Multivariate models for rainfall based on Markov modulated Poisson processes

R. Thayakaran and N. I. Ramesh

ABSTRACT

Point process models for rainfall are constructed generally based on Poisson cluster processes. Most commonly used point process models in the literature were constructed either based on Bartlett–Lewis or Neyman–Scott cluster processes. In this paper, we utilize a class of Cox process models, termed the Markov modulated Poisson process (MMPP), to model rainfall intensity. We use this class of models to analyse rainfall data observed in the form of tip time series from rain gauge tipping buckets in a network of gauges in Somerset, southwest England, recorded by the Hydrological Radar Experiment (HYREX). Univariate and multivariate models are employed to analyse the data recorded at single and multiple sites in the catchment area. As the structure of this proposed class of MMPP models allows us to construct the likelihood function of the observed tip time series, we utilize the maximum likelihood methods in our analysis to make inferences about the rainfall intensity at sub-hourly time scales. The multivariate models are used to analyse rainfall time series jointly at four stations in the region. Properties of the cumulative rainfall in discrete time intervals are studied, and the results of fitting three-state models are presented.

Key words | accumulated rainfall, likelihood function, Markov modulated Poisson process, point process, rainfall intensity

R. Thayakaran (corresponding author)
N. I. Ramesh
School of Computing and Mathematical Sciences,
University of Greenwich,
Old Royal Naval College,
Park Row,
Greenwich,
London SE10 9LS,
UK
E-mail: *R.Thayakaran@greenwich.ac.uk*

INTRODUCTION

The Cox process or the doubly stochastic Poisson process was introduced by Cox (1955), and since then, several classes of this process have been developed by many authors in different applications. This process generalizes the Poisson process by letting the mean rate of occurrence of points vary according to a stochastic process. One useful class of this process arises when the arrival rate is controlled by a finite-state Markov chain, and this special class is called a Markov modulated Poisson process (MMPP) (Meier-Hellstern & Fischer 1993; Ramesh 1995; Ryden 1996). This class of process has been utilized in several areas including telecommunication, environmental and animal population modelling. Applications of this process in telecommunications were studied, for example, by Heffes & Lucantoni (1986) and Neuts (1989). Applications of MMPPs in environmental modelling were considered by Smith & Karr (1983), Davison & Ramesh (1993, 1996), Ramesh (1998) and Onof *et al.* (2002), while some applications in animal population studies can be found in Skaug (2006).

In this paper, we use a class of MMPPs to study the pattern of rainfall bucket tipping times observed in a network of stations in a catchment area. Stochastic point rainfall models based on Poisson cluster processes were first developed by Rodriguez-Iturbe *et al.* (1987) using a rectangular pulse attached to each point of the rainfall cell arrival process. Since then, many authors have studied Bartlett–Lewis and Neyman–Scott rectangular pulse models extensively to analyse rainfall time series; see, for example, Onof & Wheater (1993, 1994), Onof *et al.* (2000), Wheater *et al.* (2005) and Cowpertwait (1994). Chandler (1997) developed a spectral method to estimate parameters in rainfall models, whereas Wheater *et al.* (2000) studied spatial–temporal structure of rainfall process by using the Hydrological Radar Experiment (HYREX) rainfall data. Recently,

doi: 10.2166/nh.2013.180

Kigobe *et al.* (2011) applied a generalized linear modelling framework to develop multi-site stochastic daily rainfall models to reproduce spatial and temporal patterns of precipitation. Following an alternative modelling approach based on a doubly stochastic Poisson process, Ramesh *et al.* (2012) described a class of bivariate MMPP models to study precipitation at fine time scales.

To develop this approach further, we first consider a univariate MMPP to study the sequence of tip times to analyse the rainfall intensity at a single location and then move on to describe the multivariate generalization. The multivariate MMPPs are used to model the pattern of tip times across four stations simultaneously. As shown in previous studies (Onof *et al.* (2002) for univariate and Ramesh *et al.* (2012) for bivariate processes), three-state models generally provide a reasonably good description of the rainfall pattern and we shall be exploring the use of the three-state process further for both univariate and multivariate models. The states of the underlying Markov chain may be interpreted as background environmental conditions that give rise to heavy rain, moderate rain and little or no rain.

Data and study area

The data sets used in our analysis, compiled under the HYREX project, were obtained from the British Atmospheric Data Centre (BADC). The HYREX project (Wood *et al.* 2000) was specially designed to measure the rainfall using a dense rain gauge network and related weather radars in the Brue experimental catchment in Somerset, southwest England. The catchment area contains 49 tipping bucket rain gauges across an area of approximately 135 km^2 and the tip time data are available between 1994 and 1999. In our multivariate model, we use rainfall data from four stations including Towers field (TOWE), New house farm field (NEWH), Dodds corner field (DODD) and Dropping lane farm (DROP), with their locations shown in Figure 1. All four stations selected are in a 2×2 km^2 grid in the region.

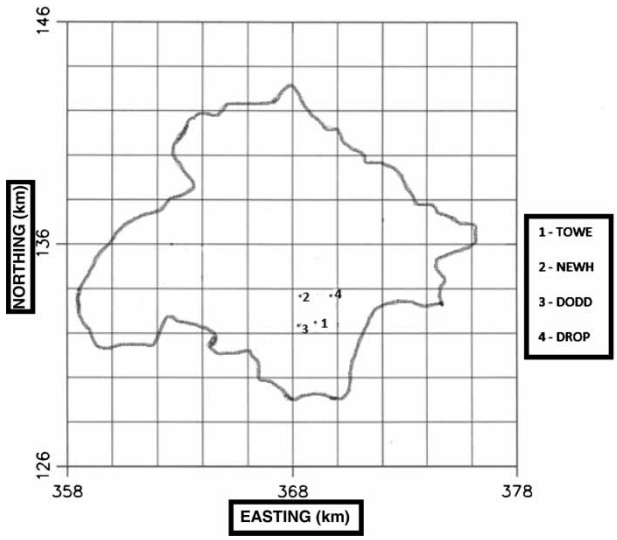

Figure 1 | Map of Brue catchment area in Somerset, southwest England, indicating location of rain gauge stations used in our study.

MMPP MODELS FOR TIP TIMES

We describe univariate and multivariate point process models for the bucket tip time series based on Markov modulated Poisson process. A common problem with most point process models is that their likelihood function is either not available in a closed form or difficult to calculate. The advantage of using the MMPP models is that their likelihood function can be written down in a closed form for both univariate and multivariate cases. We shall describe this in the following subsections.

