Stochastic models of rainfall

P. J. Northrop

Department of Statistical Science, University College London, UK, WC1E 6BT; email: p.northrop@ucl.ac.uk

Annu. Rev. Stat. Appl. 2023. 11:1–27

https://doi.org/10.1146/

Keywords

climate change, flood risk assessment, point process, Poisson cluster process, stochastic-mechanistic model, stochastic rainfall generation

Abstract

Rainfall is the main input to most hydrological systems. To assess flood risk for a catchment area, hydrologists use models that require long series of sub-daily, perhaps even sub-hourly, rainfall data, ideally from locations that cover the area. If historical data are not sufficient for this purpose, an alternative is to simulate synthetic data from a suitably-calibrated model. We review stochastic models that have a mechanistic structure, intended to mimic physical features of the rainfall processes, and are constructed using stationary point processes. We describe models for temporal and spatial-temporal rainfall and consider how they can be fitted to data. We provide an example application using a temporal model and an illustration of data simulated from a spatial-temporal model. We discuss how these models can contribute to the simulation of future rainfall that reflects our changing climate.

Contents

1. INTRODUCTION

Rainfall is a precious resource but it also presents potential risks. Characterising the variability of rainfall is vital in many areas of hydrology, including water resource management, design of drainage systems, pesticide fate management and for planning against floods, droughts and landslides. When suitable data are not available, a stochastic rainfall generator may be used to simulate realistic synthetic data. Applications like flood risk assessment and urban drainage design require long series of hourly, or even sub-hourly, rainfall data and a fine-scale spatial resolution may be needed for some catchment areas (Segond et al. 2007). These data are used as input to hydrological models that determine, based on the characteristics of the catchment, the extent and spatial distribution of excess water with which the system cannot cope. The consequences of rainfall can be sensitive to antecedent conditions, that is, the effects on the catchment area of rainfall over the preceding period. Therefore, flood risk assessments are made by running rainfall sequences through hydrological models continuously over long time periods, so that variability in conditions is represented (Wheater 2002, Pathiraja et al. 2012, Verhoest et al. 2010). For a discussion of issues concerning rainfall input to hydrological models see Chandler et al. (2014).

Empirical statistical models have been used successfully to describe the behaviour of daily rainfall. See, for example, Bellone et al. (2000), Wilks (1998) and Yang et al. (2005). These models are less feasible for sub-daily time scales, where the structure of rainfall is more complex, varying over short time scales and distances. For sub-daily data, stochastic-mechanistic models have been developed. These models attempt to represent, in an idealised way, the underlying mechanisms and structures of rainfall processes. This review focuses on

this type of stochastic model. Other approaches that can simulate rainfall at any required fine temporal and/or spatial scale are described in Section 8.

The idea of a stochastic-mechanistic model for rainfall dates back to LeCam (1961), with further contributions from Waymire & Gupta (1981), Waymire et al. (1984) and Rodriguez-Iturbe et al. (1984). A key development was made by Rodriguez-Iturbe et al. (1987) who analysed a model (described in Section 4.1) for rainfall at a point in space based on a clustered point process of pulses of rainfall with finite duration. This was the first model to provide an acceptable representation of rainfall over different timescales. Many further developments have been made and a spatial-temporal model has been used in the UK for making climate change projections (Jones et al. 2009). These models have been used extensively in temperate climates, particularly in Europe, but also in other regions and climates, including a Mediterranean climate in Cyprus (Camera et al. 2017), a semi-arid climate in Pakistan (Forsythe et al. 2014), tropical climates in Malaysia (Daud et al. 2016) and Vietnam (Vu & Mishra 2020) and various climates in Australia (Wasko et al. 2015). See Onof et al. (2000) for a review of models and Chen et al. (2021) for a review that focuses on their use in flood risk assessment.

2. STOCHASTIC-MECHANISTIC RAINFALL MODELS

The models that we consider are constructed using stochastic processes of point events. Each event gives rise to a basic rainfall element called a rain cell, or cell, for short. We review temporal models of the evolution of rainfall over time at a single site and spatial-temporal models of the distribution of rainfall in space and time. Parameters of a model govern the random process by which cells occur, the distribution of the time for which they produce rainfall and the evolution of the intensity of the rainfall that they produce over this duration. Models with a spatial component have parameters that relate to the spatial extents of cells and perhaps to the velocity with which they move.

The aim is to construct a stochastic model, with a modest number of parameters, that provides an idealised representation of the physical mechanism of the rainfall process. A key feature is a hierarchical clustering structure (Waymire & Gupta 1981) in which small ($10-50$ km^2) short-lived ($30-40$ minutes) cells cluster within a larger, longer-lasting storm, with storms clustering within a rain event that covers an area of 10^4 km^2 or more and lasts several hours. There are two broad types of rainfall: convective rainfall is short-lived and intense, whereas stratiform rainfall is more prolonged and less intense. The stochastic processes on which these models are based are stationary, so adjustments need to be made to account for seasonality, systematic spatial effects and long-term changes in climate.

Simplifying assumptions are made for mathematical convenience, to make it easier to derive properties of the model for use in fitting it to data (see Section 6.1). It is common to suppose that the intensity of a cell is constant over its duration, resulting in a rectangular pulse of rainfall. This produces realisations of rainfall intensity with an unrealistically regular appearance but this feature may not be evident in rainfall totals accumulated over a time period of an hour, say. Typically, durations of rain cells are assumed to be independently and exponentially distributed.

Temporal models are constructed in continuous time and therefore they could be used to infer properties of rainfall at any time scale of interest. For example, they are typically fitted to observed rainfall totals accumulated over discrete time intervals, using a Generalised Method of Moments (GMOM) based on selected statistical properties of observed

data. If we have available a dataset containing hourly rainfall totals then these totals can be aggregated to form six-hourly totals and daily totals. Similarly, we can perform this aggregation of scale for the model to derive key properties of hourly, six-hourly and daily rainfall totals. Data at different scales provide different information about the rainfall process and therefore we may choose to include in the fitting procedure properties from a range of scales. The construction of these models enables us to do this. Spatial-temporal models can be constructed, either in continuous space and time or in continuous time at a discrete set of sites.

A fitted model is validated by checking that it predicts well features of rainfall at scales other than those used to fit it. The model can, in principle, be used to provide inferences at scales that are shorter than the smallest scale available in the data. For example, we could fit a model to hourly rainfall totals and simulate sequences of five-minute totals. Such high resolution design rainfall data are required for applications like urban drainage design (Onof & Arnbjerg-Nielsen 2009). This kind of extrapolation may not be warranted. It would be better to have available good quality sub-hourly data with which to fit and/or validate a model and a refinement of the rectangular pulses assumption may be necessary if realistic behaviour over very short timescales is required. We return to this issue in Section 4.2 and discuss statistical inference in Section 6.

3. RAINFALL DATA

We describe the main types of rainfall data and discuss some issues to consider when modelling these data. Rain gauges measure "ground truth", the amount of water that falls to the ground at a particular location. Rainfall radar provide an indirect measurement of rainfall at a particular level of the atmosphere, via a conversion to a rainfall intensity of the energy within a radar beam reflected back to the radar station.

3.1. Rain gauge data

Early rain gauges were containers with a measurement scale in which rain water collects. The rainfall amounts are recorded manually at regular intervals, usually daily, when the container is emptied. To fit the models described in Section 4, data of higher temporal resolution, for example, hourly, are required. These are provided by automated gauges. A tipping-bucket rain gauge contains a small bucket that tips and empties when it fills with a fixed amount of rain, typically 0.1mm or 0.2mm. The number of tips within a time interval provides an estimate of the rainfall total for the interval, which enables series of rainfall totals of a given temporal resolution to be constructed. Some more modern gauges use a similar principle, but are based on the weight of rainfall collected in a larger bucket, while others produce a continuous trace on a recording chart as rainfall displaces a float within the gauge. For these gauges, discreteness of the data is less evident than for tipping-bucket gauges.

As an example of rain gauge data, we use 10-minute rainfall totals provided by German Weather Service (Deutscher Wetterdienst, DWD) for Bremen, Germany from September 1995 to March 2023. **Figure 1** shows the full series and subsets for January 1998 and January 2023. These subsets have been selected to contain at least a moderate amount of rain. We will fit a model to January data in Section 7.1. Seasonality is evident in the full time series, with the largest totals tending to occur during June to August. We

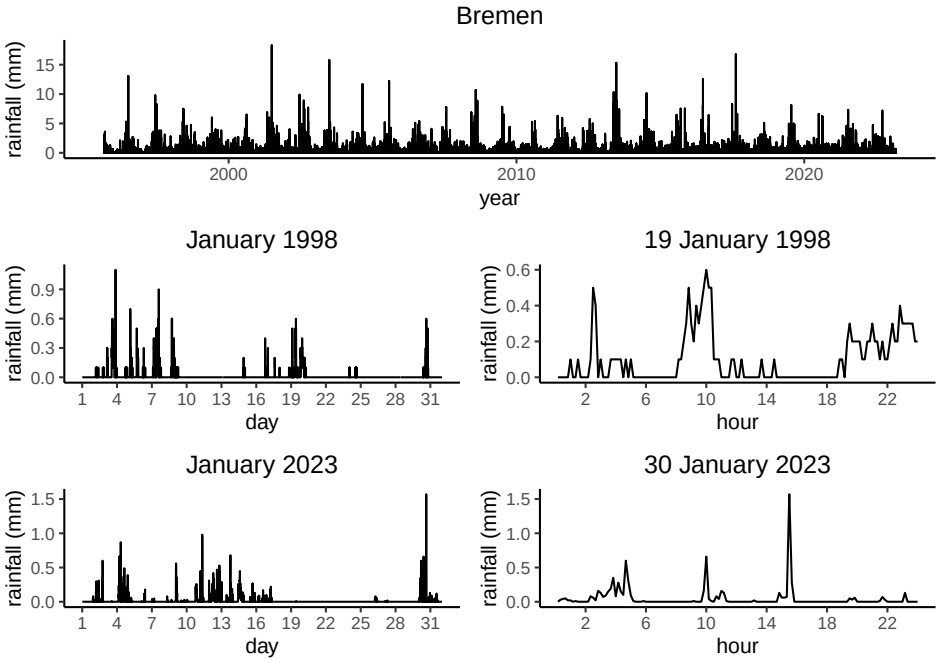

Figure 1

Series of 10-minute rainfall totals from Bremen, Germany (DWD station 691). *Top*: 1 September 1995 to 9 March 2023. *Middle left*: January 1998. *Middle right*: 19 January 1998. *Bottom left*: January 2023. *Bottom right*: 30 January 2023. Source: Deutscher Wetterdienst (DWD).

performed a basic check of these 10-minute totals by aggregating them to hourly totals and comparing them to an equivalent hourly dataset provided by DWD. Seven 10-minute totals, from 15 and 18 October 2004, were corrected, because they were a factor of ten too large. Otherwise, the correspondence was close enough to be explained by rounding of values in the datasets. If data are available from a network of gauges experiencing the same weather systems then a plot of the rainfall accumulated over a month for each gauge may reveal likely errors (Wheater et al. 2000). Until October 2008, a 0.1mm tipping-bucket gauge was used. The fact that measurements are multiples of 0.1mm is evident in **Figure 1** in the plot for 19 January 1998. Subsequently, weight-based gauges provide measurements that are multiples of 0.01mm, resulting in smoother time series. Overall, approximately 1% of the observations are missing. There is no obvious pattern in proportion of missing values over months but there is a downwards trends over years with most years after the installation of the weight-based gauge having no missing values.

