Implementing a space-time rainfall model for the Sydney region

M. Leonard*, A. Metcalfe**, M. Lambert* and G. Kuczera***

*School of Civil and Environmental Engineering, University of Adelaide, Adelaide, Australia
**School of Mathematical Sciences, University of Adelaide, Adelaide, Australia
***School of Civil, Surveying & Environmental Engineering, University of Newcastle, Australia

Abstract This paper investigates a Spatial Neyman–Scott Rectangular Pulse (SNSRP) model, which is one of only a few models capable of continuous simulation of rainfall in both space and time. The SNSRP is a spatial extension of the Neyman–Scott Rectangular Pulse model at a single point. The model is highly idealized having six parameters: storm arrival, cell arrival, cell radius, cell lifetime and two cell intensity parameters. A spatial interpolation of the scale parameter is used so that the model can be simulated continuously in space, rather than as a multi-site model. The parameters are calibrated using least-squares fits to statistical moments based on data aggregated to hourly and daily totals. The SNSRP model is calibrated to a very large network of 85 gauges over metropolitan Sydney and shows a good agreement to calibrated statistics. A simulation of 50 replicates over the region compares favourably to several observed temporal statistics, with an example given for one site. A qualitative discussion of the simulated spatial images demonstrates the underlying structure of non-advecting cylindrical cells.

Keywords Continuous simulation; Neyman–Scott; space-time rainfall

Introduction

Rainfall is the driving process of water movement on a catchment and is an important input for hydrological and environmental studies and risk assessments. Rainfall is highly variable in space and time and can contribute significant uncertainty and bias in runoff studies if the spatio-temporal structure is not adequately modelled (Shah et al., 1996b; Chaubey et al., 1999; Willems, 2001b). Further to this, small catchments (and urban catchments in particular) demonstrate quick response times to rainfall input; therefore there is the need to model rainfall accurately at a high resolution in both space and time.

In recent decades, stochastic rainfall modelling has focused on the development of point rainfall models. Point rainfall models are calibrated to rainfall from a single site, typically daily or sub-daily rain-gauges. One popular class of point rainfall models are known as Poisson-cluster models, including the Neyman–Scott and Bartlett–Lewis Rectangular Pulse models. These models express rainfall as a process of independent storm arrivals, each generating overlapping rain-cells of varied duration and intensity. Poisson cluster models have been calibrated at a variety of scales, including sub-daily scales, and also across a variety of regions (Onof et al., 2000). Recently, a Neyman–Scott model has been demonstrated to adequately simulate the rainfall process at a selection of major Australian cities (Frost et al., 2004).

Another class of models which have been demonstrated to satisfactorily simulate at the sub-daily scale are event-based models. Event based models characterize rainfall in time as an alternating process of dry periods (inter-storm duration) and wet ‘events’ (storm-duration and intensity). The DRIP model of Heneker et al. (2001) is one such model that is capable of simulating sub-daily rainfall across the majority of Australian regions. Lastly, daily rainfall models can be a viable alternative for simulating sub-daily rainfall.
rainfall when they are used in conjunction with a disaggregation process. There is a broad range of model formulations for daily rainfall, with Markov models being a popular approach (Srikanthan and McMahon, 2001). Numerous disaggregation methods have been developed for this purpose (Gutierrez-Magness and McCuen, 2004). However, no matter the approach, point rainfall models lack the ability to model the spatial variability of the rainfall process and require the unsatisfactory assumption that rain is constant over the area of a catchment.

A common extension to point rainfall models is the development of multi-site rainfall models such that the spatial correlation of rainfall at a series of points is conserved, in addition to the temporal process at each site. As one example, Onof et al. (2000) demonstrate a multi-site formulation of a Poisson cluster model. However, whilst multi-site models improve the spatial variability of rainfall in contrast to point models, they do not model rainfall as a continuous variable over the entire space.

The advent of radar measurements of rainfall in the 1960s has provided insight into the spatio-temporal evolution and highly structured nature of rainfall (Crane, 1990). As a consequence, radar has provided an avenue for developing space-time rainfall models. Recent examples include: the Modified Turning Bands model (Mellor, 1996), String of Beads model (Pegram and Clothier, 2001), a spatial Bartlett–Lewis model (Northrop, 1998) and a cascade model (Seed et al., 1999). Whilst some of the models have a facility for continuous simulation, the main research focus has been on reproducing the spatial complexity observed in a series of images (say for one storm). A result is that these models do not yet match the maturity of point rainfall models for use in continuous simulation over a region and over large periods of time (say 100 years). In addition to this, radar–rainfall measurements are inferential by nature; thus, even after error correction algorithms are applied, there is a large degree of uncertainty and bias in the measured values.

