}$	$\hat{\xi}$
20	-6.4E-05	0.08675	-0.00752	-0.12751	-0.08513	4.25827
40	-7.0E-05	0.07846	-0.00715	-0.14764	-0.10472	3.99343
60	-6.9E-05	0.04171	-0.00789	-0.17502	-0.13313	3.81131
69	-6.6E-05	0.01606	-0.00777	-0.19746	-0.15546	2.31354

$\sqrt{\text{MSE}}$						
$n_l$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\eta}$	$\hat{\xi}$
20	7.80E-05	0.25363	0.00784	0.30055	0.28529	8.31461
40	7.28E-05	0.26760	0.00732	0.32294	0.30633	8.00499
60	6.87E-05	0.26515	0.00803	0.33909	0.32019	8.22053
69	7.28E-05	0.27809	0.00792	0.36235	0.34201	8.37347

The bias of  $\hat{\mu}$  and  $\hat{\xi}$  became smaller when  $n_l$  became larger but both  $\hat{\phi}_2$  and  $\hat{\eta}$  seemed to show a slight negative bias. As we can see from the values,  $\hat{\lambda}$  and  $\hat{\phi}_1$  both show very small negative bias and again do not show much variation with  $n_l$ . The root mean squared errors seemed to stay more or less at the same level for most of the parameters, as  $n_l$  increased, except for  $\hat{\phi}_2$  and  $\hat{\eta}$  which showed a slight increase. One possible reason for that might be that the moment estimates may be slightly biased because of the serial correlation in the data, as noted by [COWPERTWAIT ET AL. \(2007\)](#), especially at sub-hourly aggregation levels which show high autocorrelation. Ideally we would have liked the bias and MSE of the estimators to converge to zero asymptotically for all parameters, as would those of the maximum likelihood estimators which have good large sample statistical properties. However, the means of the simulated estimates were closer to their true values and the relative sizes of the bias and root mean squared errors seemed to be reasonably small for the application.

## 4 Conclusions and future work

We have developed a doubly stochastic point process model, using instantaneous pulses, to study the fine-scale structure of sub-hourly rainfall time series. Second and third-order moment properties of the aggregated rainfall for the proposed DSP model were derived. The model was used to analyse 69 years of 5-minute rainfall data. The empirical properties of the rainfall



accumulations were shown to be in very good agreement with the fitted theoretical values over a range of sub-hourly time scales, including those that were not used in fitting. Although the use of second-order moments in estimation produced very good results, incorporation of third-order moments showed a clear improvement in fitting. Overall, the results of our analysis suggest that the proposed stochastic model is capable of reproducing the fine-scale structure of the rainfall process, and hence could be a useful tool in environmental or ecological impact studies.

The simulated extreme values at daily and 12-hourly aggregations are in very good agreement with their empirical counterparts. However, although the model reproduces the moment properties well, it underestimates the extremes at fine time-scale. The results from the analysis of the proportion of dry periods, using intervals with rainfall depths below appropriate threshold levels, show that the model generally reproduces the observed proportions well. The simulation study shows that the estimation method used is capable of reproducing the estimates closer to the true values, although it may not have the desired large sample properties which maximum likelihood estimators exhibit. Overall, the above analyses indicate that the proposed modelling approach is able to fit data over a range of sub-hourly time scales and reproduce most of the properties well. It has potential application in many areas, as it provides a fast and efficient way of generating synthetic fine-scale rainfall input to hydrological models directly from one stochastic model.

Despite this, there is potential to develop the model further to employ a 3-state doubly stochastic model for cell arrivals and also to explore its capability to handle aggregations at higher levels. The 3-state model, along with further developments to study other hydrological properties of interest, may be a fruitful area for future work on. Although our model based on one doubly stochastic process has performed well, another possibility is to consider superposing two doubly stochastic pulse processes, as in [Cowpertwait et al. \(2007\)](#), to better account for different types of precipitation, such as convective and stratiform.

**Acknowledgments** The authors would like to thank the Editor, Associate Editors and two anonymous referees for their constructive comments and suggestions on earlier versions of the manuscript. Their comments have greatly improved the quality of the manuscript.

## References

- Bardossy, A. and Plate, E. J. (1991). Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence. *Journal of Hydrology*, 122(1):33–47.
- Cowpertwait, P., Isham, V., and Onof, C. (2007). Point process models of rainfall: developments for fine-scale structure. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 463(2086):2569–2587.
- Cowpertwait, P., Xie, G., Isham, V., Onof, C., and Walsh, D. (2011). A fine-scale point process model of rainfall with dependent pulse depths within cells. *Hydrological Sciences Journal*, 56(7):1110–1117.
- Cowpertwait, P. S. (1994). A generalized point process model for rainfall. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 447(1929):23–37.
- Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. CRC Press.