We should be aware of potential problems when using high resolution rain gauge data, such as 5-minute or 10-minute data. A tipping-bucket gauge may attribute rainfall to a time period later than when it actually fell, especially during periods of light rainfall. Weight-based gauges are not perfect (Saha et al. 2021) but will tend to be more reliable over short time scales. The coarseness of the discrete scale on which rainfall totals are recorded may have a non-negligible impact. Making inferences using summary statistics calculated from

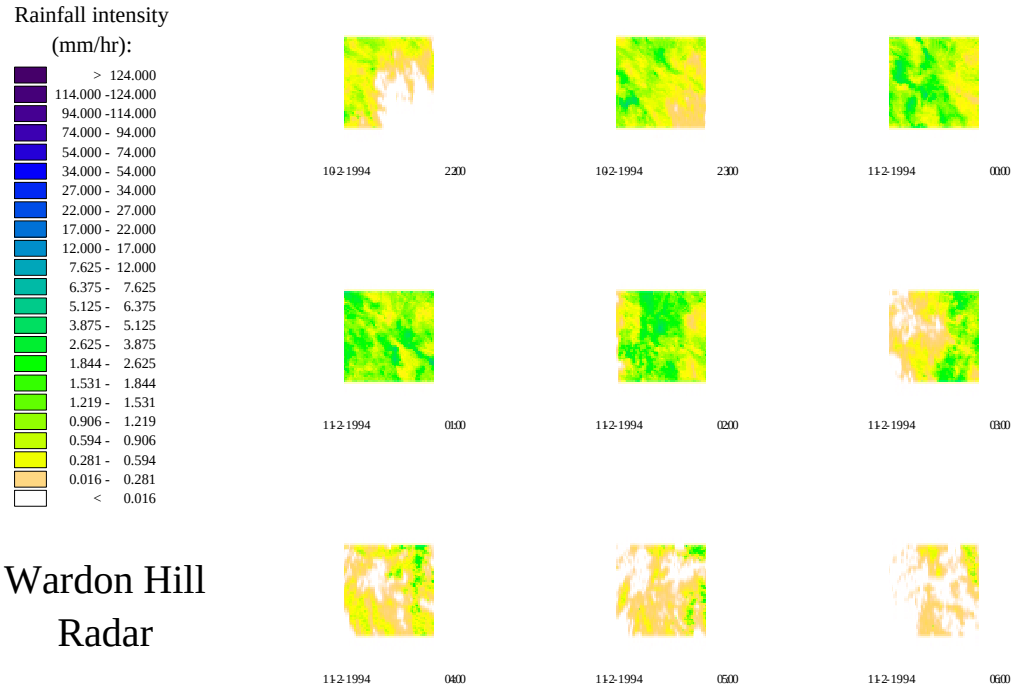

Figure 2

Radar images from the Wardon Hill Radar station in Dorchester, UK from 10pm on 10 February 1994 to 6am on 11 February 1994. Successive images are one hour apart. Each image is a 52 by 52 array of 2km by 2km pixels. The colour key indicates the average rainfall intensity for each pixel. Source: CEH Wallingford (2007).

raw data may provide some protection from these issues.

3.2. Radar data

A radar image provides estimates of average rainfall intensity over spatial grid squares called pixels. Pixel sizes are typically of size 1 km^2, 4 km^2 or 25 km^2. Normally, each image is the result of a 1-minute sweep of the radar and is therefore treated as an instantaneous snapshot of a rainfall field. **Figure 2** shows a sequence of radar images recorded by the Wardon Hill radar in South-West England over an 8-hour period in February 1994. One image is presented each hour. In the raw data an image is available every 5 minutes.

Even after processing to remove anomalies and artefacts, radar measurements may not be consistent with those provided by comparable rain gauges. Agreement may be improved by calibration of the radar images using gauge data, but this is non-trivial problem, in

part because radar images provide spatial averages and gauges point values. See Ochoa-Rodriguez et al. (2019) for a review of calibration and related methods.

Calibration issues aside, radar data have the advantage that they provide rich spatial information of high temporal resolution over a large area. This is useful if we wish to study spatial-temporal rainfall fields. The area covered in **Figure 2** is a little over 10^4 km^2, which is large enough to capture much of the interior of a rain event. There are applications, such as storm drainage modelling for urban catchments, where even 1 km^2 resolution radar data is not of sufficiently high resolution and point data provided by rain gauges are required (Ramesh et al. 2013).

4. TEMPORAL MODELS

4.1. Poisson cluster models

Recall that models are based on a stochastic point process of arrival times of rain cells. For most applications, it is necessary for cells to cluster. We call a cluster of cells a storm. The simplest clustered models use a stationary Poisson cluster process of cells represented by rectangular pulses of rainfall, with the following structure (Rodriguez-Iturbe et al. 1987).

1. The times at which storms originate are generated by a Poisson process of rate λ.
2. Each storm origin produces a random number C of cells.
3. Within a storm, the arrival times of cells are distributed randomly in time relative to the storm origin.
4. Each cell is a rectangular pulse of rainfall, with random intensity X lasting for a random duration D.

The clusters of cells associated with distinct storms are independent. Within a particular storm, distinct cells are independent, as are X and D for a given cell. Cells can overlap in time, even those generated from different storms, and there may be periods of zero rainfall intensity within a storm. The total rainfall intensity $Y(t)$ at time t is the sum of intensities from all cells present at time t.

Two main clustering mechanisms have been used. For a Neyman–Scott process, we specify a distribution for C and another distribution from which the temporal displacements of cell arrivals from their respective storm origins are independently and identically sampled. In a Bartlett–Lewis process, the intervals between the starting times of successive cells in a storm are independent and identically distributed. In the formulation typically used, the lifetime L of a storm has an exponential distribution with mean $1/\gamma$ and the first cell in a storm is located at the storm origin with further cells arriving at a rate β in a Poisson process that starts at the storm origin and terminates at the storm lifetime. It follows that C has a geometric distribution with mean $\mu_C = 1 + \beta/\gamma$. We refer to the resulting models as the Neyman–Scott Rectangular Pulses Model (NSRPM) and the Bartlett–Lewis Rectangular Pulses Model (BLRPM). It is convenient to assume an exponential distribution, with mean $1/\eta$, for D. In particular, this preserves the Markovian structure built into the specification of the BLRPM.

Figure 3 summarises the structure of the BLRPM. It shows three storm origins and their respective termination times, the arrival times of the cells within these storms, the rectangular pulses and the superposition of these intensities to produce the total rainfall intensity at a given time. The first (red) storm has a period of zero intensity within it and a rain cell that continues to produce rain beyond its termination point. There are several

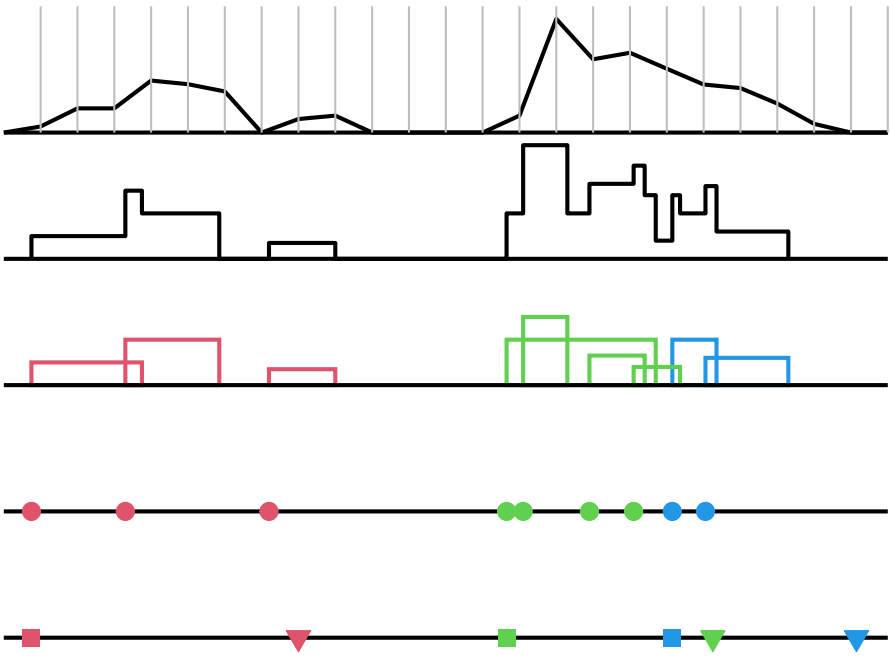

Figure 3

Illustration of the structure of the BLRPM. There are three storms, differentiated by red, green and blue colouring. *Bottom*: storm processes start at the squares and terminate at the triangles. *Second bottom*: rain cells start at the circles. *Middle*: rectangular pulses for each cell. *Second top*: rainfall intensities from the superposition of cells. *Top*: aggregation of intensities to form totals for each of 24 periods, indicated by vertical grey lines.

cells that overlap and the last cell of the second (green) storm overlaps with the first cell of the third (blue) storm. The top of the schematic shows the rainfall totals produced by aggregating rainfall intensities over disjoint time intervals to form rainfall totals, such as 24 hourly totals for one day. These are the values that we observe in rain gauge data. We see that aggregation can hide some of the fine scale structure present in the intensity process.

The NSPRM and BLRPM each have a minimum of five parameters. They have been used widely to model rainfall data, but there is no clear empirical evidence that one clustering mechanism should be preferred over the other. In Section 7.1 we use the BLRPM in an example application, where we suppose that X has an exponential distribution. The properties typically used to fit this type of model depend only on the moments of X, so it is not difficult to relax this assumption.