There are also a small number of models that aim to simulate a rainfall field and are continuous in time, but solely calibrated via rain-gauges (Shah et al., 1996a; Willems, 2001a; Cowpertwait et al., 2002). Of these models, the spatial Neyman–Scott model of Cowpertwait et al. (2002) has been chosen for further investigation in this study because of its demonstrated ability to preserve observed temporal statistics and because of the robust analytical framework of Poisson cluster models. This model is discussed further in the following section. Subsequent sections will outline the Sydney case-study, present results and discuss findings.

Methods

The Spatial Neyman–Scott Rectangular Pulse (SNSRP) model was introduced by Cowpertwait (1995); however, the formulation and notation adopted in this study follows Cowpertwait et al. (2002). Storms are assumed to arrive according to a Poisson process with a rate of $\lambda$ storms per hour. Each storm yields a set of rain-cells that are exponentially displaced in time relative to the start of the storm with a mean displacement of $\beta^{-1}$ hours. The storm cells occur over the region as a spatial Poisson process with a rate of $\phi$ per km$^2$, which gives the mean number of cells per storm as the parameter, $\mu_C$. This formulation yields cells that are clustered in time, but not in space.

Each cell is represented as a circular disc having a specified radius, duration and intensity. Radii and cell durations are both assumed to be exponentially distributed with mean values of $\phi^{-1}$ km and $\eta^{-1}$ hours respectively, whilst cell intensity is assumed to follow a Weibull distribution, $P(X > x) = \exp(-x^{1/\alpha}/\theta)$, which has two parameters $\alpha$ and $\theta$. This approach requires six parameters ($\lambda, \alpha, \mu_C, \beta, \eta, \phi$) to simulate rainfall over a homogeneous region, with a different $\theta$ calculated for each observation site to model the between site variability. Seasonality of the data is taken into account by fitting a different...
set of parameters for each month; thus for $M$ sites, there is a total of $12 \times 6$ parameters $+ 12 \times M$ scaling parameters.

The above model assumptions enable analytical expressions to be derived for various properties of the model in terms of the model parameters. The properties used for calibration in this study were the coefficient of variation ($v_1$, hourly and $v_{24}$, daily), lag one autocorrelation ($p_1$, hourly and $p_{24}$, daily), hourly skewness ($\kappa_1$) and hourly lag zero cross correlations between each pair of sites. Since the model assumes the rainfall process to be stationary in space, the model properties are calculated for a non-dimensional case. This study scaled the data such that the hourly mean at all given sites was 1 mm. This choice was arbitrary as any other aggregate or value could have been selected without affecting model performance. The same non-dimensional statistics are calculated from the observed data for each site and fitting proceeds via a least-squares fit of the analytical properties to the observed statistics. Lastly, the parameter $\theta$ is calculated such that it scales from the non-dimensional case of 1 mm to the correct observed hourly mean rainfall at a given site.

The full equations for the calculation of analytical and observed statistics are given in Cowpertwait et al. (2002); note that equation six is typed incorrectly missing a factor of two in the subtraction term. The adopted fitting procedure for this study closely follows that of Cowpertwait et al. (2002) with the following differences: (i) a global optimization routine was used, (ii) all of the parameters were fitted simultaneously rather than in a series of separate steps, (iii) the least-squares contributions were not non-dimensionalised to avoid numerical divide-by-zero issues, (iv) the least-squares contribution from cross-correlations was weighted with respect to the number of pairs of sites. Weighting the cross-correlogram is important, since for any given month there are only five statistics averaged over the entire region ($v_1$, $\kappa_1$, $p_1$, $v_{24}$, $p_{24}$) whereas there are $C_N^2$ cross-correlation statistics between each pair of $N$ sites, i.e. fitting the cross-correlogram will dominate over fitting the regional statistics if weighting is not taken into account.

Once calibrated, the model can be simulated by successively sampling from each distribution for the arrival of storms, arrival of cells and cell properties over a given region of interest. The simulation proceeds such that the mean rainfall depth at any given point in the region is the same. To obtain the corrected time-series the scaling factor $\theta$ at a given site is employed. Since the scaling factor is calculated only at the sites of observed data, this restricts the model to being simulated in a multi-site capacity. To simulate an entire field of rainfall values, it is necessary to have a scaling factor at each point in the field. For this purpose, the set of scaling factors are kriged over the simulation region to yield a scaling surface across the region for each month of the year. An exponentially decaying spatial correlation was assumed for the kriged field of mean rainfall.