- Daley, D. J. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*, volume 2. Springer Science & Business Media.
- Davison, A. and Ramesh, N. (1996). Some models for discretized series of events. *Journal of the American Statistical Association*, 91(434):601–609.
- Davison, A. and Ramesh, N. (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):191–208.
- Jesus, J. and Chandler, R. E. (2011). Estimating functions and the generalized method of moments. *Interface focus*, page rsfs20110057.
- Leiva, V., Ferreira, M., Gomes, M., and Lillo, C. (2016). Extreme value birnbaum-saunders regression models applied to environmental data. *Stochastic Environmental Research and Risk Assessment*.
- Nelder, J. and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, 7:308–313.
- Onof, C., Chandler, R., Kakou, A., Northrop, P., Wheeler, H., and Isham, V. (2000). Rainfall modelling using Poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 14(6):384–411.
- Onof, C. and Wheeler, H. S. (1994). Improvements to the modelling of british rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology*, 157(1):177–195.
- Onof, C., Yameundjeu, B., Paoli, J., and Ramesh, N. (2002). A Markov modulated Poisson process model for rainfall increments. *Water Science & Technology*, 45(2):91–97.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ramesh, N. (1995). Statistical analysis on Markov-modulated Poisson processes. *Environmetrics*, 6(2):165–179.
- Ramesh, N. (1998). Temporal modelling of short-term rainfall using Cox processes. *Environmetrics*, 9(6):629–643.
- Ramesh, N. and Davison, A. (2002). Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, 256(1):106–119.
- Ramesh, N. and Onof, C. (2014). A class of hidden Markov models for regional average rainfall. *Hydrological Sciences Journal*, 59(9):1704–1717.
- Ramesh, N. and Thayakaran, R. (2012). Stochastic point process models for fine-scale rainfall time series. *In Proceedings of the International Conference on Stochastic Modelling Techniques and Data Analysis*, pages 635–642.
- Ramesh, N., Thayakaran, R., and Onof, C. (2013). Multi-site doubly stochastic Poisson process models for fine-scale rainfall. *Stochastic Environmental Research and Risk Assessment*, 27(6):1383–1396.
- Ramesh, N. I., Onof, C., and Xie, D. (2012). Doubly stochastic Poisson process models for precipitation at fine time-scales. *Advances in Water Resources*, 45:58–64.

- Rodriguez-Iturbe, I., Cox, D., and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 410(1839):269–288.
- Smith, J. A. and Karr, A. F. (1983). A point process model of summer season rainfall occurrences. *Water Resources Research*, 19(1):95–103.
- Thayakaran, R. and Ramesh, N. (2013). Multivariate models for rainfall based on Markov modulated Poisson processes. *Hydrology Research*, 44(4):631–643.
- Verhoest, N., Troch, P. A., and De Troch, F. P. (1997). On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point. *Journal of hydrology*, 202(1):108–120.
- Wheater, H., Chandler, R., Onof, C., Isham, V., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M.-L. (2005). Spatial-temporal rainfall modelling for flood risk estimation. *Stochastic Environmental Research and Risk Assessment*, 19(6):403–416.

## Appendix A

The derivation of the third moment involves considering the product densities of three pulses at times  $x = t$ ,  $y = t + u$  and  $z = t + u + v$  and taking account of the contribution from three different cases. The three cases are where the pulses come from three different cells, two of them from the same cell and one from another, and all three come from the same cell. Following this, we find that the third moment about the origin of the aggregated rainfall, in intervals of length  $h$ , can be written as

$$\begin{aligned}
\mathbb{E} \left[ \left[ Y_i^{(h)} \right]^3 \right] &= \int_{t=0}^h \int_{y=0}^h \int_{z=0}^h \mathbb{E} [X(t)X(y)X(z)dN(t)dN(y)dN(z)] \\
&= 6 \int_{t=0}^h \int_{u=0}^{h-t} \int_{v=0}^{h-t-u} \mathbb{E} [X(t)X(t+u)X(t+u+v)dN(t)dN(t+u)dN(t+u+v)] \\
&+ 6 \int_{t=0}^h \int_{u=0}^{h-t} \mathbb{E} [X(t)(X(t+u))^2dN(t)dN(t+u)] \\
&+ \int_{t=0}^h \mathbb{E} [(X(t))^3dN(t)].
\end{aligned}$$