Univariate model

Let $N(t)$ be the process of tip times, which is taken as an MMPP with a corresponding underlying process $X(t)$. The process $X(t)$ is assumed to be an irreducible continuous time Markov chain on m states with a generator Q, see Ramesh (1995) for further details. In general, for a model with m states for the Markov chain, there will be $m(m-1)$ transition rate parameters and m Poisson arrival rate parameters, taking the total number of parameters of the model to m^2. This can be large for higher values of m, but quite often a model with $m = 3$ or $m = 4$ would be sufficient for the application of this type (Ramesh *et al.* 2012). Stochastic models based on three-state Markov process have also been utilized in the analysis of daily rainfall, using a hidden Markov model, by many authors (e.g. Zucchini & Guttorp 1991; Thompson *et al.* 2007). We shall be using $m = 3$ in our models and for this case, the parameters are given by the

arrival rate matrix L of the point process and the infinitesimal generator matrix Q of the underlying Markov chain:

$$L_{3\times3} = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 0 \\ 0 & 0 & \phi_3 \end{bmatrix} \quad \text{and}$$

$$Q_{3\times3} = \begin{bmatrix} -(q_{12}+q_{13}) & q_{12} & q_{13} \\ q_{21} & -(q_{21}+q_{23}) & q_{23} \\ q_{31} & q_{32} & -(q_{31}+q_{32}) \end{bmatrix} \quad (1)$$

The parameters $q_{12}, q_{13}, q_{21}, q_{23}, q_{31}$ and q_{32} in the matrix Q describes the transition rates of the Markov chain from one state to another and the elements ϕ_1, ϕ_2 and ϕ_3 of L are the Poisson arrival rates in the three states. The stationary distribution of the Markov chain can be obtained in the usual way by solving $\pi Q = 0_{1\times3}$, where $\pi_{1\times3} = (\pi_1, \pi_2, \pi_3)$. The solution can be written as:

$$\pi_1 = (1 - \pi_2 - \pi_3) \quad (2)$$

$$\pi_2 = \frac{(q_{12}+q_{13})(q_{32}-q_{12}) + q_{12}(q_{31}+q_{12}+q_{13})}{(q_{21}+q_{12}+q_{13})(q_{32}-q_{12}) + (q_{31}+q_{12}+q_{13})(q_{21}+q_{23}+q_{12})} \quad (3)$$

$$\pi_3 = \frac{(q_{21}+q_{23}+q_{12})\pi_2 - q_{12}}{(q_{32}-q_{12})} \quad (4)$$

Likelihood function

To calculate the likelihood function of the point process $N(t)$ whose underlying process $X(t)$ has m states, we define the following conditional probabilities, as shown in Smith (1984) and Ramesh (1995, 1998):

$$\psi_{ij}(t) = P(X(t) = j, N(t) = 0 | X(0) = i, N(0) = 0), \quad i, j = 1, 2, \cdots, m \quad (5)$$

Let $\Psi(t)$ be the matrix function with entries $\psi_{ij}(t)$, then by writing the Chapman–Kolmogorov forward differential equations (Ross 1983) for the process, we can show that:

$$\Psi(t) = e^{(Q-L)t} = \sum_{n=0}^{\infty} \frac{t^n (Q-L)^n}{n!}, \quad \text{where} \quad (Q-L)^0 = I \quad (6)$$

Suppose that the process $N(t)$ is observed in the interval $[0, t_n]$, and points occurred at times $t_1 < t_2 < \cdots < t_n$. The likelihood function of Q and L, given the data, is then given by (Ramesh 1998; Onof *et al.* 2002):

$$\mathrm{Li}(Q, L | t_1, t_2, \cdots, t_n) = \pi_{1\times m} \left[\prod_{i=1}^{n} [\Psi(t_i - t_{i-1})L] \right] 1_{m\times1} \quad (7)$$

where $1_{m\times1}$ is a column vector of ones, $t_0 = 0$, $L_{m\times m}$ and $Q_{m\times m}$ are model parameter matrices and $\pi_{1\times m} = (\pi_1, \pi_2, \cdots, \pi_m)$ is the equilibrium distribution of the underlying Markov chain $X(t)$. This likelihood function can be maximized numerically to obtain maximum likelihood estimates (MLEs) of the parameters.

Multivariate MMPP models

Multivariate MMPP models can be used to analyse the rainfall tip time series at multiple sites. These models are developed, following on from the idea used in univariate MMPP, to jointly model the rainfall patterns at more than one station by assuming that the occurrence process at the stations are all governed by the same underlying Markov chain. In our analysis, we consider the joint modelling of rainfall data from four adjacent stations, so the common Markov chain assumption is sensible. In the three-state multivariate models, the Markov chain has the same six parameters as in univariate MMPP, but the number of arrival rate parameters depends on the number of stations used in the analysis. Within this setup, two types of models are considered to accommodate additional features: the first one is a full model with 18 parameters and the second one is a reduced or linked model with 12 parameters.

Multivariate model M1

In this model, we shall assume that all the four stations are governed by the same underlying Markov chain $X(t)$, but they have their own arrival rate matrices. These matrices are denoted by L_1 for TOWE station, L_2 for NEWH station,

L_3 for DODD station and L_4 for DROP station:

$$L_1 = \begin{bmatrix} \phi_{11} & 0 & 0 \\ 0 & \phi_{12} & 0 \\ 0 & 0 & \phi_{13} \end{bmatrix} \quad L_2 = \begin{bmatrix} \phi_{21} & 0 & 0 \\ 0 & \phi_{22} & 0 \\ 0 & 0 & \phi_{23} \end{bmatrix}$$
$$L_3 = \begin{bmatrix} \phi_{31} & 0 & 0 \\ 0 & \phi_{32} & 0 \\ 0 & 0 & \phi_{33} \end{bmatrix} \quad L_4 = \begin{bmatrix} \phi_{41} & 0 & 0 \\ 0 & \phi_{42} & 0 \\ 0 & 0 & \phi_{43} \end{bmatrix} \tag{8}$$

Now, by following the same argument as before, we can write down the likelihood function of this multivariate MMPP model as (see Ramesh *et al.* 2012):

$$\mathrm{Li}(Q, L_1, L_2, L_3, L_4 | t_1, t_2, \cdots, t_n) = \pi_{1\times 3}\left[\prod_{i=1}^{n}\left[e^{(Q-L)(t_i - t_{i-1})}L(i)\right]\right]1_{3\times 1} \tag{9}$$

where $L = L_1 + L_2 + L_3 + L_4$ and $L(i)$ is the arrival rate matrix corresponding to the type of the ith event in the pooled process of events of all four types, i.e.:

$$L(i) = \sum_{j=1}^{4} \delta_{jk} L_j \quad \text{and} \quad \delta_{jk} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \tag{10}$$

where k refers to the type of the ith event in the pooled process and its possible values are $k = 1$ for TOWE, $k = 2$ for NEWH, $k = 3$ for DODD and $k = 4$ for DROP.

The multivariate model with an m states Markov chain has $m(m-1)$ transition rate parameters, as before, and sm arrival rate parameters (s is the number of stations in the multivariate model), giving a total of $m^2 + (s-1)m$ parameters.