Case-study

The SNSRP was calibrated to a network of 85 pluviograph gauges across metropolitan Sydney. The gauges range from North Sydney to Wollongong (North–South extent), and from Bondi Beach to Katoomba (East–West extent). The region is approximately 100 $\times$ 100 km$^2$, with the majority of gauges distributed along the coastline, as shown in Figure 1.

The average observation length of the gauges is 15.5 years, with durations ranging from 6 years to 41 years. The elevation across the region is undulating from the coastline to approximately 150.6° longitude, then rising to 1,000 m above sea level due to the Blue Mountains. Based on the observed data, the mean annual rainfall over the region is 914 mm. The monthly rainfall in the region is slightly higher during summer months (90 mm) and lower during winter months (60 mm), and also varies with respect to
elevation, as shown in Figure 2. Figure 2(a) shows that January rainfall (summer) is higher over the Blue Mountains but July rainfall (winter) is higher along the coastline.

The observed and fitted statistics used in the calibration of the SNSRP are presented in Table 1. Note that the statistics are non-dimensional, using equations specified by Cowpertwait et al. (2002). The model shows a very good agreement to all of the statistics, with the exception of the high hourly auto-correlation ($r_1$) for August. Figure 3 shows the observed and fitted cross-correlations for two different months (a) January and (b) July. In addition to the other calibration statistics, the model provides a good fit to the cross-correlation between sites.

It can be noted from Figure 3 that the model does not fit the tail of the cross-correlation well. That is, the fitted model does not readily decay to zero as the distance increases, as is observed from the observed rainfall at an hourly scale and is intuitive from the physical properties of storms. This arises because the model does not specify a spatial extent for the storm. Therefore, when a storm occurs, cells can simultaneously occur at any point within the simulation region, even at large distances. Therefore, the non-zero correlation observed in the model at large distances in Figure 3 is due to the co-occurrence of raincells at points separated by large distances within the region. This is a limitation of the model that occurs when large catchments are considered.

Nonetheless, for practical application of the model, the lack of fit to cross-correlations at larger distances is not a significant limitation. This is because (i) there is no need or desire to extrapolate the simulation region to distances larger than those used in the calibration, (ii) a typical application would more likely be interested in the impact of rainfall on given sub-catchments within the $100 \times 100$ km$^2$ region for which
cross-correlations at shorter distances are of greater importance. Therefore, with regards to the calibration statistics, the model is considered to provide a very good fit.

Results and discussion

Table 2 gives the calibrated parameters for each month. Rainfall for the Sydney region is highly diverse, such that any given month can experience a combination of convective

