A Neyman–Scott model with continuous distributions of storm types

Paul S. P. Cowpertwait[1]

April 10, 2010

Abstract

In previous studies, different types of precipitation (for example convective and stratiform) were modelled using superposed Poisson cluster processes. When the underlying processes are independent, statistical properties, up to third order, are obtained by aggregation of the properties of each independent point process. However, each superposition introduces further parameters, which can result in too many parameters. A continuum of storm types z is proposed, where z comes from a continuous probability distribution, and selected model parameters are taken to be functions of z. This has the effect of allowing for different types of storms through superposition whilst retaining a moderate number of model parameters. Using a uniform distribution for z, properties up to third order are re-derived for the Neyman–Scott model, and used to fit the model to a sixty year record from Wellington, New Zealand. The parameterization enables the exploration of whether storms with fewer cells, on average, tend to have heavier or lighter rainfall.

http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/3025 gives this article, © Austral. Mathematical Soc. 2010. Published April 10, 2010. ISSN 1446-8735. (Print two pages per sheet of paper.)

Contents

1 Introduction

Applications of stochastic rainfall models based on a Neyman–Scott or Bartlett–Lewis point process (Rodriguez-Iturbe et al., 1987; Cox and Isham, 1980) are now abundant in the literature, and the models are proving to be of value to problems in engineering hydrology (Wheater et al., 2005; Burton et al., 2008). For example, in the recent Engineering Mathematics and Applications Conference, held at the University of Adelaide in December 2009,[1] some applications of a Neyman–Scott spatial-temporal rainfall model to problems in urban drainage were outlined. These include applications to multi-million dollar engineering projects, such as the Thames Tideway Tunnels project, the Glasgow Metropolitan Strategic Drainage Plan, and the Auckland City Integrated Catchment Study (Cowpertwait, 2009). In most drainage studies, for the purpose of long-term planning and design, the stochastic rainfall models simulate long series of data at sites that lack sufficient historical records.

Past studies have indicated the value of using independent superposed processes to account for different types of precipitation, such as convective or stratiform rain (Cowpertwait, 2004). However, superposing multiple Neyman–Scott processes, for example, can result in too many model parameters. (Current empirical experience suggests that any more than about eight param-

[1] http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/2909

eters per season is likely to be excessive, since the sample moments used in model fitting are highly correlated.) Hence, there is a need to develop a parsimonious model structure that is still capable of allowing for different types of precipitation. The purpose of this paper is to outline one approach to this problem, based on using a superposition of independent and continuous storm types. Cox and Isham (1980) give a general discussion on the superposition of point processes.

2 Superposed continuous NSRP processes

As in the original Neyman–Scott Rectangular Pulses model (Rodriguez-Iturbe et al., 1987), let storm origins occur in a Poisson process with rate λ. Now suppose each storm origin is of random type z that has continuous probability density function f_z so that a type z storm origin follows an independent Poisson process with rate $\lambda f_z(z)\,dz$. Each type z storm origin is followed by a random number $C(z)$ of cell origins that have waiting times, relative to the storm origin, that are independent exponential random variables with parameter $\beta(z)$. Hence, type z cell origins follow a Neyman–Scott point process with rate $\lambda\nu(z)f_z(z)\,dz$, where $\nu(z) = E[C(z)]$. Type z cell origins form a marked point process of "rain cells" where each cell has a random lifetime $L(z)$, taken to be an independent exponential random variable with parameter $\eta(z)$, and a random intensity $X(z)$ that remains constant throughout the cell lifetime. Without loss of generality, we take $X(z)$ to be an independent exponential random variable with mean $\theta(z)$, and take $C(z)$ to be a Poisson random variable with mean $\nu(z)$. The original Neyman–Scott Rectangular Pulses (NSRP) model occurs as the special case of a single storm type, which follows by setting f_z equal to the Dirac delta function and the model parameters to be constants: $\beta(z) = \beta$, $\eta(z) = \eta$, $\nu(z) = \nu$, $\theta(z) = \theta$.

