A spatial-temporal rainfall generator for urban drainage design

Fiona H. McRobie, Li-Pen Wang, Christian Onof and Stephen Kenney

ABSTRACT

The work presented here is a contribution to the Thames Water project of improving the Counters Creek catchment sewerage system in London. An increase in the number of floods affecting basements in the area has indicated the need for improvements to the system. The cost of such improvements could be very high, and as such it is important to determine whether the traditional approach of applying 30-year spatially uniform design storms results in substantial overestimation. The first step in this is to generate simulations of spatially distributed rainfall events, from which 30-year storms can be extracted. Storms are modelled as clusters of Gaussian rainfall cells, extending the earlier Willems method to radar rainfall data. The parameters describing the cells and their motion are sampled from probability distributions derived from parameter estimates gained from 45 historical storm events within the catchment for the period 2000–2011. This spatial-temporal stochastic rainfall generator produces a two-dimensional time series of simulated storm events, from which events of given return period can be identified.

Key words | Gaussian, radar rainfall, rainfall cell, spatial-temporal, stochastic, urban hydrology

INTRODUCTION

The total volume of rainfall is generally recognised as the driving factor determining runoff; however, the spatial and temporal variations of storm events have been shown to provide additional insight into the catchment response (Singh 1997). In particular, the speed and direction of a storm event have been found to influence runoff peak and variability (de Lima & Singh 2002; Pechlivanidis et al. 2008).

Studies on spatial variations in the rainfall field and the subsequent runoff have produced catchment-specific results. Syed et al. (2003) found that, in a catchment in Arizona, the location of the storm ‘core’ is most significant for larger catchments, as storms located furthest from the catchment outlet generate more attenuated flows. In contrast, Pechlivanidis et al. (2008) found no change in the importance of the spatial properties of rainfall with varying catchment scales in the Upper Lee catchment in the UK. Considering the same catchment, Segond et al. (2007) highlighted that urban areas were more sensitive to spatial variations in rainfall due to greater runoff coefficients. This provides the impetus to incorporate spatially varying precipitation into network analysis, requiring a program able to simulate spatially distributed rainfall calibrated to the specific catchment.

Willems (2001) produced a spatial rainfall generator for a 100 km² urban catchment covering Antwerp, Belgium. Calibrated by a network of 12 rain gauges, ‘small mesoscale area’ storms of size $10^2$ to $10^3$ km² were modelled as a cluster of rainfall cells, each of a bivariate Gaussian form. Nimrod radar-generated precipitation fields have shown greater skill in modelling flow measurements in an urban catchment in northern England than rain gauge data (Harison et al. 2009). This work, therefore, based upon the conceptual model of Willems (2001), proposes a new fine resolution spatial-temporal rainfall generator, whereby the parameters that characterise rainfall storms are derived from historical radar-generated rainfall fields.

Catchment and data set

The region of interest is an 11 by 12 km² area covering the Counters Creek catchment in north-west London. Counters Creek is a sub-catchment of the Beckton catchment, and spatially distributed rainfall fields encompassing this wider catchment are required in order to allow for backflow. Precipitation fields generated from Nimrod radar data (Golding
(1998) are available at 1 x 1 km² and 5-minute resolution from 2000. Eighty storm events from the period 2000 to 2011 were highlighted by MWH Global (MWH) as significant with respect to network flooding, and, of these, the radar data of only 57 storms were considered sufficiently high quality for use (MWH 2011). The model is calibrated using rainfall fields from a larger area fully encompassing the Beckton catchment, to ensure that the full shape of storms is captured. This region covers 8,400 km² for post-November 2006 data and 5,525 km² prior to November 2006. A complete rainfall field is available for only 45 of the 57 selected storm events. T-tests verified these to be a fairly representative sample of the 57 selected storms.

**METHODOLOGY**

**Model definition**

Under the assumption of constant velocity, as per Willems (2001), each simulated storm event is generated by translation of the ‘simulation area’ across the catchment area as depicted in Figure 1.