Linked multivariate model M2

To incorporate additional dependence between the component processes in the multivariate model, we now introduce a new linked MMPP where the arrival rate matrices for different stations are linked by a linear relationship. In this setting, we take the Poisson arrival rate matrices of the four stations as:

$$L_1 = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 0 \\ 0 & 0 & \phi_3 \end{bmatrix} \quad L_2 = \alpha_2 L_1 \quad L_3 = \alpha_3 L_1 \quad L_4 = \alpha_4 L_1 \tag{11}$$

In this special model, we have only six arrival rate parameters and six transition rate parameters. In a general case, for an m state Markov chain model, there will be $m(m-1)$ transition rate parameters and $m + (s-1)$ arrival rate parameters, giving a total of $m^2 + s - 1$ parameters for this linked model.

DATA ANALYSIS

Univariate MMPP results

We shall begin this section by fitting the univariate models separately for the four stations TOWE, NEWH, DODD and DROP. The data on the rain gauge bucket (0.2 mm) tip time series cover a period of about 6 years from 1994 to 1999. The tip times are viewed as a point process evolving in time, and we analyse this using our MMPP models. Their likelihood function, given in Equation (7), is maximized numerically to obtain MLEs of the parameters. The models are fitted separately for each month. The MLEs of the parameters (not shown due to lack of space) show that state 1 is the dry state with little or no arrival of rain, state 2 is a moderate-intensity state, and state 3 is the high-intensity state. When the process is in the high-intensity state, the arrival rate of rain varies from about seven tips (per hour) to 96 tips (per hour). The higher rates are mainly attributable to the summer months as the rainfall is mostly of the thunderstorm type, with high intensity and short duration. This pattern is in line with results reported in previous studies (Onof *et al.* 2002; Ramesh *et al.* 2012).

Using this model, we studied various properties of rainfall accumulations at discrete time intervals of length $h = 5$, 10, 20, 30 and 60 min for all 12 months. The empirical values of these were compared with their corresponding simulated values from the fitted model. We simulated 100 copies of the data from the fitted model for each month and calculated their properties to make comparison with their empirical counterparts. Figure 2 shows the empirical plots of mean rainfall intensity, standard deviation (SD), autocorrelation at lag 1 (AC1) and proportion of dry period (PDry) at different aggregation levels $h = 5$, 20 and 60 min for NEWH station. Also displayed in the plots are simulation bands using the minimum and maximum values of the 100

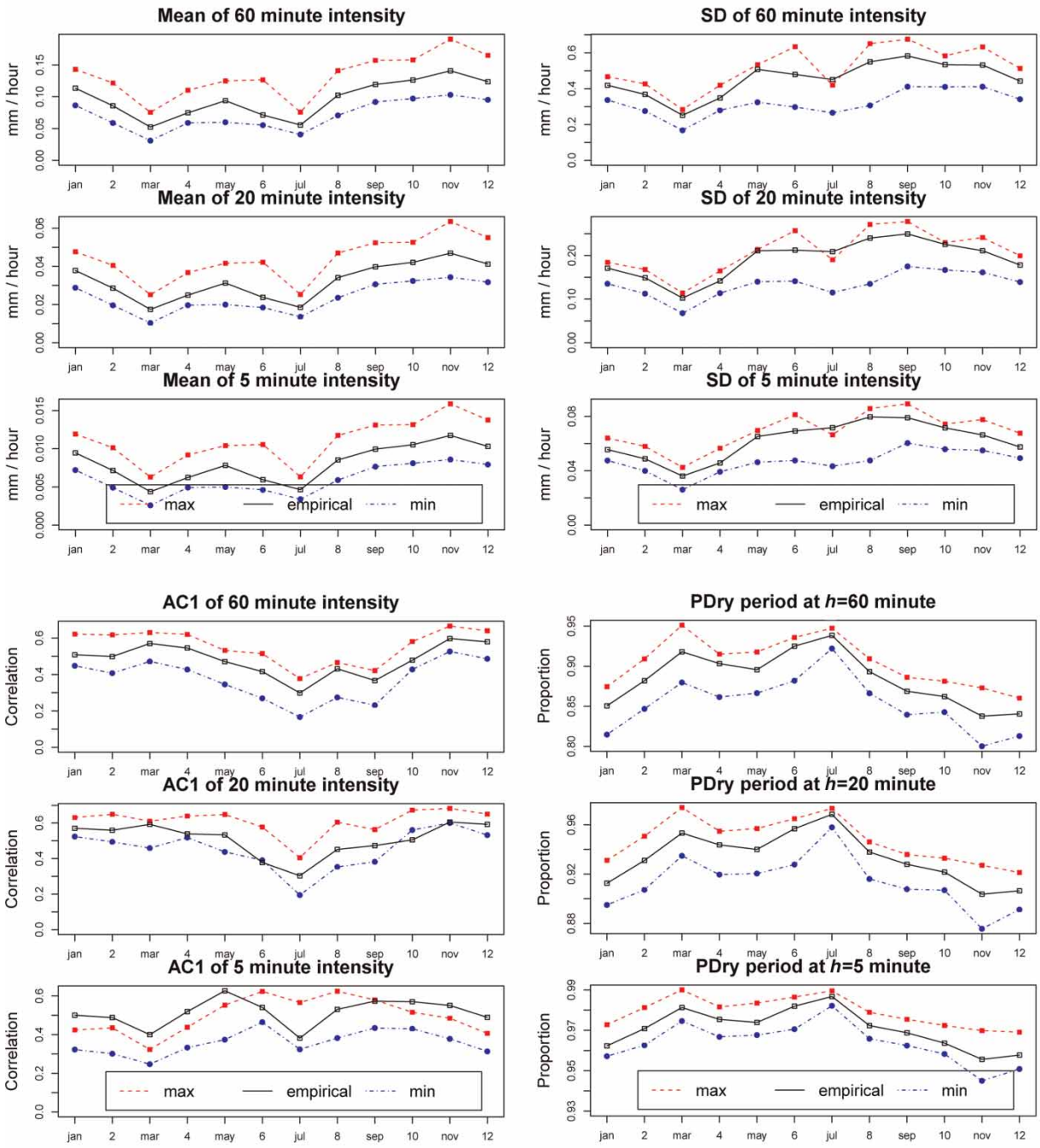

Figure 2 | Mean, standard deviation (SD), autocorrelation (AC1) of rainfall intensity and proportion of dry period (PDry) plots, with simulation bands from the fitted univariate model, at different levels of aggregation for NEWH station.

simulations from the fitted univariate model. The dark solid lines are empirical data and the dashed lines are for the simulation bands in all plots. All of these plots provide very satisfactory results, with AC1 giving an improved fit when moving from smaller sub-hourly aggregations ($h = 5$, 20) to the hourly level. The results obtained by fitting the model to the other three stations are very similar, with only a slight difference in the arrival rate parameters.