We give a brief outline of the mathematical representation of these model processes. See Rodriguez-Iturbe et al. (1987) and Rodriquez-Iturbe et al. (1988) for details. Suppose that cells arrive at times $\tau_1, \tau_2, \ldots$. The total rainfall intensity $Y(t)$ at time t sums contributions from these cells. It can be expressed as

$$Y(t) = \sum_{\tau_i \leq t} I(\tau_i; t)\, X(\tau_i),$$

$$= \int_{\tau=-\infty}^{t} I(\tau;t)\, X(\tau)\, \mathrm{d}N(\tau), \qquad 1.$$

where $I(\tau;t) = 1$ if a cell that arrived at time τ is alive at time t and $I(\tau;t) = 0$ otherwise, $X(\tau)$ is the intensity of this cell, and $\{N(t)\}$ counts occurrences in the clustered process of cell arrival times. Owing to the assumed independence of cell durations, intensities and the cell arrival process, we have

$$\begin{aligned}
\mathrm{E}[Y(t)] &= \int_{\tau=-\infty}^{t} \mathrm{E}[I(\tau;t)]\, \mathrm{E}[X(\tau)]\, \mathrm{E}[\mathrm{d}N(\tau)], \\
&= \lambda\mu_C \mathrm{E}(X) \int_{\tau=-\infty}^{t} P(D > t-\tau)\, \mathrm{d}\tau, \\
&= \lambda\mu_C \mathrm{E}(X)\mathrm{E}(D),
\end{aligned}$$

where $\mathrm{E}[\mathrm{d}N(\tau)] = \lambda\mu_C\, \mathrm{d}\tau$ because the arrival rate of cells is $\lambda\mu_C$ per unit time. Rodriguez-Iturbe et al. (1987) derive the variance, skewness and autocovariance function of $Y(t)$ and $P(Y(t) = 0)$ and adapt these properties to the aggregated process, $\{Y_i^{(h)}, i = 1, 2, \ldots\}$, where

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)\, \mathrm{d}t, \qquad 2.$$

because rain gauge data provides cumulative rainfall totals over intervals of length h.

4.2. Developments and extensions

In the original specification of the BLRPM and NSRPM the parameters β, γ and η that govern the temporal structure within a storm are constant across storms. Rodriquez-Iturbe et al. (1988) extended the BLRPM in a way that preserves the within-storm structure, but produces storms that occur on different timescales. They fix the dimensionless parameters $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$ and randomise the cell duration parameter η between storms according to a gamma distribution. Within a storm, all cells have the same mean duration. Fixing κ and ϕ means that storms with a relatively short mean cell duration, $1/\eta_s$, say, have a correspondingly short mean storm lifetime $1/\gamma_s$ and mean cell inter-arrival time $1/\beta_s$. The random η BLRPM has one extra parameter compared to the fixed η version and resulted in an improved representation of wet/dry rainfall properties in a case study presented in Rodriquez-Iturbe et al. (1988). The NSRPM can be modified in the same way (Entekhabi et al. 1989).

We provide a non-exhaustive summary of other contributions to the literature for stochastic-mechanistic temporal models of rainfall. See also the review of Onof et al. (2000). One area is the derivation of new properties for model fitting, such the skewness of marginal rainfall totals and properties of wet and dry periods. See, for example, Onof et al. (1994), Cowpertwait (1998) and Kaczmarska (2013).

Ramesh (1998) replaces the Poisson cluster process of rectangular rain cell arrivals with a Cox process, in the form of a Markov-modulated Poisson process (MMPP) that switches between two states, with respective high and low rates of rain cell arrival. This model does not involve explicit mechanistic features like rain cells, but it does involve interpretable features like the movements between characteristic rainfall intensity states and mean times spent in those states. Ramesh et al. (2012) use a model of the same form, generalised to allow more than two states, to model directly rain gauge tip-times. It provides effective

modelling of rainfall over timescales as short as 5 minutes. Issues concerning the conversion of tip-times to rainfall totals are avoided and, unlike models for rainfall totals based on rainfall pulses, likelihood-based inference can be used.

Many refinements have been proposed with the aim of improving realism and/or the ability to reproduce particular observed rainfall properties. We categorise them according to their nature and cite example literature: inducing dependence between cell intensity and duration, via a joint distribution (Kakou 1997, Evin & Favre 2008) or using multiple cell types (Cowpertwait 1994); using non-rectangular cell shapes (Park et al. 2021, Ramesh et al. 2018); extending the random η BLRPM to include randomisation of cell intensities over storms (Kaczmarska et al. 2014); representing cells as a random number of instantaneous bursts of rainfall (Cowpertwait et al. 2007). Some approaches are aimed at improving the reproduction of extreme values, an issue considered in Verhoest et al. (2010). In Cameron et al. (2000) large rain cell intensities are modelled using a generalised Pareto distribution. Cross et al. (2018) develop a general strategy in which small rainfall totals are censored. This reduces the influence of data from periods of light rainfall and focuses modelling on the larger totals.

4.3. Covariate effects

The construction of these models uses a stationary process. Seasonality can be represented by fitting separately to data from different months, but further adjustment is required to incorporate the trends anticipated owing to our changing climate. Wasko & Sharma (2017) use a log-linear regression to model the effect of temperature on rainfall statistics to be used in fitting. Evin & Favre (2013) model the storm arrival rate parameter as increasing linearly in time. Thayakaran & Ramesh (2013a) build a MMPP model for rain gauge tip-times with log-linear regression effects for temperature, sea level pressure and relative humidity. Simple parametric relationships may not be realistic. Kaczmarska et al. (2015) develop a nonparametric regression approach that allows estimates of BLRPM parameters to change smoothly with the monthly means of atmospheric variables, in a way that is determined directly by the data. We describe this in Section 6.1. A key finding is that atmospheric variables involved in the physical processes that generate rain explain far more variation in rainfall observations than a categorisation by month. Cross et al. (2020) uses a similar approach to represent effects of temperature.

5. SPATIAL-TEMPORAL MODELS

Models for the temporal evolution of rainfall at a single-site are useful, but to perform flood risk assessment for a particular catchment area how rainfall is distributed over space and time is often important (Segond et al. 2007). Depending on the nature and size of the catchment, it can matter greatly both where and when rain falls. We expect to observe spatial dependence in rainfall sequences at nearby locations as they will tend to experience similar rainfall patterns when rain events move over them. In a multisite approach, modelling is performed in discrete space at a set of sites, with each site having its own temporal process. The aim is to link processes at different sites to reflect systematic differences between sites and induce an appropriate spatial correlation structure. We return multisite modelling briefly in Section 5.2, but first we consider an alternative approach in which the rainfall process is modelled in continuous space and time, which has the advantage that we

can simulate rainfall at any space-time resolution.

5.1. Continuous space-time

Models are direct extensions of those described in Section 4.1, based now on a stochastic process of points occurring in time and two-dimensional space. Each point provides the time at which a rain cell starts to produce rainfall and the location of the centre of the cell at this time. The centre of the cell may move over time to represent the movement of a rain event. As before, cells have random intensities and durations. During its lifetime, a cell deposits rainfall over an area surrounding its centre.

Cowpertwait (1995) (see also Cowpertwait et al. (2002), Leonard et al. (2008) and Cowpertwait (2010)) and Northrop (1998) (see also Chandler et al. (2007)) have developed models of this type, building on Cox & Isham (1988). These models have much in common, for example, they both suppose that a cell's intensity is constant over its area, but with differences in their construction influenced by the type of data to which they are typically fitted, that is, time series of rainfall totals from multiple rain gauges for Cowpertwait (1995) and sequences of radar images for Northrop (1998). We describe the structure of these models and then discuss the differences.

In the Cowpertwait (1995) model, cells are clustered within storms in time according to an (exponential) Neyman–Scott process. Cells do not move. For each storm, cells are distributed over space in a Poisson process, the rate of which determines the mean number of cells per storm. Cells are circular and categorised into n_c types, each type having a set of parameters to describe its distribution of intensity, duration and size and its contribution to the overall rate of the spatial Poisson process. The minimum number of parameters is $2 + 4n_c$.

In the Northrop (1998) model, storms arrive in a Poisson process. We call the spatial location at which a storm arrives its storm centre. Storm centres, and all cells, move with velocity $\boldsymbol{v}$. Cells are elliptical in shape, with eccentricity e, and orientated at an angle ω, to reflect anisotropy in rainfall structure. Cells are clustered within storms in time using a Bartlett–Lewis process and in space using a Neyman–Scott mechanism. For the latter, cell centres are displaced from the current location of their storm centre by a vector drawn from a bivariate Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, which is parameterised to produce elliptical Gaussian density contours with eccentricity e and orientation ω. Thus, storms and cells have the same shape and orientation. The remaining free parameter in $\boldsymbol{\Sigma}$ reflects storm size. Distributions are needed for cell intensity, duration and size and for storm size. The minimum number of parameters in the model is 11. **Figure 4** illustrates the spatial structure of this model, including averaging of intensities over radar pixels.

For both models, cell durations are assumed to be exponentially distributed and strong independence assumptions, analogous to those in Section 4.1, are made for mathematical tractability. The representation of the total rainfall intensity $Y(\boldsymbol{u}, t)$ at location $\boldsymbol{u} = (u_1, u_2)$ and time t is analogous to Equation 1, with an extra two-dimensional integral over space. Aggregation over time for modelling rain gauge totals is performed using the equivalent of Equation 2 and averaging over space for modelling radar data using

$$Y_{i,j}^{(h)}(t) = \frac{1}{h^2} \int_{(i-1)h}^{ih} \int_{(j-1)h}^{jh} Y(\boldsymbol{u}, t)\, \mathrm{d}u_1 \mathrm{d}u_2. \qquad 3.$$

The main ways in which these models differ are in their treatment of spatial structure

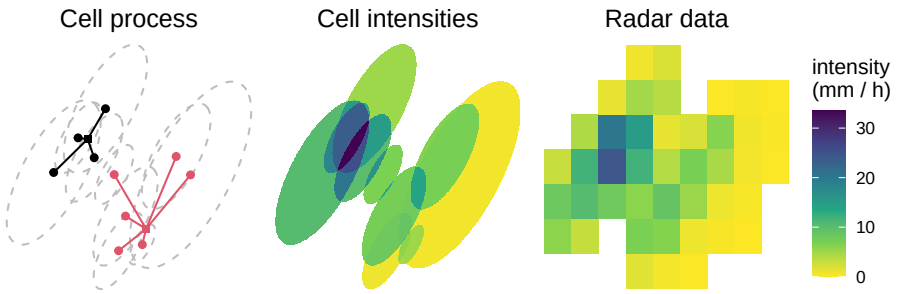

Figure 4

Illustration of the spatial structure of the Northrop (1998) model, containing two storms. *Left*: locations of cells. The squares are the storm centres and the circles the cell centres. The black storm has four cells and the red storm six cells. The dashed curves indicate the extents of the elliptical cells. *Middle*: summing cell intensities from different cells. *Right*: spatial aggregation to create average intensities over radar pixels.