The dimensions of the simulation area are defined by the dimensions of the catchment area (wa and la) as well as the duration, d, and speed, v, of the storm:

\[
ws = wa + 4 E(s_2)
\]

\[
l_s = vd - la
\]

The equation for ws includes an adjustment to include cells whose centres fall outside the catchment area, but which still contribute to rainfall within the catchment.

Within the simulation area, rainfall cells are generated at a spatial resolution of 100 x 100 m², reducing error when translating to the catchment area of 1 x 1 km² resolution. Based on the assumption that rainfall cells are uniformly distributed across a mesoscale rainfall event (Willems 2001), the number n of rainfall cells within the simulation area is found by sampling from the Poisson distribution \[P(\lambda ws)\]. These n cell centres are located across the...
simulation area using the uniform property of Poisson variables. Sampling from U([0,λ]) and U([0,ω]) yields the coordinates x_c and y_c of each cell centre.

Following Willems (2001), the cells are modelled as ellipses with one axis in the direction of movement and the other perpendicular. Rainfall intensity is represented by a Gaussian function along each axis, centred on (x_c, y_c). This differs from the approach of Einfalt et al. (1990) where the main axes of inertia are used, but fits with empirical observation that cell movement is orthogonal to any cell banding (Wheater et al. 2000). Cells are constructed by sampling for the peak cell intensities r_{max}, attributed to the point (x_c, y_c), and cell spread parameters s_1 and s_2, which describe the rate of decay of the rainfall intensity in the longitudinal and transverse planes respectively.

Finally, the whole simulation area moves across the catchment area according to the sampled speed v and angle θ. At each time step t, the rainfall intensity in the catchment area is obtained by a mapping onto the moving simulated rainfall field based on geometric relationships. This mapping relates each 1 × 1 km² grid square within the catchment area to a group of 100 × 100 m² grid points in the simulation field, which are then averaged to yield the rainfall intensity in the catchment grid square.

Model calibration

Cell identification

The cell identification method of Willems (2001) uses paired rain gauges and the relative changes in precipitation measured at each in order to identify cells and estimate shape and velocity parameters. This method is not directly applicable to the precipitation fields derived from radar data, so instead the hierarchical threshold segmentation (HTS) method of Peak & Tag (1994) was implemented. HTS was developed to identify cloud features in satellite images, extending the basic approach of partitioning a data field by a single threshold level of image brightness, and instead takes a sequence of threshold levels and retrospectively selects the most appropriate level for different regions of the image (Peak & Tag 1994). Applied here, for a number of threshold precipitation intensities, τ_i rainfall cells are defined as a connected group of pixels within the precipitation field with values exceeding the threshold τ_i if, for all thresholds τ_i with τ_i < τ_f only a single, smaller, rainfall cell is identified. This algorithm was applied to the precipitation field at each time step, outputting a set of cells at each 5-minute interval.

Parameter estimation

The distributions from which model parameters are sampled are fitted to estimates derived from the 45 observed storm events.

The estimator of the number of cells per km², λ, is the number of cells identified at each time step. Duration, d, is estimated by taking the time when the first cells are identified until the last cells are identified.

The speed of the storm event, v, and, the direction of movement, θ, are estimated using spatial correlations relative to a central pixel in the precipitation field over a series of time-step lags. This allows for an estimate of the distance travelled during the lag to be determined, and thus velocity (magnitude and direction) can be estimated for each storm event.