Multivariate MMPP results

We started off with the multivariate model M1 to perform a joint analysis of the pattern of tip times at the four stations together. Our multivariate models assume a common underlying Markov process for the occurrences of points at all stations. This assumption is quite reasonable when modelling rainfall at multiple recording stations within a catchment area, as the underlying Markov process controlling the occurrence process is expected to represent the environmental weather conditions of the area. In addition, the four stations selected in our analysis are all situated within a $2 \times 2\,\text{km}^2$ grid in the region and, therefore, their rainfall features would be expected to depend on the same weather patterns. The estimates of the Markov chain parameters, when the univariate models are fitted separately for the four stations, suggest that our assumption is a sensible one for this application. It is also worth noting that the common Markov chain assumption is usually utilized in the analysis of multi-site rainfall occurrence processes using hidden Markov models (Zucchini & Guttorp 1991).

The tip time series at the four sites were labelled as type 1, type 2, type 3 and type 4, and were then pooled together to form a multivariate pooled process of events. The multivariate models are applied to this pooled process representing the tip times at all four stations. The likelihood function of this multivariate process, given in Equation (9), is maximized numerically to obtain parameter estimates.

The MLEs of this first multivariate model (M1) showed a linear association between the arrival rate parameters. Although the results reproduced the statistical properties of cumulative rainfall reasonably well at different aggregation levels, they did not sufficiently capture the correlation between the component processes. As the four stations are all located in a $2 \times 2\,\text{km}^2$ grid of the catchment area, a good deal of association between the rainfall pattern would be expected. To accommodate this and to incorporate sufficient dependency between the four component processes, we introduced the second multivariate model (M2), where the arrival rate parameters are linked by a linear relationship, as given in the previous section.

This linked multivariate model with linearly dependent arrival rates is fitted to the pooled data from the four stations separately for all 12 months. The results of fitting this model M2 are presented in Table 1, which shows the MLEs of the parameters with standard errors of estimates in parentheses. The MLEs and their standard errors suggest that the transition rate parameter $q_{31} = 0$ for some months. A reduced model with $q_{31} = 0$ did not have any noticeable effects on the overall results, and therefore we keep this parameter in our model to maintain a consistent model across the months. Here again, the results show a broadly similar pattern to that of the univariate models in that state 1 is a dry state with little or no rain, state 2 has moderate or light rain and state 3 has high-intensity rain. The estimates of ϕ_3 are highest for the summer months and smallest for the winter months. The final column of the table shows the number of tips observed for each month over the period of study for the pooled process. It shows that the month November has the highest amount of rain and July has the lowest rainfall. This is natural as July is in the middle of the summer period in England and November falls in the wet period.

Table 2 shows the mean sojourn times, the amount of time the Markov chain spends on average in a state ($1/q_i$ where $q_i = \sum_{j \neq i} q_{ij}$), and the equilibrium probabilities of the three states $i = 1, 2, 3$ for each month. It shows that the underlying Markov chain only spends, on average, about 1–3% of the time in the high-intensity rain state and about 87–97% of the time in the dry state, while the remaining time is spent in the moderate rain state. The average sojourn times in state 1 varies from 9 to 17 hours, whereas in state 3 it varies between 9 and 36 min over the months.

Here again, we studied properties of rainfall accumulations at $h = 5$, 10, 20, 30 and 60 min for all 12 months using the linked multivariate model with linearly dependent arrival rates. The empirical values of the accumulated rainfall in discrete intervals at the four stations were compared with the corresponding simulated ones from the fitted multivariate model M2. We simulated 100 copies of the data from the fitted multivariate model for each month and calculated their statistical properties to make comparison with their empirical counterparts for the four stations. The upper panels of Figure 3 show empirical plots of mean rainfall intensity at different levels of aggregation for the stations TOWE and NEWH, together with the simulation band constructed as the minimum and maximum values from the 100

Table 1 | Maximum likelihood estimates and their standard errors for the linked multivariate model M2

Month	q_{12}	q_{13}	q_{21}	q_{23}	q_{31}	q_{32}	ϕ_1	ϕ_2	ϕ_3	α_1	α_2	α_3	No. of points
JAN	0.081	0.013	0.882	0.427	0.145	2.105	0.017	2.018	13.238	1.156	1.077	0.941	9,146
	(0.006)	(0.003)	(0.069)	(0.047)	(0.106)	(0.217)	(0.001)	(0.079)	(0.365)	(0.034)	(0.032)	(0.029)	
FEB	0.069	0.008	0.772	0.301	0.479	1.744	0.026	2.931	16.457	0.862	0.774	0.811	6,956
	(0.006)	(0.002)	(0.067)	(0.043)	(0.154)	(0.252)	(0.002)	(0.111)	(0.554)	(0.028)	(0.026)	(0.027)	
MAR	0.055	0.012	0.496	0.351	0.267	1.323	0.009	1.052	8.358	0.739	0.980	0.932	5,784
	(0.006)	(0.003)	(0.056)	(0.042)	(0.087)	(0.149)	(0.001)	(0.076)	(0.293)	(0.028)	(0.035)	(0.034)	
APR	0.053	0.017	0.705	0.325	0.581	2.031	0.016	2.425	15.371	0.836	0.797	0.946	6,912
	(0.005)	(0.003)	(0.066)	(0.048)	(0.149)	(0.239)	(0.001)	(0.099)	(0.490)	(0.028)	(0.027)	(0.031)	
MAY	0.063	0.003	1.056	0.403	0	3.971	0.024	3.708	34.201	1.023	0.911	0.910	7,868
	(0.005)	(0.001)	(0.086)	(0.053)	–	(0.464)	(0.002)	(0.169)	(1.329)	(0.032)	(0.029)	(0.029)	
JUN	0.063	0.004	1.014	0.409	0	4.573	0.016	4.0264	40.705	0.687	0.965	0.769	7,672
	(0.005)	(0.002)	(0.076)	(0.052)	–	(0.480)	(0.001)	(0.143)	(1.374)	(0.023)	(0.029)	(0.025)	
JUL	0.062	0.005	2.201	0.530	0.174	6.716	0.012	4.6134	40.896	1.044	0.969	1.024	4,800
	(0.005)	(0.002)	(0.174)	(0.087)	(0.426)	(0.949)	(0.001)	(0.233)	(1.978)	(0.042)	(0.040)	(0.042)	
AUG	0.058	0.003	1.073	0.399	0	4.342	0.031	4.711	44.782	0.993	1.021	0.728	8,612
	(0.005)	(0.001)	(0.082)	(0.051)	–	(0.489)	(0.002)	(0.152)	(1.462)	(0.029)	(0.029)	(0.023)	
SEP	0.106	0.003	1.644	0.429	0	5.279	0.027	5.093	50.316	0.985	0.893	0.974	10,083
	(0.006)	(0.002)	(0.100)	(0.054)	–	(0.559)	(0.002)	(0.156)	(1.509)	(0.027)	(0.025)	(0.027)	
OCT	0.097	0.004	1.534	0.7520	0	3.438	0.029	4.006	24.333	1.011	0.871	0.809	10,294
	(0.007)	(0.002)	(0.099)	(0.074)	–	(0.298)	(0.002)	(0.171)	(0.729)	(0.026)	(0.024)	(0.023)	
NOV	0.092	0.010	0.998	0.431	0.090	2.496	0.027	3.302	20.414	1.012	0.937	0.954	11,734
	(0.006)	(0.002)	(0.067)	(0.047)	(0.092)	(0.243)	(0.002)	(0.102)	(0.509)	(0.026)	(0.025)	(0.025)	
DEC	0.084	0.008	0.978	0.419	0	1.678	0.019	2.262	13.199	1.134	1.068	0.953	10,106
	(0.006)	(0.002)	(0.066)	(0.046)	–	(0.148)	(0.001)	(0.083)	(0.327)	(0.032)	(0.030)	(0.028)	