(multiple cell types vs clustering), cell movement (none vs some) and cell/storm shape (circular vs elliptical). The Cowpertwait (1995) model has been fitted to monthly sequences of hourly totals from networks of rain gauges. These data are the result of rainfall accumulations from multiple rain events. The model is able to capture the effect of these accumulations without isolating individual rain events for fitting. The Northrop (1998) model has been fitted to sequences of radar images from the interiors of individual rain events, such as those in **Figure 2**, enabling the capture of spatial clustering, movement of rain events and anisotropy in these images. Further modelling is needed to generate realistic sequences of events, using a renewal process to generate wet and dry spells and basing the orientation of leading and trailing edges of rain events on those of historical events (Wheater et al. 2006, Chandler et al. 2007).

5.2. Multisite

Cox & Isham (1994) discuss extensions of the models of Section 4.1 to multiple sites. A master process of cell arrivals is decomposed into subprocesses. All cells in a given subprocess have the property that they affect the same combination of sites. Constructing a parsimonious model with a realistic space-time correlation structure is challenging, arguably more so than developing a fully continuous spatial-temporal model. Kakou (1997) extends this work and fits models to monthly sequences of hourly rain gauge totals from three sites. Favre et al. (2002) develop a two-site model in which correlation between sites is induced using bivariate distributions for the numbers of cells at these sites, their intensities and durations. Thayakaran & Ramesh (2013b) and Ramesh et al. (2013) develop multisite extensions of the Cox processes models for tip-times described in Section 4.2, the latter incorporating covariate effects of atmospheric variables. Models are fitted to data from four sites. These approaches are useful when a relatively small number of locations are of interest.

6. STATISTICAL INFERENCE

Consider a collection $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$ of rainfall observations, regarded as the realised value of corresponding random variables $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$ under a model. For rain gauge data $\boldsymbol{y}$ is a time series of rainfall totals and for radar data it is a sequence of arrays of gridded average rainfall intensities. To specify a full likelihood for a parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ we would need to derive the full joint distribution of all random variables contained in $\boldsymbol{Y}$. Apart from the MMPP models for tip-time data described in Section 4.2, this is not feasible, even for a simple temporal model based on a Poisson process of rectangular rain cells. Moreover, seeking to perform inference based on this likelihood may not be sensible. The locally-deterministic representations used for rain cells mean that at very local scales rainfall intensities tend to be more strongly dependent than observed in data. It is reasonable to expect these models to reproduce key statistical properties of rainfall data over different time scales, but not joint distributions of local collections of observations. A full likelihood-based approach would place undue emphasis on features that these models are not intended to reproduce. For an extended discussion of these points see Jesus & Chandler (2011). However, Jesus & Chandler (2017) and Northrop (2006) have performed inferences using a likelihood constructed based on particular aspects of rainfall data. In Section 6.2 we describe an approach that uses Approximate Bayesian computation (Beaumont 2019).

The method typically used to fit stochastic-mechanistic rainfall models to data is a Generalised Method of Moments (GMOM), in which parameter values are chosen to minimise an objective function based on differences between observed and fitted values of selected statistical properties. This approach has been used since these models were developed. Jesus & Chandler (2011) consider how best to implement GMOM for a given set of properties and produce estimates of parameter uncertainty. An important issue is how contributions from different properties should be weighted in the objective function.

6.1. Generalised method of moments

Initially, we consider making inferences about the parameters of a temporal rainfall model using data from a single rain gauge. Rainfall is a seasonal phenomenon, so it is common to fit models separately to data originating from a single month. For example, in Section 7.1 we extract from the full set of rainfall totals recorded at Bremen only those values recorded during the 28 Januaries present in the data. More generally, we suppose that m years of data provides m independent replicates of the rainfall process during a month of interest.

We specify a vector $\boldsymbol{T}$ of k summary statistics of observed rainfall. Example sets of statistics are given in Section 7.1. We must have $k \geq p$ and it is common to use $k > p$. We define separate summary statistic vectors $\boldsymbol{T}_1, \ldots, \boldsymbol{T}_m$ for each of m replicates and set $\overline{\boldsymbol{T}} = m^{-1} \sum_{i=1}^{m} \boldsymbol{T}_i$. We denote by $\boldsymbol{\mu}(\boldsymbol{\theta})$ the expectation of $\boldsymbol{T}$ under the model. The estimator $\widehat{\boldsymbol{\theta}}$ minimises the quadratic form

$$Q(\boldsymbol{\theta}; \overline{\boldsymbol{T}}) = \left[\overline{\boldsymbol{T}} - \boldsymbol{\mu}(\boldsymbol{\theta})\right]' \boldsymbol{W} \left[\overline{\boldsymbol{T}} - \boldsymbol{\mu}(\boldsymbol{\theta})\right], \qquad 4.$$

where $\boldsymbol{W}$ is a $k \times k$ matrix of weights. The choice of weighting matrix in Equation 4 will affect the performance of the estimator $\widehat{\boldsymbol{\theta}}$. Jesus & Chandler (2011) study the asymptotic distribution of $\widehat{\theta}$ as m tends to infinity, showing that the optimal choice, in terms of precision of estimation of θ, for $\boldsymbol{W}$ is $\boldsymbol{S}^{-1}$, where $\boldsymbol{S}$ is the covariance matrix of $\boldsymbol{T}$. In practice, $\boldsymbol{S}$ is

estimated using the sample covariance matrix of $\boldsymbol{T}_1, \ldots, \boldsymbol{T}_m$, that is,

$$\widehat{\boldsymbol{S}} = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\boldsymbol{T}_i - \overline{\boldsymbol{T}})(\boldsymbol{T}_i - \overline{\boldsymbol{T}})'. \quad 5.$$

Theoretically optimal performance may not be achieved if $\boldsymbol{S}$ is estimated inaccurately. Jesus & Chandler (2011) find that $\boldsymbol{W} = \text{diag}(\widehat{\boldsymbol{S}})$ produces comparable performance to $\boldsymbol{W} = \widehat{\boldsymbol{S}}$ and that it is necessary to perform large scale simulations to improve the estimation of $\boldsymbol{S}$ if theoretical superiority of the optimal choice of $\boldsymbol{W}$ is to be realised. Using $\boldsymbol{W} = \text{diag}(\widehat{\boldsymbol{S}})$, which gives greater weight to statistics that are estimated more precisely from the data, may be a good choice in practice. Both these forms of $\boldsymbol{W}$ perform better than using an identity matrix, which has been a common choice in the hydrological literature. The application might influence the importance of different statistics. Kim & Olivera (2012) studied this empirically, providing recommendations concerning the weighting of statistics for different types of application.

Kaczmarska et al. (2015) use a kernel-based local polynomial regression to model the effects of covariates on model parameters, with candidate covariate data sourced from the National Centers for Environmental Prediction (NCEP) reanalysis data (Kistler et al. 2001). We describe the special case of a local mean approach using a single covariate, for example, temperature. Suppose that the full data record spans n months. Let $\boldsymbol{T} = (\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n)$ denote the summary statistic vectors, $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ the corresponding monthly mean values of the covariate and x_0 a covariate value of interest. The objective function to be minimised is

$$Q_h(\boldsymbol{\theta}; \boldsymbol{T}, \boldsymbol{X}, x_0) = g(\boldsymbol{\theta}; \boldsymbol{T}, \boldsymbol{X}, x_0)' \, \boldsymbol{W}(x_0) \, g(\boldsymbol{\theta}; \boldsymbol{T}, \boldsymbol{X}, x_0), \quad 6.$$

where $g(\boldsymbol{\theta}; \boldsymbol{T}, \boldsymbol{X}, x_0) = n^{-1} \sum_{t=1}^{n} K_b(X_t - x)[\boldsymbol{T}_t - \boldsymbol{\mu}(\boldsymbol{\theta})]$, $K_b(x) = b^{-1} K(x/b)$ and $K(\cdot)$ is a kernel function to weight the contribution from $\boldsymbol{T}_t$ according to the proximity of $\boldsymbol{X}_t$ to x_0. The weighting matrix $\boldsymbol{W}(x_0)$ depends on x_0. This enables model parameters to be modelled flexibly as smooth functions of covariates. The degree of smoothing is determined by the bandwidth b, which is selected using cross-validation. An alternative approach using regression splines is discussed in Kaczmarska (2013).

6.1.1. Implementation issues. The choice of statistical properties and time scales included in $\boldsymbol{T}$ will depend on the intended application, but some general considerations apply. Algebraic expressions should be available for the expected properties under the model, to avoid the need to use simulation to estimate them. Collectively, they should contain information about the components of $\boldsymbol{\theta}$ but, to lessen potential problems with estimating $\boldsymbol{W}$, selecting a large number of properties should be avoided and each property should be estimable with reasonable precision. The closer the properties are to being statistically independent of each other, that is, the closer $\widehat{\boldsymbol{S}}$ in Equation 5 is to being diagonal, the stronger argument for using a diagonal $\boldsymbol{W}$.

The objective function is minimised numerically. This can be challenging and there is the possibility that it has local minima. Reparameterising the model, for example, working with the logs of the parameters, can improve stability of the minimisation and speed of convergence. Aryal & Jones (2020) and Aryal & Jones (2021) perform approximate Bayesian inference for rainfall models using a parameterisation that reduces posterior dependence between the parameters. Choosing good starting values for the minimisation algorithm can be important. The MOMFIT software (Chandler 2016) implements an effective strategy that repeats multiple minimisations from several random starting points.

A fitted model will tend to reproduce well the properties used to estimate its parameters. Validation of a model can be based on properties not explicitly used to fit it. See Section 7.1 for an example. At this stage, we may use simulation to estimate model properties that are important practically but for which algebraic expressions are not available, for example, properties of extreme rainfall.