<table>
<thead>
<tr>
<th>Parameter set</th>
<th>Fitted distribution</th>
<th>Distribution parameters</th>
<th>Standard error (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storm duration (d)</td>
<td>Gamma</td>
<td>a (-) = 2.44, b (hrs) = 17.92</td>
<td>SE_a = 0.529, SE_b = 4.234</td>
</tr>
<tr>
<td>Number of cells per square kilometre (λ)</td>
<td>Normal</td>
<td>μ (km⁻²) = 0.0099, σ (km⁻²) = 0.0052</td>
<td>SE_μ = 0.000519, SE_σ = 0.000374</td>
</tr>
<tr>
<td>Velocity magnitude (v)</td>
<td>Log normal</td>
<td>μ (km/hr) = 78.30, σ (km/hr) = 23.64</td>
<td>SE_μ = 0.0555, SE_σ = 0.0387</td>
</tr>
<tr>
<td>Velocity direction (θ)</td>
<td>Log normal</td>
<td>μ (rads) = 0.508, σ (rads) = 1.164</td>
<td>SE_μ = 0.150, SE_σ = 0.108</td>
</tr>
<tr>
<td>Cell spread in direction of motion (s₁)</td>
<td>Log normal</td>
<td>μ = 0.700, σ = 1.060 (mean = 3.52 km; var = 25.63 km²)</td>
<td>SE_μ = 0.00324, SE_σ = 0.00229</td>
</tr>
<tr>
<td>Cell spread in direction perpendicular to motion (s₂)</td>
<td>Log normal</td>
<td>μ = 0.704, σ = 1.001 (mean = 3.34 km; var = 19.17 km²)</td>
<td>SE_μ = 0.00297, SE_σ = 0.0021</td>
</tr>
<tr>
<td>Maximum intensity in cell (r_{max})</td>
<td>Generalised Pareto</td>
<td>λ = 0.586, σ = 1.040, θ = 0.2 mm (mean = 2.71 mm)</td>
<td>SE_λ = 0.00233, SE_σ = 0.00272, SE_θ = 0</td>
</tr>
</tbody>
</table>
Figure 2 | Histogram of estimated parameters and fitted probability distribution: (a) storm duration, d, and gamma distribution, (b) number of storm cells per km², λ, and normal distribution, (c) scaled velocity parameter, 150-v, and log-normal distribution, (d) scaled direction parameter, π/2 – θ, and log-normal distribution, (e) longitudinal cell spread, s₁, and log-normal distribution, (f) transverse cell spread, s₂, and log-normal distribution, and (g) maximum intensity, r_max, and Generalised Pareto distribution.
The cell parameters, \( s_1 \) and \( s_2 \), and \( r_{\text{max}} \) are estimated for all cells in each time step, as no suitable method for tracking the rainfall cells over time was found. Estimating the cell spreads relies on the assumption that in the plane of the direction of motion (given by \( \theta \)), and the transverse, rainfall cells are described by a Gaussian function:

\[
f(u_i) = \frac{A}{\sqrt{2\pi s_i^2}} \exp\left(-\frac{u_i^2}{2s_i^2}\right)
\]

where \( u_i \) is distance in the plane of the direction of motion or transverse, \( s_i \) is cell spread, and \( A \) is a proportionality constant. This allows \( s_1 \) and \( s_2 \) to be independently solved for using the rainfall intensity at the point where the cell is cut off due to the HTS method and the peak intensity.

The assumption that rainfall cells remain a constant shape and size over time is not true, and the estimated values of \( s_1 \) and \( s_2 \) will include all the fluctuating sizes of the cells. This results in a large sample of spread estimates containing not only a useful set of spread estimates but also some unrealistically small numbers that would have been averaged out over the life-time of the cell. This is corrected for by transforming the estimates of the spread variables such that they are distributed according to the average of the spread variables.

**Distribution fitting**

Table 1 gives the distributions fitted to each set of parameter estimates. The distributions were fitted by maximum likelihood estimation, and the fit of these distributions is displayed in Figure 2. Although the velocity magnitude parameter, \( v \), is fitted to a lognormal distribution, this distribution is later discarded in favour of the Weibull distribution as per Willems (2001). This is covered in detail in the model validation section.

**Model validation**

Model validation focuses on two central aspects, assessing cell identification and parameter estimation, and analysing the statistical structure of individual storm events relative to those observed.

To review the cell identification and parameter estimation process, the HTS algorithm is applied to a number of simulated storms; parameter estimates of \( v, \theta, s_1, s_2 \) and \( r_{\text{max}} \) are derived and are then compared to the known storm and cell parameters used to generate the synthetic storm. Figure 3 highlights a key problem with the velocity estimation process: for a simulated storm with all cells travelling at constant velocity, the range of estimates of speed and direction is broad. The estimates of cell spread are seen to be more reliable, as in the example in Figure 4, although a slight skew towards smaller cell sizes is noticeable.