Table 2 | Mean sojourn times (in hours) and equilibrium probabilities of the underlying Markov chain in the three states for the multivariate model M2 (TOWE, NEWH, DODD and DROP). Here, S_i is the mean sojourn time ($1/q_i$) and $\hat{\pi}_i$ is the equilibrium probability for state i of the Markov chain for $i = 1, 2, 3$

Multivariate model M2

Month	S_1	S_2	S_3	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$
JAN	10.638	0.764	0.444	0.887	0.091	0.022
FEB	12.953	0.932	0.449	0.904	0.081	0.014
MAR	14.992	1.179	0.629	0.869	0.101	0.029
APR	14.286	0.926	0.383	0.907	0.076	0.017
MAY	15.290	0.686	0.252	0.936	0.058	0.006
JUN	14.881	0.702	0.219	0.932	0.062	0.006
JUL	14.771	0.366	0.145	0.967	0.029	0.003
AUG	16.502	0.679	0.230	0.941	0.053	0.006
SEP	9.225	0.482	0.189	0.933	0.062	0.006
OCT	9.823	0.437	0.291	0.924	0.061	0.015
NOV	9.746	0.699	0.387	0.892	0.089	0.018
DEC	10.822	0.716	0.596	0.890	0.084	0.026

simulations. All values of the observed mean rainfall intensity are well within the simulation band at all levels of aggregations. There appears to be a notable dip in rainfall intensity for the summer month of July, especially for TOWE and DODD, and the intensity is highest during November for all four stations. Plots for the proportion of dry period at $h = 30$, 20 and 10 min aggregation levels are displayed in the lower panels of Figure 3 for the stations DODD and DROP. Again, this is reproduced well at all levels by the multivariate model, which shows that the month of July has the highest proportion of dry periods. A clear peak for July is visible in most plots, which mirrors the dip in the mean intensity curve for July in the upper panels.

The upper panels of Figure 4 show the plots for the SD and AC1 of the rainfall intensity at different aggregation levels h for the stations TOWE and NEWH, respectively. The SD appears to be underestimated by the multivariate model, especially for the summer months, although it improves at larger values of h. AC1 also shows underestimation for the first 6 months of the year, but improves again as h increases. The lower left panel shows that autocorrelation at lag 2 (AC2) has been reproduced reasonably well at lower aggregation levels for station DODD. The mean duration of all dry periods (MDDry) for the station DROP is reproduced well in the lower right panel at aggregation levels $h = 5$, 10 and 20 min, and shows a noticeable peak in July. The month of July has the highest mean duration of dry period for all stations, which agrees with the findings from Figure 3.

The cross-correlations at lag 0 (CC0) of the rainfall at $h = 60$ min aggregation level, which measure the linear association between the rainfall amount at two sites within the same hour, are reproduced well by the linked multivariate model M2. The cross-correlations at lag 1 (CC1) appear to vary a lot from 1 month to another in the upper panels of Figure 5, but fall mostly within the simulation band. The cross-correlations at lag 2 (CC2) are also captured reasonably well by this multivariate model in the lower panels of Figure 5, although they too exhibit greater variability across the months.

Extreme values

As the knowledge of the properties of extreme rainfall is important in many applications, in this subsection, we examine the capability of our model to generate rainfall extremes. We have only 6 years of historical data for our analysis, from the HYREX project, which is in general not enough to embark on a full-scale extreme value performance of our fitted model. However, we used the rainfall data from one of the four stations (DODD) to study the extreme rainfall generated by the fitted univariate model for the station. The yearly extreme rainfall values generated by the model are compared with those of the observed rainfall extremes over this 6 year period for this station. From the fitted model, 100 copies of the 6 year long tip time series were first simulated. For each of these simulated series, the annual maxima over $h = 1$, 6, 12, 24 hour intervals were calculated. They were then ordered to plot against the reduced Gumbel variate along with their empirical counterparts. The ordered annual maxima of the historical extremes, at different aggregation levels, are plotted against their ranks and the corresponding Gumbel reduced variates in Figure 6. The simulated extremes are displayed as interval plots showing the maximum, minimum and the mean from 100 replications for each aggregation. The plots suggest that, in general, the extremes are underestimated by the fitted model at small time scale ($h = 1$ and 6 hours); however, at higher levels of aggregation ($h = 12$ and 24 hours), the