6.1.2. Spatial-temporal models. In principle, spatial-temporal models can be fitted by selecting a set of summary statistics and minimising numerically the objective function given in Equation 4. However, authors have found that this minimisation tends to be unstable and that it is preferable to split the estimation into a sequence steps, with each step focused on a subset of the parameters. For example, Wheater et al. (2006) estimate the rain cell velocity $\boldsymbol{v}$ directly from a sequence of radar images, using an estimate of the maximum of the sample space-time cross-correlation function at a time lag of 5 minutes. The remaining parameters are partitioned into subsets, relating to spatial aspects, temporal aspects and marginal rainfall occurrence and amount, with a bespoke fitting step for each of these subsets. Cowpertwait et al. (2002) and Burton et al. (2010) devise approaches that are carefully structured to provide stability in the fitting of the model of Cowpertwait (1995) to rain gauge data, including scaling adjustments to account for systematic between-site variation. Willems (2001) and Muñoz Lopez et al. (2023) take a different approach, using an algorithm to identify individual cells using data from a network of rain gauges and from sequences of radar images, respectively.

6.2. Approximate Bayesian computation

Approximate Bayesian computation (ABC) is a method of inference for models where a likelihood is not available but it is easy to simulate from the model. In simple terms, ABC is a Bayesian implementation of GMM that uses an acceptance-rejection algorithm based on a criterion like $Q(\boldsymbol{\theta}; \overline{\boldsymbol{T}})$ in Equation 4. We simulate datasets using different proposal values of $\boldsymbol{\theta}$, use each of these datasets to estimate the expectations $\boldsymbol{\mu}(\boldsymbol{\theta})$ of the summary statistics under the model, and accept only those parameter vectors for which $Q(\boldsymbol{\theta}; \overline{\boldsymbol{T}})$ is smaller than a threshold ϵ. We accept only those values of $\boldsymbol{\theta}$ that result in a set of summary statistics that are close to those observed in the data. If the proposed values of $\boldsymbol{\theta}$ are simulated from a prior distribution $\pi(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ then the accepted values have a distribution that approximates the posterior distribution for $\boldsymbol{\theta}$ given the data. The nature of the approximation will depend on the choice of summary statistics and ϵ. For an overview of ABC see Beaumont (2019).

Aryal & Jones (2020) use ABC to fit the BLRPM to 6-minute rainfall totals from Victoria, Australia, including a random walk Metropolis–Hastings update step to propose preferentially values of $\boldsymbol{\theta}$ that have a non-negligible probability of being accepted. They include in $Q(\boldsymbol{\theta}; \overline{\boldsymbol{T}})$ the means and standard deviations of the lengths of wet and dry periods. Expressions are available for some of these properties (Onof et al. 1994), but not in closed form. Aryal & Jones (2021) use ABC to fit extended versions of the spatial-temporal model of Northrop (1998) to 1 km^2 resolution radar data recorded at Melbourne, Australia. The extensions relax assumptions made originally to enable properties of the model to be derived. For example, the intensity of a cell decays from its centre to its edge. To reduce computation time, the cell velocity and shape parameters are fixed at estimates calculated directly from the data. Properties relating the distributions of wet and dry areas in radar images are included in $Q(\boldsymbol{\theta}; \overline{\boldsymbol{T}})$. It is shown that particular extensions of the model, including the new

representation of rain cells, improve the fit of the model to data.

These papers highlight the potential benefits of ABC. At the expense of extra computation time, we are able to relax unrealistic assumptions without needing to derive, and code, expressions for model properties, and we have greater freedom to use summary statistics that provide information about the parameters. We can fit models that we would really like to fit using the properties that we would really like to use. It may be possible to use the data to select the best properties to use (Fearnhead & Prangle 2012). In common with GMOM fitting, an effective strategy for setting the starting value for $\boldsymbol{\theta}$ is required. Additionally, the value for the threshold ϵ and tuning parameters in the simulation algorithm need to selected to achieve a reasonable acceptance rate.

7. APPLICATIONS

We present an illustrative analysis of a subset of the rain gauge data described in Section 3.1 using the BLRPM temporal model. The aim is to describe the general approach, rather than to seek a best-fitting model. We also show an example of data simulated from a spatial-temporal model fitted to the radar data from Section 3.2. The purpose is to show that this type of model can produce realisations that looks like rain.

7.1. Temporal model

We fit the BLRPM to a subset of the data shown in **Figure 1**, specifically the 10-minute rainfall totals recorded in Bremen during the Januaries of 1992 to 2023. For simplicity, we suppose that rain cell intensities have an exponential distribution, with mean μ_X. We use the GMOM methodology described in Section 6.1, based on two sets of fitting properties.

- **Set 1**. The 1-hour mean, 1, 6 and 24-hour variance, 1 and 24-hour lag-1 autocorrelation and 1 and 24-hour proportion of wet periods.
- **Set 2**. The 1-hour mean, and the variance, lag-1 autocorrelation and coefficient of skewness at timescales of 10 minutes, 1 hour, 6 hours and 24 hours.

The former is the set of properties used by Jesus & Chandler (2011) and in the manual for the MOMFIT software (Chandler 2016). The latter is based on Cowpertwait et al. (2007) but variances are used instead of the coefficients of variation for ease of comparison with Fit 1. The key differences between these sets are the inclusion of 10-minute properties in Set 2 and the use of the coefficient of skewness in Set 2 instead of the proportion of wet periods used in Set 1. The inclusion of the coefficient of skewness, in conjunction with the use of a more flexible distribution for rain cell intensity such as a gamma or Weibull distribution, has been used to improve the reproduction of extreme values (Cowpertwait 1998).

Table 1 summarises the model fits from Set 1 and Set 2. The confidence intervals are calculated for the logs of the parameters, from the estimated asymptotic normal distribution of their estimators, before transforming back to the original scale. The estimated rates of storm arrival (λ) are similar, but the storm structures implied by the point estimates are rather different. Recall from Section 4.1 that the mean number of cells per storm is $\mu_C = 1 + \beta/\gamma$, the mean storm lifetime, the time for which it can produce new cells, is $1/\gamma$ and the mean cell duration is $1/\eta$. In comparison to Set 1, the fit from Set 2 produces storms that tend to be shorter (mean storm duration 4 hours vs 11 hours), contain more

Table 1 Estimates, standard errors and 95% confidence intervals for two fits of the BLRPM[a]

	Set 1				Set 2			
	estimate	SE	lower	upper	estimate	SE	lower	upper
λ	0.026	0.003	0.020	0.033	0.025	0.003	0.019	0.031
μ_X	0.799	0.080	0.658	0.972	1.295	0.170	1.001	1.676
β	0.393	0.134	0.201	0.768	4.300	0.590	3.286	5.626
γ	0.092	0.024	0.055	0.153	0.233	0.023	0.192	0.282
η	1.408	0.137	1.164	1.704	8.202	0.585	7.131	9.433

[a] Data: from 29 Januaries in 1995–2023 recorded in Bremen, Germany.

cells (a mean of 19 vs 5), produced at approximately ten times the rate (β) per hour and these cells are, on average, shorter (mean cell duration 7 minutes vs 42 minutes) and with a slightly higher mean intensity (1.3 mm vs 0.8 mm). Informal inspection of the estimates of uncertainty in **Table 1** suggests that these differences are real. This is a common finding. To closely reproduce statistical properties at shorter time scales using rectangular pulses of rainfall, storms need to contain more cells to represent the temporal profiles of rainfall well at those scales.

Properties used to fit the data will tend to be reproduced well, but it is instructive to see how well other properties are reproduced. **Figure 5** and **Figure 6** compare the fits of the BLRPM to the Bremen data using Sets 1 and 2, based on properties that are included in one or both of these sets or in neither. We have not included the hourly mean in this comparison because both models fit the observed value closely. As expected, properties that are used to fit the model tend to be reproduced better than those that are not. For example, statistics of 10-minute totals are reproduced better by Set 2 than by Set 1. However, extrapolation of the fit from Set 1 to 10-minute properties is reasonable, especially for the variance and proportion of wet periods, particularly when we take into account the variability in the observed values of the statistics over different years. In **Figure 5** there is no obvious systematic difference in the summary statistics from tipping bucket (pre 2009) and weight-based gauges nor a discernible trend in these statistics over time. This is also the case for the sample autocorrelations (not shown). In **Figure 6**, some of the sample autocorrelations for daily totals are negative and the sample mean of these is slightly negative for lag 3. The model cannot produce negative autocorrelations, but the fitted values at lags 2 and 3 are close to zero.

Figure 7 shows plots of 10-minute rainfall totals simulated from the BLRPM based on these two fits. The simulated series have been selected by trial and error to have similar amounts of rain to the lower plots in **Figure 1**. The appearance of these plots varies between different simulations, but we can see that in the top row, the longer-lived storms and cells produce a time series that is smoother than in the bottom row. A comparison with **Figure 1** should be based not on exactly when the rain falls but whether real and simulated series appear to have been produced by the same process. The simulated values do contain features that look like parts of the data but they are perhaps less noisy than the data.

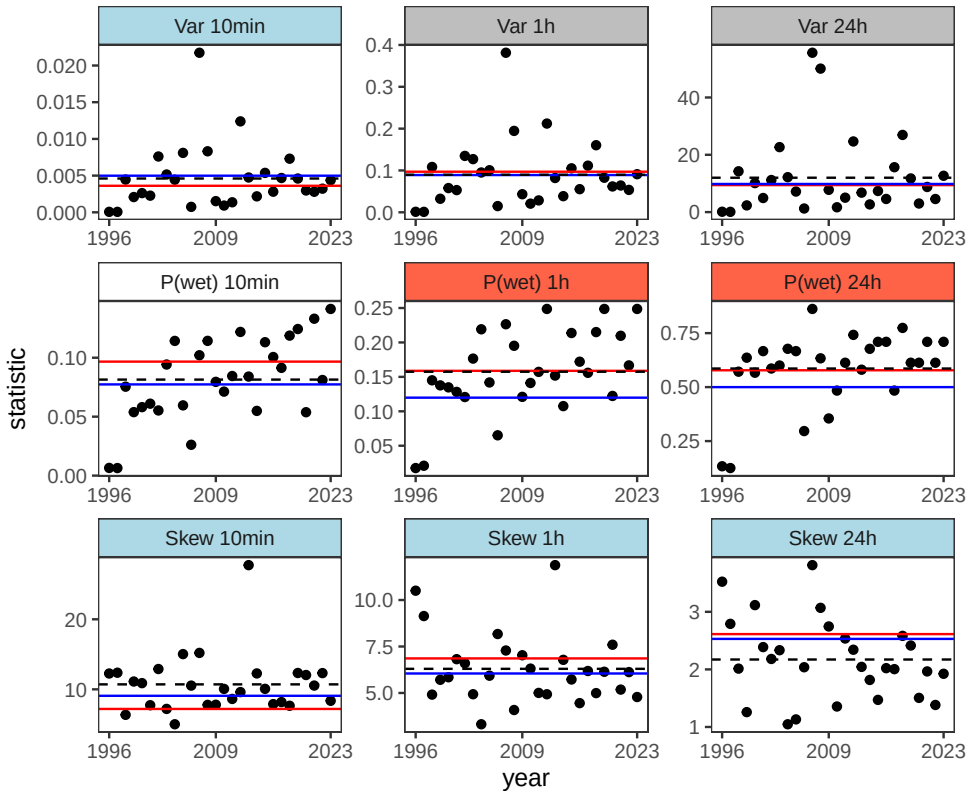

Figure 5

Comparison of observed and fitted properties based on the BLRPM fitted to the Bremen data using Sets 1 and 2. *Top row*: variance. *Middle row*: proportion of intervals that are wet. *Bottom row*: coefficient of skewness. *Left column*: 10-minute totals. *Middle column*: hourly totals. *Right column*: daily totals. The black circles are the observed statistic values for each year. The horizontal lines show the means of these values (black, dashed) and the fitted values using Set 1 (solid, red) and Set 2 (solid, blue). The colour of the header indicates whether the property is included in both Set 1 and Set 2 (grey), in Set 1 only (light red), in Set 2 only (light blue), or in neither set (white).