Let $N_z(t)$ be the counting process of type z cell origins and $Y_z(t)$ be the total rain intensity due to type z storms. Then,

$$Y_z(t) = \int_{u=0}^{\infty} X_{z,t-u}(u)\, dN_z(t-u) \tag{1}$$

where $X_{z,t-u}(u)$ is the intensity at time t due to a type z cell with origin at time $t-u$, which takes the value $X(z)$ with probability $e^{-\eta(z)\,u}$ (or zero otherwise). The total rain intensity $Y(t)$ due to the superposed process of all possible storm types is $Y(t) = \int_z Y_z(t)$, from which the mean of the total intensity process follows as

$$\mathrm{E}\,[Y(t)] = \int_z \int_{u=0}^{\infty} \mathrm{E}\,[X_{z,t-u}(u)]\ \mathrm{E}\,[dN_z(t-u)] = \lambda \int_z \frac{\nu(z)\theta(z)f_z(z)}{\eta(z)}\, dz\,. \tag{2}$$

With the parameters set to constant values and f_z to the delta function, the mean intensity $\lambda\nu\theta/\eta$ of the original NSRP model is recovered. A further special case, that is particularly tractable and is used in the empirical study of Section 3, is to take z to be a uniform random variable ($f_z(z) = 1\,; 0 < z < 1$), and set the parameter functions to $\eta(z) = \eta$, $\nu(z) = \nu_0 + z\nu_1$, and $\theta(z) = z\theta$, in which case the mean total intensity is

$$\mathrm{E}\,[Y(t)] = \frac{\lambda\theta(\frac{1}{2}\nu_0 + \frac{1}{3}\nu_1)}{\eta}\,. \tag{3}$$

The relationships from which (3) is derived are linear and give insight into whether storms with low cell intensities on average tend to correspond to storms with low or high average numbers of rain cells. For example, if $\nu_1 > 0$ then storms with smaller expected cell numbers, given by ν_0 as $z \to 0$, have lower expected cell intensities since $z\theta \to 0$ as $z \to 0$.

Let $\{Y_{z,i}^{(h)} : i = 1, 2, \ldots\}$ be the aggregated rainfall series sampled over an interval of width h due to a type z storm, so that $Y_{z,i}^{(h)} = \int_{(i-1)h}^{ih} Y_z(t)\, dt$. The total rainfall in the ith time interval due to the superposition of all storm types is $Y_i^{(h)} = \int_z Y_{z,i}^{(h)}$. Properties, up to third order, of the aggregated superposed process are the sum of the equivalent properties for each independent type z process

$$\mathsf{E}\left[\{Y_i^{(h)} - \mu^{(h)}\}^k\right] = \int_z \mathsf{E}\left[\{Y_{z,i}^{(h)} - \mu_z^{(h)}\}^k\right], \quad k \leqslant 3, \tag{4}$$

where $\mu^{(h)} = \mathsf{E}\left[Y_i^{(h)}\right]$ and $\mu_z^{(h)} = \mathsf{E}\left[Y_{z,i}^{(h)}\right]$. Aggregated properties up to third order then follow by substituting $\nu(z)$, $\beta(z)$, $\eta(z)$ and $\theta(z)$ for ν, β, η and θ respectively, and $\lambda f_z(z)\,dz$ for λ in the statistical properties derived in the previous work (Rodriguez-Iturbe et al., 1987; Cowpertwait, 1998) and using (4) to give

$$\begin{aligned}
\mu^{(h)} &= \mathsf{E}[Y_i^{(h)}] = \lambda h \int_z \frac{\nu(z)\theta(z)f_z(z)}{\eta(z)}\,dz; && (5)\\
\gamma(h,l) &= \mathrm{Cov}[Y_i^{(h)}, Y_{i+l}^{(h)}] \\
&= \lambda \int_z f_z(z)\,\nu(z)\theta(z)^2 \Big[A(h,l,z)\eta(z)^{-3} \\
&\qquad \times \left\{4 + \beta(z)^2\nu(z)\left[\beta(z)^2 - \eta(z)^2\right]^{-1}\right\} \\
&\qquad - B(h,l,z)\nu(z)\left\{\beta(z)\left[\beta(z)^2 - \eta(z)^2\right]\right\}^{-1}\Big]\,dz && (6)
\end{aligned}$$

where

$$\begin{aligned}
A(h,l,z) &= \begin{cases} h\eta(z) + e^{-\eta(z)h} - 1\,, & l = 0\,, \\ \frac{1}{2}\left[1 - e^{-\eta(z)h}\right]^2 e^{-\eta(z)h(l-1)}, & l = 1, 2, \ldots\,, \end{cases} \\
B(h,l,z) &= \begin{cases} h\beta(z) + e^{-\beta(z)h} - 1\,, & l = 0\,, \\ \frac{1}{2}\left[1 - e^{-\beta(z)h}\right]^2 e^{-\beta(z)h(l-1)} & l = 1, 2, \ldots\,; \end{cases}
\end{aligned}$$