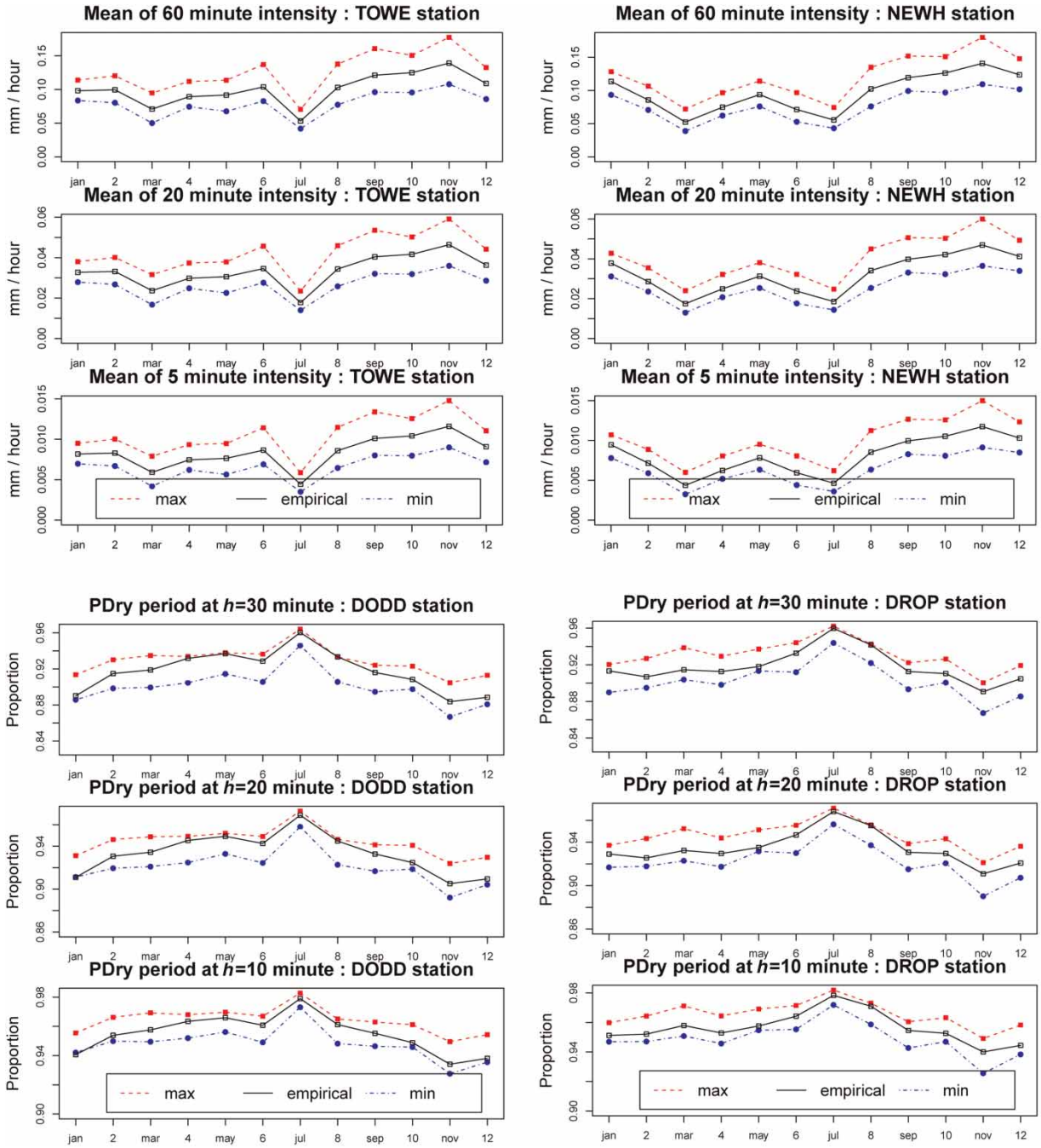

Figure 3 | Mean rainfall intensity and proportion of dry period (PDry) plots, with simulation bands from multivariate model M2, at different time scales. The top panels show the mean rainfall intensity at time scales $h = 60$, 20, 5 min for TOWE and NEWH stations, whereas the bottom panels show the plots of PDry at $h = 30$, 20, 10 min aggregations for stations DODD and DROP, respectively.

simulated extremes are in reasonably good agreement with empirical extremes, especially at higher values of the reduced variate. The model does not seem to reproduce the extremes well at fine time scales (sub-hourly). The underestimation of extremes may have resulted from the small number of historical data available, in the form of tip times in HYREX data, but we intend to investigate this further with longer series of rainfall in future work.

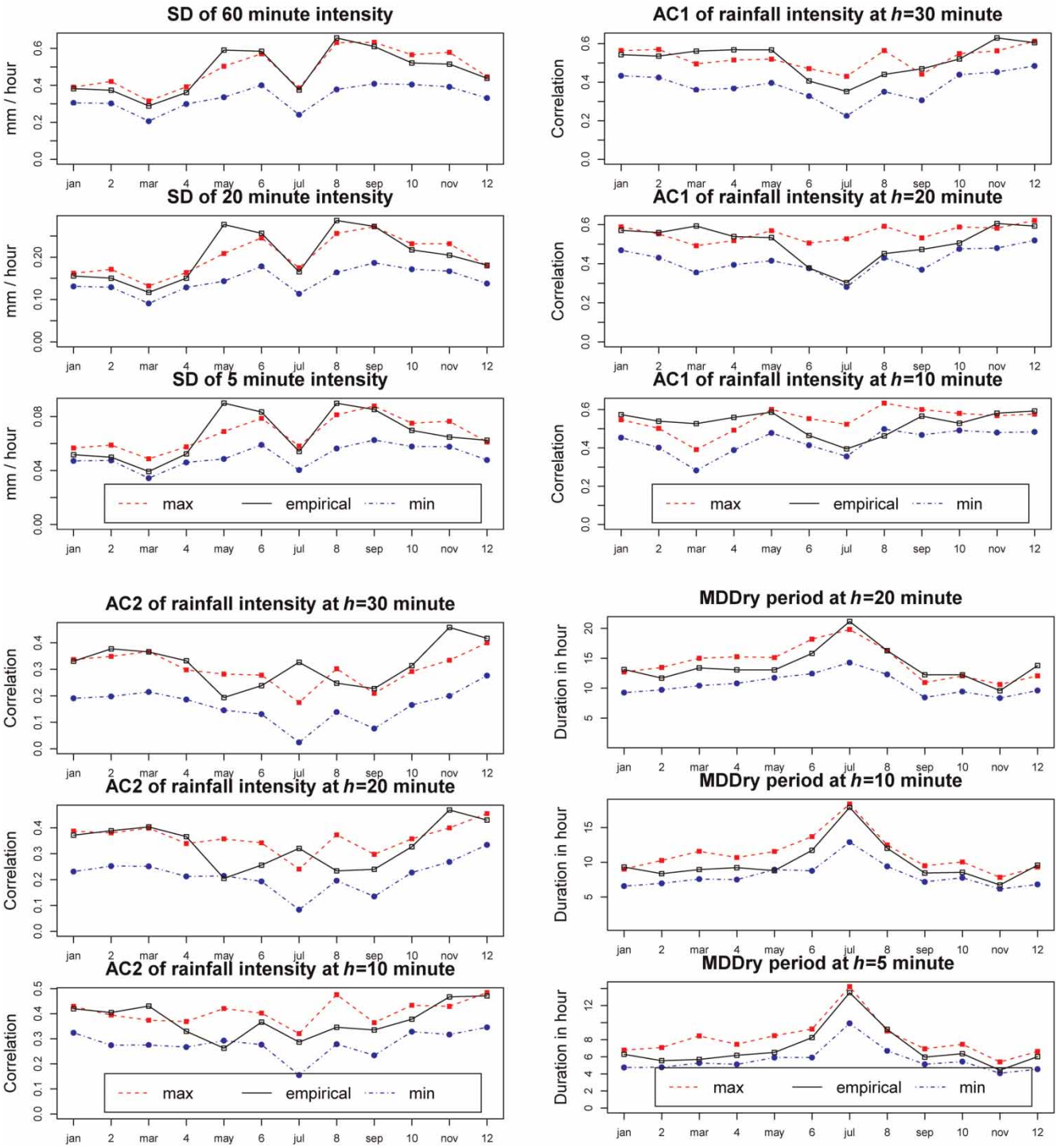

Figure 4 | Empirical quantities and simulation bands from the fitted multivariate model M2 using four stations. The top left panel shows SD plots for TOWE station and the top right panel displays AC1 plots for NEWH station. The bottom left panel shows AC2 plots for DODD station and the bottom right panel gives mean duration Dry (MDDry) plots for DROP station.