7.2. Spatial-temporal model

Figure 8 shows a simulation from the Northrop (1998) model fitted to the rain event depicted in **Figure 2**. The movement of the simulated rain event across the radar window is performed using an approach described in Chandler et al. (2007). The duration of the simulated rain event over the radar window has not been matched to that of the observed event, the former is longer than the latter, and the leading and trailing edges of the simulated event are rather more regular than those of the observed event. Otherwise, the features of the simulated data resemble those of the real data.

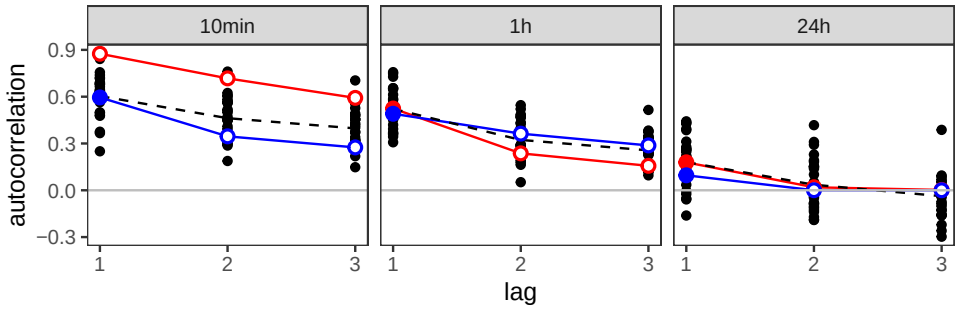

Figure 6

Comparison of the fitted autocorrelation functions at lags 1, 2 and 3 based on the BLRPM fitted to the Bremen data using Sets 1 and 2. *Left*: 10-minute totals. *Middle*: hourly totals. *Right*: daily totals. The black circles are the observed statistic values, one for each year in the data. The lines show the means of these values (dashed, black), the fitted values using Set 1 (solid, red) and Set 2 (solid, blue). For the fitted values, the plotting circle indicates whether the property is included (filled circle) in the set or not included (hollow circle).

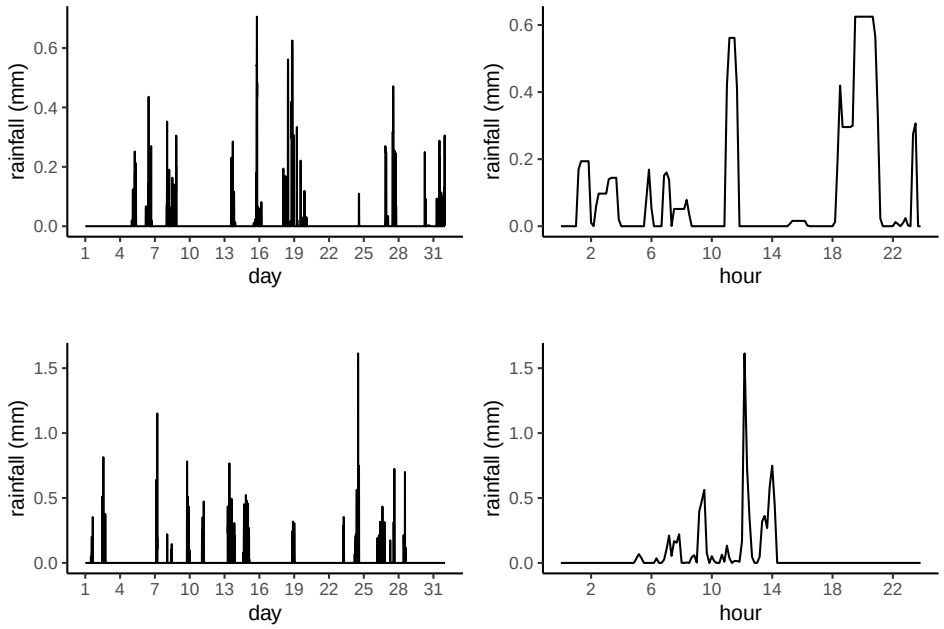

Figure 7

10-minute rainfall totals simulated from the BLRPM based on the fits to the Bremen data using Set 1 (*top row*) and Set 2 (*bottom row*). *Left column*: one month of data. *Right column*: one selected day of data (day 18 on the top row and day 24 on the bottom row).

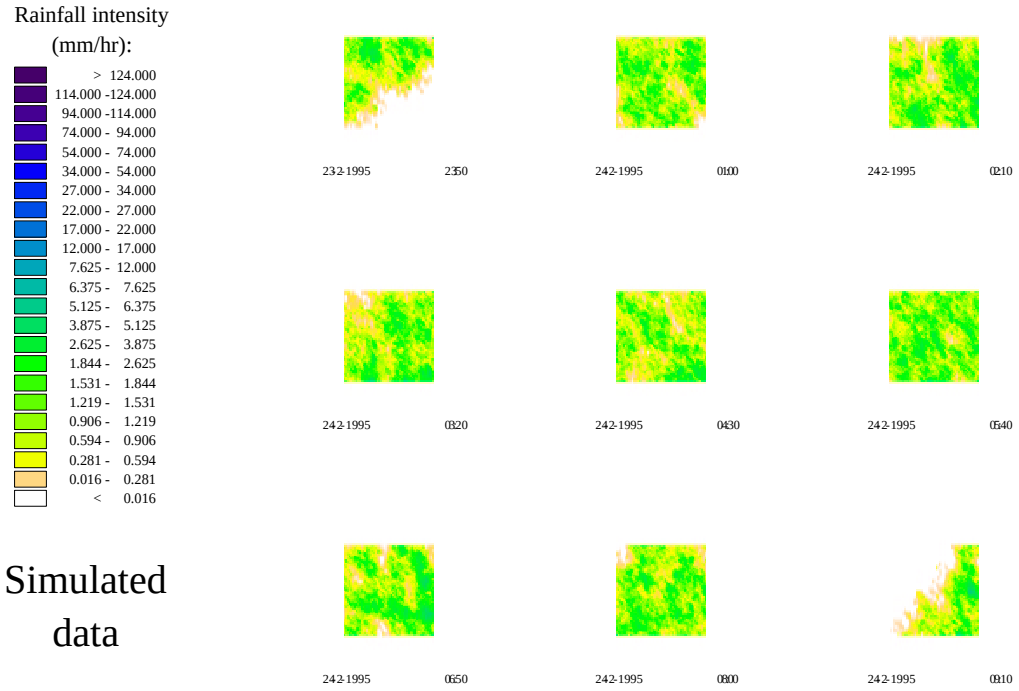

Figure 8

Simulated rainfall data. Each image is a 52 by 52 array of 2km by 2km pixels. The colour key indicates the average rainfall intensity for each pixel.

8. CONCLUDING REMARKS

We have described models that can simulate rainfall at any scale of interest. They are based on stationary processes and therefore some adjustment is required in circumstances where stationarity is unrealistic. A key issue is how to simulate from these models in a way that reflects anticipated effects of climate change on future rainfall. Studies of future climate impacts typically use projections from Global Climate Models (GCMs) based on scenarios of socio-economic development and greenhouse gas emissions (Collins et al. 2012). GCMs provide projections of precipitation for spatial and temporal scales that are far too coarse for hydrological applications. However, the temporal models described in Section 4.3, in which the effects of large-scale atmospheric covariates on model parameters are modelled using historical data, could utilise these projections. We could simulate hourly, say, rainfall data from such a fitted model conditional on projections of the relevant covariates from a GCM or multiple GCMs, ideally accounting for statistical uncertainty in the rainfall model's

parameters. In principle, a similar approach could be devised using a spatial-temporal model, but this would be a challenging problem.

An alternative is to condition simulation from these models on information about future non-stationarity in rainfall values. For example, suppose that we have available a series of daily rainfall totals that reflect anticipated future changes in rainfall. These could be produced by a GCM or simulated from a statistical model, for example, a Generalised Linear Model (Yang et al. 2005), fitted to GCM output. In one type of approach (Burton et al. 2010, Leith 2008, Kilsby et al. 2007, Marani & Zanetti 2007), empirical relationships are used to infer from daily summary statistics sets of sub-daily statistics for individual months. A stochastic-mechanistic model is fitted to these statistics, from which rainfall can be simulated. In a similar approach, Segond et al. (2006) first simulate hourly totals unconditionally from a stochastic-mechanistic model and adjust these totals, using the method of Koutsoyiannis & Onof (2001), to replicate the daily totals. These techniques are examples of statistical downscaling. Fowler et al. (2007) and Maraun et al. (2010) provide reviews that focus on rainfall. We describe two other methods that simulate rainfall data at fine scales and can be used to perform statistical downscaling.