After a lengthy algebraic calculation, considering various cases noted above, we find that the third moment about the origin of the aggregated process is given by

$$\begin{aligned}
&\mathbb{E} \left[ \left[ Y_i^{(h)} \right]^3 \right] = \\
&= \frac{6\xi^3}{\eta^2} \left[ \mathbb{E}[X_{ij}X_{ik}X_{il}] \frac{m}{\eta} - \mathbb{E}[X_{ij}X_{ik}] \mu_X \left( \frac{2m^2}{\eta^2} + \frac{A}{\eta^2 - (\lambda + \mu)^2} \right) + \mu_X^3 \left( \frac{A_1}{\eta^3} + \frac{A_3}{\eta(\eta^2 - (\lambda + \mu)^2)} \right) \right] \\
&\quad \times \left[ h - \frac{2}{\eta} + \frac{2e^{-\eta h}}{\eta} + he^{-\eta h} \right] \\
&\quad + \frac{6\xi^3}{\eta(\lambda + \mu + \eta)(\eta^2 - (\lambda + \mu)^2)} \left[ (A_4 + A_2) \frac{\mu_X^3}{\eta} - A\mu_X \mathbb{E}[X_{ij}X_{ik}] \right]
\end{aligned}$$

$$\begin{aligned}
& \times \left[ h - \frac{\eta + 2(\lambda + \mu)}{(\lambda + \mu)(\lambda + \mu + \eta)} + \frac{(\lambda + \mu + \eta)e^{-\eta h}}{(\lambda + \mu)\eta} - \frac{\eta e^{-(\lambda + \mu + \eta)h}}{(\lambda + \mu)(\lambda + \mu + \eta)} \right] \\
& + \frac{6\xi^3}{(\lambda + \mu)(\lambda + \mu + \eta)(\eta^2 - (\lambda + \mu)^2)} \left[ A\mu_X E[X_{ij}X_{ik}] - (A_4 + A_2)\frac{\mu_X^3}{\eta} \right] \\
& \times \left[ h - \frac{\eta + 2(\lambda + \mu)}{(\lambda + \mu)(\lambda + \mu + \eta)} + \frac{(\lambda + \mu + \eta)e^{-(\lambda + \mu)h}}{(\lambda + \mu)\eta} - \frac{(\lambda + \mu)e^{-(\lambda + \mu + \eta)h}}{\eta(\lambda + \mu + \eta)} \right] \\
& + \frac{6\mu_X^3 \xi^3 A_4}{\eta(\lambda + \mu)^2(\eta^2 - (\lambda + \mu)^2)} \left[ h - \frac{2}{\lambda + \mu} + \frac{2e^{-(\lambda + \mu)h}}{\lambda + \mu} + he^{-(\lambda + \mu)h} \right] \\
& + \frac{6\xi^3}{\eta(\lambda + \mu)(\eta^2 - (\lambda + \mu)^2)} \left[ A\mu_X E[X_{ij}X_{ik}] - (A_3 + A_4)\frac{\mu_X^3}{\eta} \right] \\
& \times \left[ h - \frac{(\lambda + \mu + \eta)}{\eta(\lambda + \mu)} + \frac{(\lambda + \mu)e^{-\eta h}}{\eta(\lambda + \mu - \eta)} - \frac{\eta e^{-(\lambda + \mu)h}}{(\lambda + \mu)(\lambda + \mu - \eta)} \right] \\
& + \frac{12\xi^3}{\eta} \left[ E[X_{ij}X_{ik}]\mu_X \frac{m^2}{\eta^2} - \frac{A_1\mu_X^3}{\eta^3} - \frac{(A_2 + A_3)\mu_X^3}{2\eta(\eta^2 - (\lambda + \mu)^2)} \right] \times \left[ \frac{h^2}{2} - \frac{h}{\eta} + \frac{1}{\eta^2} - \frac{e^{-\eta h}}{\eta^2} \right] \\
& + \frac{6\mu_X^3 \xi^3 (A_2 + A_3)}{\eta(\lambda + \mu)(\eta^2 - (\lambda + \mu)^2)} \times \left[ \frac{h^2}{2} - \frac{h}{\lambda + \mu} + \frac{1}{(\lambda + \mu)^2} - \frac{e^{-(\lambda + \mu)h}}{(\lambda + \mu)^2} \right] \\
& + \mu_X^3 \frac{A_1}{\eta^3} \xi^3 h^3 + 6E[X_{ij}^2 X_{ik}]\xi^2 \frac{m}{\eta^2} \times \left[ h - \frac{1}{\eta} + \frac{e^{-\eta h}}{\eta} \right] \\
& + 6E[X^2]\mu_X \xi^2 \left[ \left( \frac{A}{(\lambda + \mu)(\eta^2 - (\lambda + \mu)^2)} \right) \left( h - \frac{1}{\lambda + \mu} + \frac{e^{-(\lambda + \mu)h}}{\lambda + \mu} \right) \right] \\
& - 6E[X^2]\mu_X \xi^2 \left[ \left( \frac{m^2}{\eta^3} + \frac{A}{\eta(\eta^2 - (\lambda + \mu)^2)} \right) \left( h - \frac{1}{\eta} + \frac{e^{-\eta h}}{\eta} \right) \right] + 3E[X^2]\mu_X \frac{m^2 \xi^2}{\eta^2} h^2 + E[X^3]h \frac{m\xi}{\eta}.
\end{aligned}$$