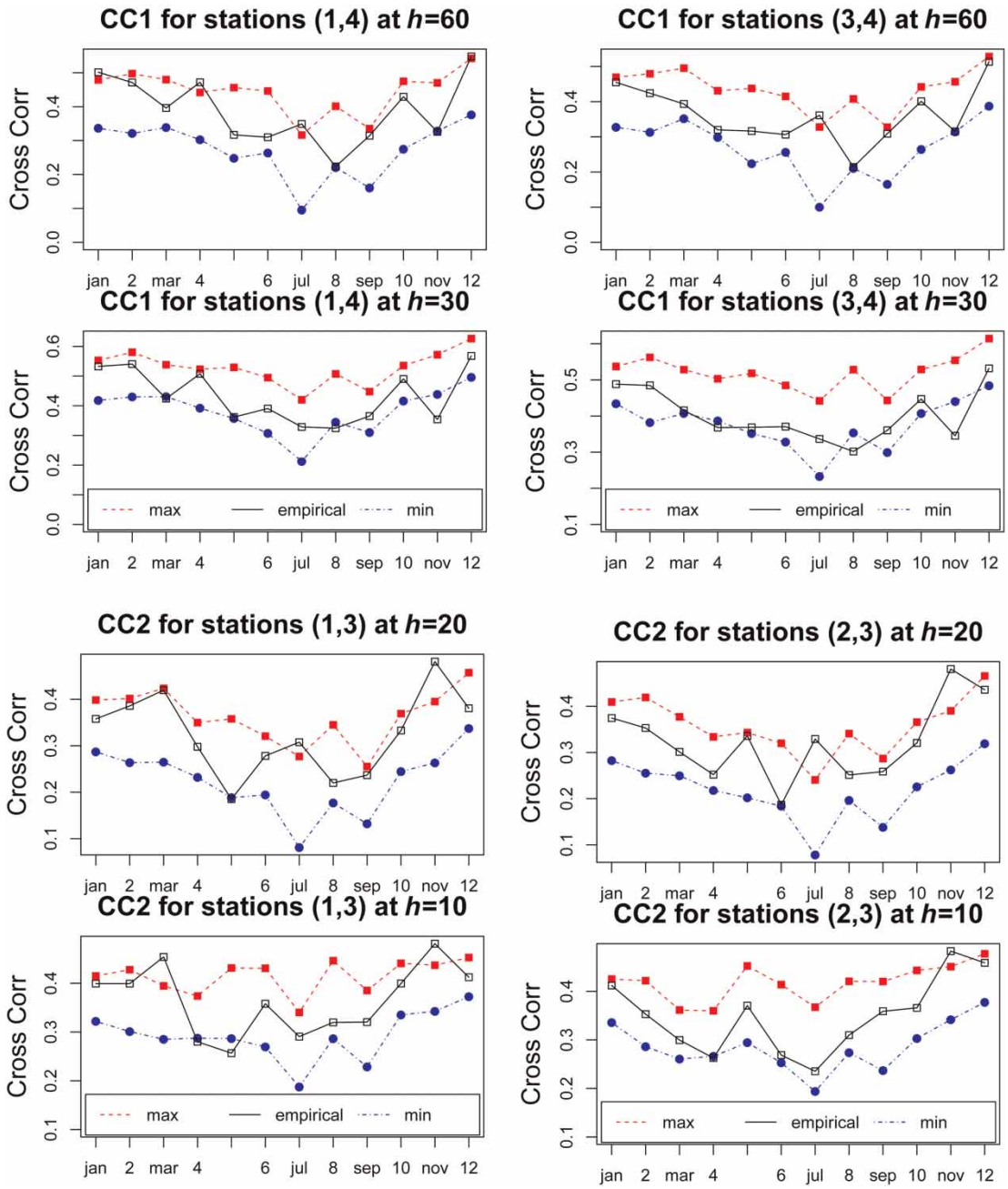

Figure 5 | Cross-correlation plots at lag 1 (CC1) and lag 2 (CC2) with simulation bands. The top panels show CC1 plots at $h = 60$, 30 min for the pair of stations (1, 4) and (3, 4), respectively. The bottom panels show CC2 plots at $h = 20$, 10 min for the pair of stations (1, 3) and (2, 3), where the station labels are TOWE (1), NEWH (2), DODD (3), DROP (4).

CONCLUSIONS

We have considered univariate and multivariate MMPP models to analyse rainfall data collected in the form of bucket tip time series. The univariate model with three states gave a good description of the structure and properties of the rainfall process individually at the four stations studied. The characteristics of the cumulative rainfall were all reproduced well by this model at all aggregation levels considered. The multivariate model reveals its potential to