Approaches based on a multiplicative random cascade process (Schertzer & Lovejoy 1987) repeatedly redistribute rainfall of coarse temporal and/or spatial scale to create values of progressively finer resolution. For example, the average intensity Y of a 2-hour time interval generates two 1-hour average intensities YW_1 and YW_2, where W_1, W_2 are random variables, often assumed to be independent and identically distributed, that define the cascade generator. Common choices of generator distribution are log-Poisson (Deidda et al. 1999) and log-Lévy (Schertzer & Lovejoy 1997). Moments of average rainfall intensity of positive order q, the q-moments, have a log-log linear relationship with scale whose gradient depends on q and parameters of the cascade model. Model calibration is based on the behaviour of sample q-moments aggregated over several scales, for a selected set of values of q (Pecknold et al. 2001, Gaume et al. 2007). The cascade structure enables simulation of intensities at scales that are finer than those observed in the data. See, for example, Gires et al. (2012), who consider both spatial and spatial-temporal downscaling and note that normalisation could be used to preserve at the extrapolated scales the same average intensity at the finest observed scale (rainfall disaggregation). The universal multifractal model (Lovejoy & Schertzer 1990, Schertzer & Lovejoy 2011), which can have as few as two parameters, has been used successfully in many hydrological applications. Other cascade models have been developed for modelling rainfall. For example, Olsson (1998) built dependence between the cascade generator variables to reflect better observed properties of rainfall and account explicitly for zero rainfall intensities.

An alternative approach, sometimes described as meta-Gaussian, models rainfall in space and time using a transformed Gaussian process (Lebel et al. 1998). The transformation must reflect the mass of zero rainfall values in the marginal distribution and produce an appropriate dependence structure. Building on developments by Guillot & Lebel (1999), Onibon et al. (2004) and Vischel et al. (2009), Wilcox et al. (2021) construct a 20-parameter meta-Gaussian model to simulate data that replicates the properties of convective storms in West Africa. Some parts of the model's structure are targeted at semi-arid regions, but adjustments could be made for implementation in other regions. Onibon et al. (2004) provides an algorithm for simulating conditionally on an average areal value, such as GCM output.

Finally, we note some relevant software. MOMFIT (Chandler 2016) fits single-site rain-

fall models to rain gauge data. RainSim (Burton et al. 2008, Fowler et al. 2005) simulates from a spatial–temporal NSRPM. LetItRain (Kim et al. 2017a,b) simulates from the random η BLRPM at sites across the USA, using kriging to interpolate parameter values to ungauged sites. STORAGE (De Luca & Petroselli 2000, 2021) performs fitting and simulation for the NSRPM. Rglimclim (Chandler 2020b,a), RMAWGEN (Cordano & Eccel 2017, 2016) and MulGETS (Chen et al. 2014) simulate daily multisite rainfall data. HYETOS (Koutsoyiannis & Onof 2000, 2001) disaggregates daily totals to form hourly totals by adjusting simulations from the BLRPM.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGEMENTS

The author is grateful to an anonymous reviewer for very helpful comments on a draft.

LITERATURE CITED

Aryal NR, Jones OD. 2020. Fitting the Bartlett–Lewis rainfall model using Approximate Bayesian Computation. *Mathematics and Computers in Simulation* 175:153–163

Aryal NR, Jones OD. 2021. Spatial-temporal rainfall models based on Poisson cluster processes. *Stochastic Environmental Research and Risk Assessment* 35(12):2629–2643

Beaumont MA. 2019. Approximate Bayesian computation. *Annual Review of Statistics and Its Application* 6(1):379–403

Bellone E, Hughes JP, Guttorp P. 2000. A hidden Markov model for downscalling synoptic atmospheric patterns to precipitation amounts. *Climate Research* 15(1):1–12

Burton A, Fowler HJ, Blenkinsop S, Kilsby CG. 2010. Downscaling transient climate change using a Neyman–Scott rectangular pulses stochastic rainfall model. *Journal of Hydrology* 381(1-2):18–32

Burton A, Fowler HJ, Kilsby CG, O'Connell PE. 2010. A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts. *Water Resources Research* 46(11):W11501

Burton A, Kilsby CG, Fowler HJ, Cowpertwait PS, O'Connell PE. 2008. RainSim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling and Software* 23(12):1356–1369

Camera C, Bruggeman A, Hadjinicolaou P, Michaelides S, Lang MA. 2017. Evaluation of a spatial rainfall generator for generating high resolution precipitation projections over orographically complex terrain. *Stoch Environ Res Risk Assess* 31:757–773

Cameron D, Beven K, Tawn J. 2000. Modelling extreme rainfalls using a modified random pulse Bartlett–Lewis stochastic rainfall model (with uncertainly). *Advances in Water Resources* 24(2):203–211

CEH Wallingford. 2007. HYREX project: Met Office C-band Radar data from Wardon Hill, UK. NCAS British Atmospheric Data Centre

Chandler RE. 2016. MOMFIT: Moment-based calibration of point process models for single-site rainfall sequences. `https://www.ucl.ac.uk/~ucakarc/work/momfit.html`

Chandler RE. 2020a. Multisite, multivariate weather generation based on generalised linear models. *Environmental Modelling & Software* 134:104867

Chandler RE. 2020b. Rglimclim: A multisite, multivariate weather generator based on generalised linear models. `https://www.ucl.ac.uk/~ucakarc/work/glimclim.html`

Chandler RE, Isham V, Bellone E, Yang C, Northrop P. 2007. Space-time modelling of rainfall for continuous simulation. In *Statistical Methods for Spatio-Temporal Systems*, eds. B Finkenstadt, L Held, V Isham, chap. 5. Chapman and Hall/CRC, 177–215

Chandler RE, Isham VS, Northrop PJ, Wheater HS, Onof CJ, Leith NA. 2014. Uncertainty in rainfall inputs. In *Applied Uncertainty Analysis for Flood Risk Management*, eds. K Beven, J Hall, chap. 7. Imperial College Press, 101–152

Chen J, Brissette F, Zhang XJ. 2014. A multi-site stochastic weather generator for daily precipitation and temperature. *Transactions of the ASABE* 57(5):1375–1391

Chen Y, Paschalis A, Wang LP, Onof C. 2021. Can we estimate flood frequency with point-process spatial-temporal rainfall models? *Journal of Hydrology* 600:126667

Collins M, Chandler RE, Cox PM, Huthnance JM, Rougier J, Stephenson DB. 2012. Quantifying future climate change. *Nature Climate Change* 2(6):403–409

Cordano E, Eccel E. 2016. Tools for stochastic weather series generation in R environment. *Italian Journal of Agrometeorology* 3:31–42

Cordano E, Eccel E. 2017. RMAWGEN: Multi-Site Auto-Regressive Weather GENerator. R package

Cowpertwait P, Isham V, Onof C. 2007. Point process models of rainfall: developments for fine-scale structure. *Proc. R. Soc. Lond. Ser. A* 463(2086):2569–2587

Cowpertwait PSP. 1994. A generalized point process model for rainfall. *Proc. R. Soc. Lond. Ser. A* 447(1929):23–37

Cowpertwait PSP. 1995. A generalized spatial-temporal model of rainfall based on a clustered point process. *Proc. R. Soc. Lond. Ser. A* 450(1938):163–175

Cowpertwait PSP. 1998. A Poisson-cluster model of rainfall: some high-order moments and extreme values. *Proc. R. Soc. Lond. Ser. A* 454(1971):885–898

Cowpertwait PSP. 2010. A spatial-temporal point process model with a continuous distribution of storm types. *Water Resources Research* 46(12)

Cowpertwait PSP, Kilsby CG, O'Connell PE. 2002. A space-time Neyman–Scott model of rainfall: Empirical analysis of extremes. *Water Resources Research* 38(8):6–1–6–14

Cox DR, Isham VS. 1988. A simple spatial-temporal model of rainfall. *Proc. R. Soc. Lond. Ser. A* 415(1849):317–328

Cox DR, Isham VS. 1994. Stochastic models of precipitation. In *Statistics for the Environment 2: Water Related Issues*, eds. V Barnett, KF Turkman, chap. 1. Wiley, 3–18

Cross D, Onof C, Winter H. 2020. Ensemble estimation of future rainfall extremes with temperature dependent censored simulation. *Advances in Water Resources* 136:103479

Cross D, Onof C, Winter H, Bernardara P. 2018. Censored rainfall modelling for estimation of fine-scale extremes. *Hydrology and Earth System Sciences* 22(1):727–756

Daud ZM, Rasid SMM, Abas N. 2016. A regionalized stochastic rainfall model for the generation of high resolution data in peninsular malaysia. *Modern Applied Science* 10(5):77–86

De Luca D, Petroselli A. 2000. STORAGE (STOchastic RAinfall GEnerator) : A user-friendly software for generating long and high-resolution rainfall time series. `https://sites.google.com/unical.it/storage`

De Luca DL, Petroselli A. 2021. STORAGE (STOchastic RAinfall GEnerator): A user-friendly software for generating long and high-resolution rainfall time series. *Hydrology* 8(2):76

Deidda R, Benzi R, Siccardi F. 1999. Multifractal modeling of anomalous scaling laws in rainfall. *Water Resources Research* 35(6):1853–1867

Entekhabi D, Rodriguez-Iturbe I, Eagleson PS. 1989. Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: parameter estimation and validation. *Water Resources Research* 25(2):295–302

Evin G, Favre AC. 2008. A new rainfall model based on the Neyman–Scott process using cubic copulas. *Water Resources Research* 44(3):W03433

Evin G, Favre AC. 2013. Further developments of a transient Poisson-cluster model for rainfall. *Stochastic Environmental Research and Risk Assessment* 27(4):831–847

Favre AC, Musy A, Morgenthaler S. 2002. Two-site modeling of rainfall based on the Neyman–Scott process. *Water Resources Research* 38(12):43–1–43–7

Fearnhead P, Prangle D. 2012. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *J. R. Stat. Soc. Ser. B* 74(3):419–474

Forsythe N, Fowler H, Blenkinsop S, Burton A, Kilsby C, et al. 2014. Application of a stochastic weather generator to assess climate change impacts in a semi-arid climate: The upper indus basin. *Journal of Hydrology* 517:1019–1034

Fowler HJ, Blenkinsop S, Tebaldi C. 2007. Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology* 27(12):1547–1578

Fowler HJ, Kilsby CG, O'Connell PE, Burton A. 2005. A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change. *Journal of Hydrology* 308(1):50–66

Gaume E, Mouhous N, Andrieu H. 2007. Rainfall stochastic disaggregation models: Calibration and validation of a multiplicative cascade model. *Advances in Water Resources* 30(5):1301–1319

Gires A, Onof C, Maksimovic C, Schertzer D, Tchiguirinskaia I, Simoes N. 2012. Quantifying the impact of small scale unmeasured rainfall variability on urban runoff through multifractal downscaling: A case study. *Journal of Hydrology* 442-443:117–128