In order to gauge how well the model replicates reality, the statistical structure of a number of simulated storms is compared with historically observed storms. Evaluation of the autocorrelation function for the first 10 5-minute lags shows that the stochastically generated storm events exhibit low autocorrelation, while historically observed storms display persistent autocorrelation to much larger lags, as seen in Figure 5. The persistence of the rainfall field is likely to have a significant impact on the urban catchment; however, this bias in the stochastic rainfall generator can be remedied by reducing the storm velocity. In particular, higher autocorrelation functions can be produced if the velocity parameter, \( v \), is sampled from the Weibull distribution as proposed by Willems (2001), allowing for more frequent sampling of smaller values of \( v \). The lognormal distribution fitted to the velocity parameter, \( v \), is therefore discarded in favour of the Weibull distribution (Willems 2001).

To evaluate the modelling of rainfall cells, intra-storm profiles of maximum intensity, mean intensity and spatial variance are produced. A sample of observed storms are...
selected and their velocity ($v$ and $\theta$), duration ($d$) and cell density ($\lambda$) parameter estimates are used to simulate storms which are identical apart from the location and size of cells.

Plots of the time evolution of the mean, maximum and spread of rainfall intensity over the catchment for observed (Figure 6) and simulated (Figure 7) storms indicate that the rainfall generated from the model is typically too small in...
intensity. Only rarely is a large cell intensity sampled (as expected from the Generalised Pareto distribution), but this does not lead to a storm of great volume, as the rest of the cells in the storm are likely to be small. Thus far, rainfall intensities of cells within a given storm are treated as independent from one another; however, this is an inaccurate representation. Analysis of the upper quartiles of maximum cell intensities of observed storms shows great deviation in the larger cell sizes between storms, as depicted in Table 2.

The maximum rainfall intensity of cells may be dependent on that of other cells in the same storm, creating a relationship between cell intensity and the characteristics of the storm event. This is a similar concept as that employed by Onof et al. (1996) to determine a relationship between the cell intensity distribution parameter and storm duration. As a result, the simulation program is adapted to allow for a level of ‘correlation’ between maximum cell intensities within a storm so the model is more likely to generate storms with several cells of substantial intensity. To allow for variation in the level of dependence, a percentage level $K\%$ is set such that $K\%/100$ cell peaks are sampled as independent, identically distributed variables from the Generalised Pareto function, while $(K/100 - 1)$ cell peaks are sampled from a truncated distribution based around the maximum of the independent $K\%/100$ cell peaks.

**DISCUSSION AND RECOMMENDATIONS FOR FURTHER WORK**

The aim of this study is to provide a proof of concept of the possibility of generating a wide variety of spatially
distributed storms able to emulate the observed variety of spatial storm structures. Certain issues arose in the identification of the model parameters, which are important to bear in mind, and could form the basis of further analysis.

It was hoped that velocity parameters \( v \) and \( \theta \) would be estimated directly through the use of a cell tracking algorithm, linking cells from one time step to the next. However, no such algorithm was determined, largely due to the noisy intensities of the radar data. As a result, the velocity parameters are estimated using spatial correlations over time-lagged radar-generated precipitation fields, later seen to be unreliable. Although we sought to remedy this, as described in the model validation section, this was based on the assumption that we could apply the

![Figure 7](image_url) Statistical properties of an example simulated storm generated from velocity, duration and cell density parameter estimates as storm in Figure 5: (a) mean intensity over time, (b) spatial variance over time, (c) maximum intensity over time.

<table>
<thead>
<tr>
<th>Quartile</th>
<th>0.80</th>
<th>0.85</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
<th>0.99</th>
<th>0.995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storm 37</td>
<td>2.9000</td>
<td>3.9000</td>
<td>5.7000</td>
<td>9.7000</td>
<td>15.2775</td>
<td>24.8000</td>
<td>34.1255</td>
</tr>
<tr>
<td>Storm 38</td>
<td>3.9000</td>
<td>5.8000</td>
<td>9.0000</td>
<td>17.4000</td>
<td>29.5000</td>
<td>53.1880</td>
<td>71.6735</td>
</tr>
<tr>
<td>Storm 39</td>
<td>2.0000</td>
<td>2.9000</td>
<td>4.6000</td>
<td>9.0000</td>
<td>16.3000</td>
<td>29.8000</td>
<td>42.8000</td>
</tr>
<tr>
<td>Storm 40</td>
<td>2.1000</td>
<td>2.8000</td>
<td>3.9000</td>
<td>6.3000</td>
<td>9.8000</td>
<td>17.8000</td>
<td>24.1025</td>
</tr>
</tbody>
</table>
distribution of velocities found for the Brussels area in Belgium (Willems 2001). It is worth noting here that the cell spread estimates are based upon the estimates of \( \theta \); however, as the resulting spread parameter distribution reflected actual values effectively, these are retained.