Table 1 Observed and fitted values; all statistics are dimensionless

<table>
<thead>
<tr>
<th>Month</th>
<th>$r_1$ Observed</th>
<th>$r_1$ Fitted</th>
<th>$r_2$ Observed</th>
<th>$r_2$ Fitted</th>
<th>$p_{24}$ Observed</th>
<th>$p_{24}$ Fitted</th>
<th>$p_{24}$ Observed</th>
<th>$p_{24}$ Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7.69</td>
<td>7.69</td>
<td>20.19</td>
<td>20.18</td>
<td>0.40</td>
<td>0.39</td>
<td>3.00</td>
<td>3.01</td>
</tr>
<tr>
<td>2</td>
<td>6.68</td>
<td>6.68</td>
<td>14.91</td>
<td>14.91</td>
<td>0.53</td>
<td>0.53</td>
<td>3.41</td>
<td>3.41</td>
</tr>
<tr>
<td>3</td>
<td>6.37</td>
<td>6.37</td>
<td>17.83</td>
<td>17.83</td>
<td>0.41</td>
<td>0.39</td>
<td>2.52</td>
<td>2.53</td>
</tr>
<tr>
<td>4</td>
<td>7.40</td>
<td>7.40</td>
<td>20.22</td>
<td>20.22</td>
<td>0.45</td>
<td>0.45</td>
<td>3.26</td>
<td>3.26</td>
</tr>
<tr>
<td>5</td>
<td>6.09</td>
<td>6.09</td>
<td>13.57</td>
<td>13.57</td>
<td>0.54</td>
<td>0.54</td>
<td>2.97</td>
<td>2.97</td>
</tr>
<tr>
<td>6</td>
<td>7.21</td>
<td>7.21</td>
<td>16.10</td>
<td>16.10</td>
<td>0.60</td>
<td>0.61</td>
<td>4.02</td>
<td>4.01</td>
</tr>
<tr>
<td>7</td>
<td>6.94</td>
<td>6.95</td>
<td>23.68</td>
<td>23.42</td>
<td>0.55</td>
<td>0.55</td>
<td>3.45</td>
<td>3.47</td>
</tr>
<tr>
<td>8</td>
<td>8.01</td>
<td>8.08</td>
<td>19.40</td>
<td>19.36</td>
<td>0.67</td>
<td>0.75</td>
<td>5.11</td>
<td>4.97</td>
</tr>
<tr>
<td>9</td>
<td>8.29</td>
<td>8.29</td>
<td>21.09</td>
<td>21.09</td>
<td>0.53</td>
<td>0.54</td>
<td>4.14</td>
<td>4.13</td>
</tr>
<tr>
<td>10</td>
<td>7.75</td>
<td>7.75</td>
<td>21.69</td>
<td>21.69</td>
<td>0.47</td>
<td>0.47</td>
<td>3.39</td>
<td>3.39</td>
</tr>
<tr>
<td>11</td>
<td>7.08</td>
<td>7.08</td>
<td>20.90</td>
<td>20.90</td>
<td>0.38</td>
<td>0.38</td>
<td>2.86</td>
<td>2.87</td>
</tr>
<tr>
<td>12</td>
<td>7.82</td>
<td>7.82</td>
<td>19.66</td>
<td>19.64</td>
<td>0.37</td>
<td>0.37</td>
<td>3.12</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Figure 2 Comparison of elevation and mean monthly rainfall, plotted with respect to longitude for two different months (a) January and (b) July
storms and frontal systems. Nonetheless, by comparing summer (months 12,1,2) and winter (months 6,7,8) parameters, the fitted model suggests that winter has less storms (lower $\lambda$), but that these storms are more intense, with more cells per storm (higher $\mu_C$), more intense cells (higher $\alpha$) and longer duration cells (lower $\eta$). The mean cell radius ($1/\phi$) lies approximately between 5–10 km for the various months, which is quite low, suggesting that the influence of an individual rainfall cell is quite isolated. However, this parameter is compensated for by large numbers of cells being associated with each storm. Thus, cross-correlations at large distances are more likely due to the widespread activity of a storm (a large number of small cells) rather than of storms having fewer cells but with long-range influence (large radius).

**Table 2 Monthly parameter estimates**

<table>
<thead>
<tr>
<th>Month</th>
<th>$\lambda$ $\text{hr}^{-1}$</th>
<th>$\mu_C$ $\text{hr}^{-1}$</th>
<th>$\beta$ $\text{hr}^{-1}$</th>
<th>$\eta$ $\text{hr}^{-1}$</th>
<th>$\alpha$</th>
<th>$\phi$ $\text{km}^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00387</td>
<td>34.1</td>
<td>0.0733</td>
<td>2.29</td>
<td>1.91</td>
<td>0.148</td>
</tr>
<tr>
<td>2</td>
<td>0.00193</td>
<td>101.1</td>
<td>0.0528</td>
<td>1.92</td>
<td>1.93</td>
<td>0.159</td>
</tr>
<tr>
<td>3</td>
<td>0.00435</td>
<td>49.0</td>
<td>0.0502</td>
<td>2.22</td>
<td>2.03</td>
<td>0.182</td>
</tr>
<tr>
<td>4</td>
<td>0.00212</td>
<td>81.2</td>
<td>0.0425</td>
<td>1.95</td>
<td>2.12</td>
<td>0.185</td>
</tr>
<tr>
<td>5</td>
<td>0.00262</td>
<td>67.0</td>
<td>0.0487</td>
<td>1.61</td>
<td>1.82</td>
<td>0.168</td>
</tr>
<tr>
<td>6</td>
<td>0.00129</td>
<td>158.9</td>
<td>0.0486</td>
<td>1.50</td>
<td>2.16</td>
<td>0.229</td>
</tr>
<tr>
<td>7</td>
<td>0.00184</td>
<td>275.0</td>
<td>0.0473</td>
<td>1.60</td>
<td>2.92</td>
<td>0.137</td>
</tr>
<tr>
<td>8</td>
<td>0.00073</td>
<td>267.0</td>
<td>0.0377</td>
<td>0.79</td>
<td>2.63</td>
<td>0.170</td>
</tr>
<tr>
<td>9</td>
<td>0.00175</td>
<td>153.6</td>
<td>0.0883</td>
<td>2.12</td>
<td>2.45</td>
<td>0.214</td>
</tr>
<tr>
<td>10</td>
<td>0.00291</td>
<td>76.3</td>
<td>0.0865</td>
<td>2.10</td>
<td>2.32</td>
<td>0.118</td>
</tr>
<tr>
<td>11</td>
<td>0.00357</td>
<td>63.2</td>
<td>0.0606</td>
<td>2.54</td>
<td>2.18</td>
<td>0.127</td>
</tr>
<tr>
<td>12</td>
<td>0.00334</td>
<td>37.2</td>
<td>0.0735</td>
<td>2.71</td>
<td>1.81</td>
<td>0.148</td>
</tr>
</tbody>
</table>