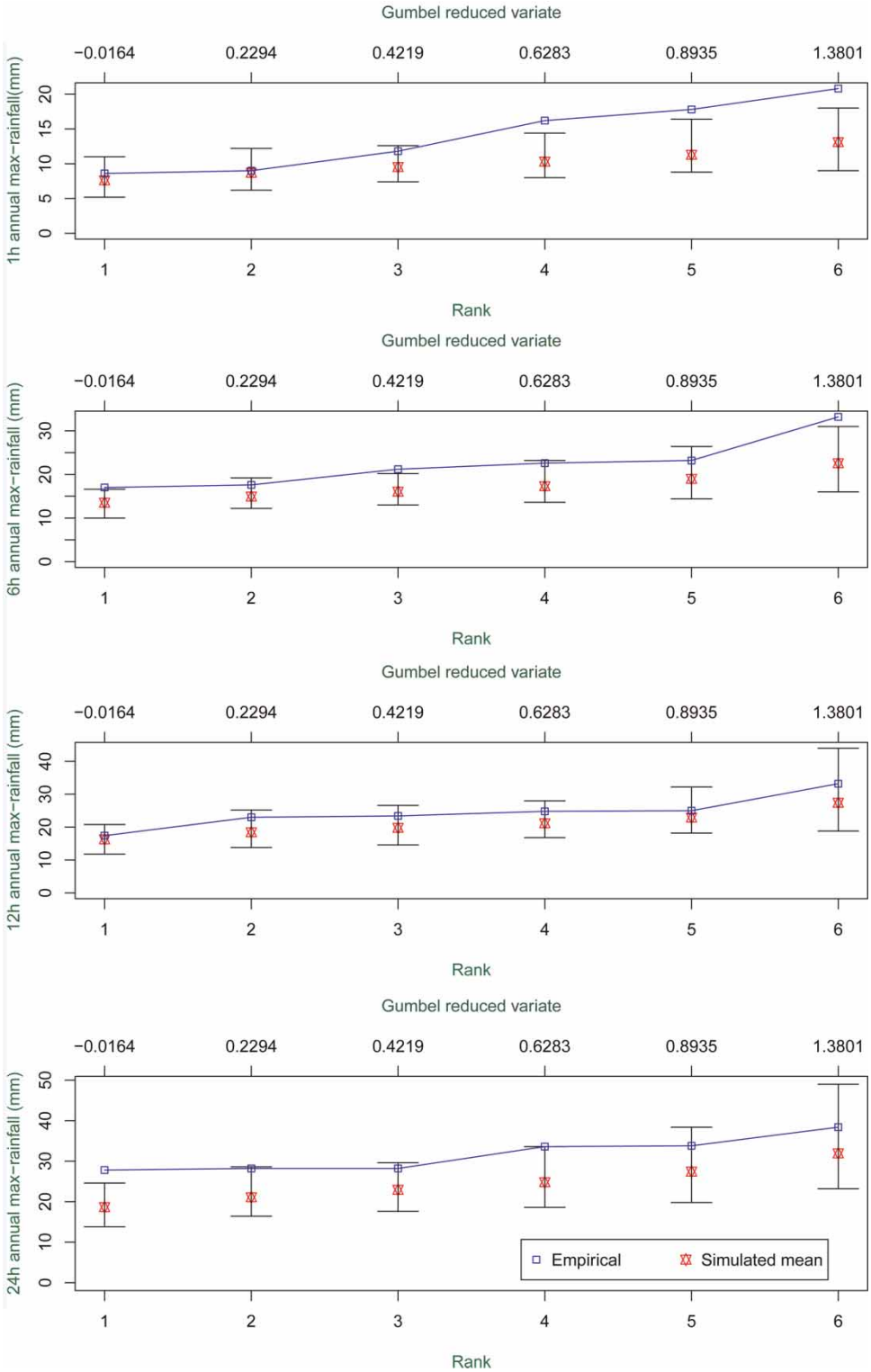

Figure 6 | Plots of the annual maximum rainfall against their Gumbel reduced variates and their ranks for DODD station at different levels (1, 6, 12 and 24 hours) of aggregation. The ranks are displayed on the *x*-axis (bottom), and the actual Gumbel reduced variates are shown at the top of the panel. Blue solid lines connecting square icons show empirical annual rainfall maxima and the interval plots display the maximum, minimum and the mean (red star icons) of the annual maxima from 100 copies of the data generated by the fitted model, at each aggregation level.

provide a strong modelling framework to jointly analyse the pattern of rainfall at multiple stations in the catchment area. The linked model reproduced the properties of aggregated rainfall well at all four stations at different aggregations. In addition, the multivariate model M2 performed well in capturing the cross-correlation features of the rainfall between the four stations. This is an area that requires further research to obtain improved fit, and we intend to incorporate covariates into our modelling to achieve this. The extreme value analysis carried out suggests that our model has limited ability to reproduce extremes, particularly at small aggregation levels. Nevertheless, the results presented in this paper indicate that this class of models is capable of providing a sound basis for spatial temporal model development.

REFERENCES

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

Cowpertwait, P. S. P. 1994 A generalized point process model for rainfall. *Proceedings of the Royal Society London A* **447** (1929), 23–37.

Cox, D. R. 1955 Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, Series B (Methodological)* **17** (2), 129–164.

Davison, A. C. & Ramesh, N. I. 1993 A stochastic model for times of exposures to air pollution from a point source. In: *Statistics in the Environment* (V. Barnett & K. F. Turkman eds). Wiley, New York, pp. 123–138.

Davison, A. C. & Ramesh, N. I. 1996 Some models for discretized series of events. *Journal of American Statistical Association, Theory and Methods* **91** (434), 601–609.

Heffes, H. & Lucantoni, D. M. 1986 A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications* **4** (6), 856–868.

Kigobe, M., McIntyre, N., Wheater, H. & Chandler, R. 2011 Multi-site stochastic modelling of daily rainfall in Uganda. *Hydrology Science Journal* **56** (1), 17–33.

Meier-Hellstern, K. S. & Fischer, W. 1993 The Markov modulated Poisson process (MMPP) cookbook. *Performance Evaluation* **18** (2), 149–171.

Neuts, M. F. 1989 *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, New York.

Onof, C. & Wheater, H. S. 1993 Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology* **149** (1–4), 67–95.

Onof, C. & Wheater, H. S. 1994 Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology* **157** (1–4), 177–195.

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S. & 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., Yameundjeu, B., Paoli, J. P. & Ramesh, N. I. 2002 A Markov modulated Poisson process model for rainfall increments. *Water Science and Technology* **45**, 91–97.

Ramesh, N. I. 1995 Statistical analysis on Markov-modulated Poisson processes. *Environmetrics* **6**, 165–179.

Ramesh, N. I. 1998 Temporal modelling of short-term rainfall using Cox processes. *Environmetrics* **9**, 629–643.

Ramesh, N. I., Onof, C. & Xie, D. 2012 Doubly stochastic Poisson process models for precipitation at fine time-scales. *Advances in Water Resources* **45**, 58–64.

Rodriguez-Iturbe, D., Cox, R. & Isham, V. 1987 Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society London A* **410** (1839), 269–288.

Ross, S. M. 1983 *Stochastic Processes*. Wiley, New York.

Ryden, T. 1996 An EM algorithm for estimation in Markov-modulated Poisson processes. *Computational Statistics & Data Analysis* **21** (4), 431–447.

Skaug, H. J. 2006 Markov modulated Poisson processes for clustered line transect data. *Environmental and Ecological Statistics* **13** (2), 199–211.

Smith, R. L. 1984 Contribution to the discussion of Stern, R.D. and Coe, R. 1984 A model-fitting analysis of daily rainfall data (with discussion). *Journal of Royal Statistical Society Series A* **147** (1), 1–34.

Smith, J. A. & Karr, A. F. 1983 A point process model of summer season rainfall occurrences. *Water Resources Research* **19** (1), 95–103.

Thompson, C. S., Thomson, P. J. & Zheng, X. 2007 Fitting a multisite daily rainfall model to New Zealand data. *Journal of Hydrology* **340** (1–2), 25–39.

Wheater, H. S., Chandler, R. E., Onof, C. J., Isham, V. S., Bellone, E., Yang, C., Lekkas, D., Lourmas, G. & Segond, M.-L. 2005 Spatial-temporal rainfall modelling for flood risk estimation. *Stochastic Environmental Research and Risk Assessment* **19** (4), 403–416.

Wheater, H. S., Isham, V. S., Cox, D. R., Chandler, R. E., Kakou, A., Northrop, P. J., Oh, L., Onof, C. & Rodriguez-Iturbe, I. 2000 Spatial-temporal rainfall fields: modelling and statistical aspects. *Hydrology and Earth System Sciences* **4** (4), 581–602.

Wood, S. J., Jones, D. A. & Moore, R. J. 2000 Accuracy of rainfall measurement for scales of hydrological interest. *Hydrology and Earth System Sciences* **4**, 531–543.

Zucchini, W. & Guttorp, P. 1991 A hidden Markov model for space time precipitation. *Water Resources Research* **27** (8), 1917–1923.

First received 15 November 2011; accepted in revised form 15 June 2012. Available online 2 January 2013