The distribution fitting for the spread parameter estimates, \( s_1 \) and \( s_2 \), assumes they are both distributed identically from one storm to the next. Although the cumulative distribution functions for each storm are visually similar, a Kolmogorov–Smirnov test is inconclusive, and for more than 95% of storm pairs, the hypothesis that they are drawn from the same distribution cannot be accepted with 90% confidence. This casts significant doubt over the fitted parameter distributions, which may be due to the lack of a satisfactory cell tracking algorithm or may be an indicator that the model is oversimplified and unable to accurately replicate storm heterogeneity.

Many of the issues above are related to the uncertainty and noise in radar rainfall measurements. It has been estimated that the Beckton catchment would require several hundred rain gauges to provide the same level of insight into spatial variability as can be captured in radar-generated precipitation estimates (Neale 2008). There are, however, well known limitations to the use of radar-generated precipitation estimates, detailed in Einfall et al. (1990). Comparison of Nimrod radar and gauge rainfall measurements in a northern UK catchment indicated no systematic bias between the two, rather the differences varied substantially over time (Schellart et al. 2012). In the London region, comparison with gauge measurements highlighted overprediction of peak intensities and total volume during large storm events in some areas, while other parts of the catchment indicated that the radar measurements underpredicted (Neale 2008). In this study, no attempt is made to improve the precipitation estimates beyond the removal of storm events with clearly erroneous values. Further developments could include the use of gauge measurements to adjust the radar-generated precipitation estimates, as in Overeem et al. (2009).

The maximum cell intensity is assumed to be independent of the cell spread in either direction. Although there is no evidence of any significant correlation, this may be due to the problem with the cell tracking, and should be considered in any future work.

The cell arrival process is Poisson, assuming mutual independence between cells. This is considered a reasonable approximation in the case of a small catchment area, but were the model to be applied to a larger area, clustering would have to be included. This could be achieved by embedding cells within clusters, represented as arriving according to a Poisson process, while the cells are displaced from the cluster centre using a Gaussian distribution (see Wheater et al. 2000). Following an analysis of radar data for 24 rainfall events in Treillières, France, Emmanuel et al. (2012) identified four different groups of rainfall events, characterised by differing spatial and temporal ranges. Such variation is not accounted for in this model, and the distributions used to sample the storm and cell parameters are assumed to be valid for all rainfall events. Future work incorporating detailed analysis into the nature of rainfall events across the catchment is recommended to assess the suitability of a single set of probability distributions.

The model assumes the underlying processes are stationary although the evidence suggests this is an oversimplification. Rainfall cells are modelled with a constant size throughout the storm, although visual inspection of the rainfall fields showed cells to grow, decay, split and merge. This could be partially accounted for by including a growth/decay constant as in Willems (2001), but estimating this constant would require cell tracking. Similarly, velocities are seen to shift over time, with storms changing both their speed and direction as they evolve. Incorporating this into the model similarly requires the cell tracking algorithm to generate a time series of velocity estimates.

## CONCLUSIONS

A stochastic generator of spatially varied rainfall for the Counters Creek catchment in the London region has been developed. This model is calibrated by radar rainfall data for 45 hydrologically significant storm events from the past 11 years. This study was based on the work undertaken by Willems (2001) and estimation of characteristic parameters was adapted to allow for the use of radar data.

## REFERENCES


MWH 2011 Counters Creek spatial rainfall study: Stage (a) How important is spatial rainfall variation? MWH Technical Note. TN/EWI/CH11/186/01.00, MWH Global, Warrington, UK.


First received 21 November 2012; accepted in revised form 4 March 2013