**Figure 3** Comparison of observed and fitted hourly cross-correlation against the distance between sites, for two different months (a) January and (b) July.
A simulation of the SNSRP proceeds as a stationary process, which in particular produces a constant mean. Therefore, a scaling factor is required to represent the observed mean at a given point. The mean hourly rainfall was used as the scaling factor, though note that this choice is arbitrary. As this statistic is only calculated at observed sites, an interpolation of the scaling factor is required. Figure 4 shows the mean hourly rainfall interpolated across the entire Sydney region, based upon the estimates at each observation location. The values were interpolated using ordinary kriging with an exponential covariance function fitted to the observed data.

Figure 5 provides a comparison of simulated and observed statistics for one gauge as an example of the temporal performance of the model. The observed statistics at this site are based on 41 years of data and the simulated upper and lower limits were based on 50 replicates having the same duration. Figure 5(a) shows a comparison of the hourly mean

![Figure 4](image1.png)

**Figure 4** Interpolated hourly mean rainfall depth, for two different months (a) January and (b) July

![Figure 5](image2.png)

**Figure 5** Comparison of simulated and observed monthly statistics at a single site. Simulated limits are based on maximum and minimum values from 50 replicates (a) mean hourly rainfall and standard deviation of rainfall and (b) probability of a dry ‘bin’ for 1 hour and 24 hour totals
and standard deviation across the 12 months with good agreement. This is expected as these statistics were used in the calibration procedure. Figure 5(b) demonstrates the probability of a dry ‘bin’ across all months for two different aggregates, hourly and daily. This statistic was not used in the calibration procedure, and the simulations tend to slightly underestimate the dry probability at an hourly level and overestimate it at the daily aggregation level. Figure 6 shows the intensity-duration-frequency (IDF) curves at three different aggregates, 1 hour, 24 hours, 72 hours. The simulation performs adequately, but at this site there is an overall underestimation of intensities, at each of the aggregates.

With respect to the spatial dimension, Figure 7 shows a simulation of the model over a 100 × 100 km² region for four successive hourly time-steps for a given storm in January. The total rainfall depth is the sum of depths from overlapping cells which is then scaled according to the scaling fields given in Figure 4. Even though the images have been scaled, the underlying structure of cylindrical cells is strongly evident. The images show some evidence of persisting raincells from one hour to the next, and some evidence of the scaling field is present. Visual inspection of the images leads to the qualitative conclusion that the rain-fields do not look ‘realistic’, but whether or not this limitation has an impact on runoff modelling would require further quantitative assessment.

Conclusions
The successful calibration and simulation of the SNSRP model was demonstrated for an extensive network of 85 gauges across the Sydney region. A good fit to calibration statistics was obtained but considerable care was required implementing least-squares with respect to the spatial cross-correlations. The scaling factor was spatially interpolated to enable non-stationary fields of rainfall to be obtained from the stationary simulation over the region. With respect to temporal statistics, a comparison of observed data to 50 simulated replicates
was given for one site as a representative example of the region, and the model showed reasonable agreement. Upon inspection of the simulated rain-fields the cylindrical cell structure was strongly evident, as this is an outcome of the highly idealized model formulation. Despite the simple formulation of the model, it is one of only a few capable of long-term continuous-simulation in both space and time. The model is capable of outputting rainfall at any desired set of points in the space, or as average values across a grid, which allows it to be inputted into conventional or distributed runoff models for practical applications. Further research would be required to assess whether limitations of the model (particularly the use of non-advecting cylinders) contribute to any noticeable discrepancy when considered in the context of a rainfall-runoff simulation.

Acknowledgements
This project was funded by an ARC discovery grant. The authors would also like to acknowledge Sydney Water for providing access to their extensive network of rainfall data used in this analysis.

References