Guillot G, Lebel T. 1999. Disaggregation of sahelian mesoscale convective system rain fields: Further developments and validation. *Journal of Geophysical Research: Atmospheres* 104(D24):31533–31551

Jesus J, Chandler RE. 2011. Estimating functions and the generalized method of moments. *Interface Focus* 1(6):871–885

Jesus J, Chandler RE. 2017. Inference with the Whittle likelihood: A tractable approach using estimating functions. *J. Time Ser. Anal.* 38(2):204–224

Jones PD, Kilsby CG, Harpham C, Glenis V, Burton A. 2009. UK climate projections science report: Projections of future daily climate for the UK from the weather generator. Technical report, University of Newcastle

Kaczmarska J, Isham V, Onof C. 2014. Point process models for fine-resolution rainfall. *Hydrological Sciences Journal* 59(11):1972–1991

Kaczmarska JM. 2013. Single-site point process-based rainfall models in a nonstationary climate. Ph.D. thesis. University College London

Kaczmarska JM, Isham VS, Northrop P. 2015. Local generalised method of moments: an application to point process-based rainfall models. *Environmetrics* 26(4):312–325

Kakou A. 1997. Point proces based models for rainfall. Ph.D. thesis. University College London

Kilsby C, Jones P, Burton A, Ford A, Fowler H, et al. 2007. A daily weather generator for use in climate change studies. *Environmental Modelling & Software* 22(12):1705–1719

Kim D, Cho H, Onof C, Choi M. 2017a. Let-It-Rain: a web application for stochastic point rainfall generation at ungaged basins and its applicability in runoff and flood modeling. `http://www.letitrain.info/`

Kim D, Cho H, Onof C, Choi M. 2017b. Let-It-Rain: a web application for stochastic point rainfall generation at ungaged basins and its applicability in runoff and flood modeling. *Stochastic Environmental Research and Risk Assessment* 31(4):1023–1043

Kim D, Olivera F. 2012. Relative importance of the different rainfall statistics in the calibration of stochastic rainfall generation models. *Journal of Hydrologic Engineering* 17(3):368–376

Kistler R, Kalnay E, Collins W, Saha S, White G, et al. 2001. The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society* 82(2):247–268

Koutsoyiannis D, Onof C. 2000. HYETOS - a computer program for stochastic disaggregation of fine-scale rainfall. `https://www.itia.ntua.gr/e/softinfo/3/`

Koutsoyiannis D, Onof C. 2001. Rainfall disaggregation using adjusting procedures on a Poisson

cluster model. *Journal of Hydrology* 246(1):109–122
Lebel T, Braud I, Creutin JD. 1998. A space-time rainfall disaggregation model adapted to sahelian mesoscale convective complexes. *Water Resources Research* 34(7):1711–1726
LeCam L. 1961. A stochastic description of precipitation. *Proc. 4th Berkeley Symp.* 3:165–186
Leith NA. 2008. Single-site rainfall generation under scenarios of climate change. Ph.D. thesis. University College London
Leonard M, Lambert MF, Metcalfe AV, Cowpertwait PSP. 2008. A space-time Neyman–Scott rainfall model with defined storm extent. *Water Resources Research* 44(9):W09402
Lovejoy S, Schertzer D. 1990. Multifractals, universality classes and satellite and radar measurements of cloud and rain fields. *Journal of Geophysical Research: Atmospheres* 95(D3):2021–2034
Marani M, Zanetti S. 2007. Downscaling rainfall temporal variability. *Water Resources Research* 43(9):W09415
Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, et al. 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* 48(3):RG3003
Muñoz Lopez C, Wang LP, Willems P. 2023. Statistical characterization of rainfall fields based upon a 12-year high-resolution radar archive of Belgium. *Atmospheric Research* 283:106544
Northrop P. 1998. A clustered spatial-temporal model of rainfall. *Proc. R. Soc. Lond. Ser. A* 454(1975):1875–1888
Northrop PJ. 2006. Estimating the parameters of rainfall models using maximum marginal likelihood. *Student* 5(3/4):173–183
Ochoa-Rodriguez S, Wang LP, Willems P, Onof C. 2019. A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resources Research* 55(8):6356–6391
Olsson J. 1998. Evaluation of a scaling cascade model for temporal rainfall disaggregation. *Hydrology and Earth System Sciences* 2(1):19–30
Onibon H, Lebel T, Afouda A, Guillot G. 2004. Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resources Research* 40(8):W08401
Onof C, Arnbjerg-Nielsen K. 2009. Quantification of anticipated future changes in high resolution design rainfall for urban areas. *Atmospheric Research* 92(3):350–363
Onof C, Chandler RE, Kakou A, Northrop P, Wheater HS, 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, Wheater HS, Isham V. 1994. Note on the analytical expression of the inter-event time characteristics for Bartlett–Lewis type rainfall models. *Journal of Hydrology* 157(1):197–210
Park J, Cross D, Onof C, Chen Y, Kim D. 2021. A simple scheme to adjust Poisson cluster rectangular pulse rainfall models for improved performance at sub-hourly timescales. *Journal of Hydrology* 598:126296
Pathiraja S, Westra S, Sharma A. 2012. Why continuous simulation? The role of antecedent moisture in design flood estimation. *Water Resources Research* 48(6):W06534
Pecknold S, Lovejoy S, Schertzer D. 2001. Stratified multifractal magnetization and surface geomagnetic fields—ii. multifractal analysis and simulations. *Geophysical Journal International* 145(1):127–144
Ramesh NI. 1998. Temporal modelling of short-term rainfall using Cox processes. *Environmetrics* 9(6):629–643
Ramesh NI, Garthwaite AP, Onof C. 2018. A doubly stochastic rainfall model with exponentially decaying pulses. *Stochastic Environmental Research and Risk Assessment* 32(6):1645–1664
Ramesh NI, Onof C, Xie D. 2012. Doubly stochastic Poisson process models for precipitation at fine time-scales. *Advances in Water Resources* 45:58–64
Ramesh NI, Thayakaran R, Onof C. 2013. Multi-site doubly stochastic Poisson process models for fine-scale rainfall. *Stochastic Environmental Research and Risk Assessment* 27(6):1383–1396

Rodriguez-Iturbe I, Cox DR, Isham V. 1987. Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. Ser. A* 410(1839):269–288

Rodriguez-Iturbe I, Gupta VK, Waymire E. 1984. Scale considerations in the modeling of temporal rainfall. *Water Resources Research* 20(11):1611–1619

Rodriquez-Iturbe I, Cox DR, Isham V. 1988. A point process model for rainfall: further developments. *Proc. R. Soc. Lond. Ser. A* 417(1853):283–298

Saha R, Testik FY, Testik MC. 2021. Assessment of OTT Pluvio2 rain intensity measurements. *Journal of Atmospheric and Oceanic Technology* 38(4):897–908

Schertzer D, Lovejoy S. 1987. Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research: Atmospheres* 92(D8):9693–9714

Schertzer D, Lovejoy S. 1997. Universal multifractals do exist!: Comments on "A statistical analysis of mesoscale rainfall as a random cascade". *Journal of Applied Meteorology* 36(9):1296–1303

Schertzer D, Lovejoy S. 2011. Multifractals, generalized scale invariance and complexity in geophysics. *International Journal of Bifurcation and Chaos* 21(12):3417–3456

Segond ML, Onof C, Wheater H. 2006. Spatial–temporal disaggregation of daily rainfall from a generalized linear model. *Journal of Hydrology* 331(3):674–689

Segond ML, Wheater HS, Onof C. 2007. The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *Journal of Hydrology* 347(1-2):116–131

Thayakaran R, Ramesh NI. 2013a. Markov modulated Poisson process models incorporating covariates for rainfall intensity. *Water Science and Technology* 67(8):1786–1792

Thayakaran R, Ramesh NI. 2013b. Multivariate models for rainfall based on Markov modulated Poisson processes. *Hydrology Research* 44(4):631–643

Verhoest NEC, Vandenberghe S, Cabus P, Onof C, Meca-Figueras T, Jameleddine S. 2010. Are stochastic point rainfall models able to preserve extreme flood statistics? *Hydrological Processes* 24(23):3439–3445

Vischel T, Lebel T, Massuel S, Cappelaere B. 2009. Conditional simulation schemes of rain fields and their application to rainfall–runoff modeling studies in the sahel. *Journal of Hydrology* 375(1):273–286Surface processes and water cycle in West Africa, studied from the AMMA-CATCH observing system

Vu T, Mishra A. 2020. Performance of multisite stochastic precipitation models for a tropical monsoon region. *Stoch Environ Res Risk Assess* 34:2159–2177

Wasko C, Pui A, Sharma A, Mehrotra R, Jeremiah E. 2015. Representing low-frequency variability in continuous rainfall simulations: A hierarchical random bartlett lewis continuous rainfall generation model. *Water Resources Research* 51(12):9995–10007

Wasko C, Sharma A. 2017. Continuous rainfall generation for a warmer climate using observed temperature sensitivities. *Journal of Hydrology* 544:575–590

Waymire E, Gupta VK. 1981. The mathematical structure of rainfall representations: 1. a review of the stochastic rainfall models. *Water Resources Research* 17(5):1261–1272

Waymire E, Gupta VK, Rodriguez-Iturbe I. 1984. A spectral theory of rainfall intensity at the meso-β scale. *Water Resources Research* 20(10):1453–1465

Wheater HS. 2002. Progress in and prospects for fluvial flood modelling. *Phil. Trans. R. Soc. A.* 360(1796):1409–1431

Wheater HS, Isham VS, Chandler RE, Onof CJ, Stewart EJ. 2006. Improved methods for national spatial-temporal rainfall and evaporation modelling for BSM. R&D Technical Report FD2105/TR, DEFRA

Wheater HS, Isham VS, Cox DR, Chandler RE, Kakou A, et al. 2000. Spatial-temporal rainfall fields: modelling and statistical aspects. *Hydrology and Earth System Sciences* 4(4):581–601

Wilcox C, Aly C, Vischel T, Panthou G, Blanchet J, et al. 2021. Stochastorm: A stochastic rainfall simulator for convective storms. *Journal of Hydrometeorology* 22(2):387–404

Wilks D. 1998. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology* 210(1):178–191

Willems P. 2001. A spatial rainfall generator for small spatial scales. *Journal of Hydrology* 252(1):126–144

Yang C, Chandler RE, Isham VS, Wheater HS. 2005. Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research* 41(11):W11415