and lastly the third moment

$$\begin{aligned}
\xi(h) &= \mathsf{E}\left[\left\{Y_i^{(h)} - \mu^{(h)}\right\}^3\right] \\
&= \lambda \int_z f_z(z)\nu(z)\theta(z)^3\eta(z)^{-4}
\end{aligned}$$

$$\times \Big[36h\eta(z) - 72 + 36h\eta(z)e^{-\eta(z)h} + 72h\, e^{-\eta(z)}$$
$$+ 3\nu(z)p\left(\eta(z), \beta(z), h\right)\beta(z)^{-1}\left(\beta(z)^2 - \eta(z)^2\right)^{-2}$$
$$+ \nu(z)^2 q\left(\eta(z), \beta(z), h\right)\left\{2\beta(z)(\eta(z)^2 - \beta(z)^2)\right.$$
$$\left. \times (\eta(z) - \beta(z))(2\beta(z) + \eta(z))(\beta(z) + 2\eta(z))\right\}\Big]\, dz \qquad (7)$$

where $p(.)$ and $q(.)$ are high order polynomials (Cowpertwait, 1998).

3 Fitted model

There are many possible choices for the density f_z and model parameter functions in Equations (5)–(7) but, as commented in Section 2, a particularly tractable choice is to take z to be a uniform random variable, $f_z(z) = 1$, $0 < z < 1$, and to set the parameter functions to $\eta(z) = \eta$, $\nu(z) = \nu_0 + z\nu_1$, and $\theta(z) = z\theta$. Based on this choice the integrals in (5)–(7) were evaluated, and the model properties implemented into a model fitting algorithm written in R (R Development Core Team, 2009).

Using data from a site in Kelburn (near Wellington) and the previous fitting procedure (Cowpertwait, 2004), the model was fitted to sample values of the coefficient of variation α_h, the lag one autocorrelation ρ_h, and the skewness κ_h at aggregation levels $h = 1$, 6 and 24. A good fit was obtained to these properties. Table 1 lists the resulting parameter estimates.

From Table 1 a clear seasonal variation can be discerned in the estimates, with higher storm rates λ over the winter months, for example. However, it is appropriate to focus on the new parameters, ν_0 and ν_1, since these have not been studied before. For most of the year $\widehat{\nu}_1 > 0$, which implies storms with high expected numbers of cells have higher mean cell intensities. The converse is also implied: storms with lower mean cell intensities are expected to have fewer rain cells. Hence, for this data, there is a positive correlation between cell numbers and cell intensity. Two months (April and May) have

Table 1: Parameter estimates for Kelburn.

Month	$\widehat{\lambda}$ (h^{-1})	$\widehat{\nu}_0$ (cells per storm)	$\widehat{\nu}_1$ (cells per storm)	$\widehat{\beta}$ (h^{-1})	$\widehat{\eta}$ (h^{-1})	$\widehat{\theta}$ ($mm\,h^{-1}$)
1	0.00591	5.54	8.80	0.0897	1.02	3.13
2	0.00687	2.51	12.50	0.1008	1.04	2.92
3	0.00774	4.15	11.46	0.1061	1.20	3.07
4	0.00724	15.25	−1.54	0.0958	1.36	3.42
5	0.00768	16.83	−0.78	0.0828	1.46	3.75
6	0.00831	14.37	5.70	0.0782	1.41	3.61
7	0.00955	10.33	11.24	0.0842	1.21	2.55
8	0.0117	6.18	13.10	0.0975	1.03	1.90
9	0.0130	4.44	12.25	0.1084	1.05	1.72
10	0.0107	7.83	5.65	0.1029	1.12	2.66
11	0.00799	8.91	3.81	0.0903	1.15	3.25
12	0.00626	7.85	5.84	0.0849	1.10	3.44

negative values for ν_1 which suggests for these months higher cells numbers have lower than average cell intensities. However, this is only slight, because the estimates for ν_1 are small ($\nu_1 < 2$); hence the original model, which is obtained as the special case $\nu_1 = 0$, may be adequate for April and May (Table 1).

To assess the fit to extreme values that are not used in the fitting procedure but which are important in many hydrological applications, two hundred samples were simulated using the fitted model (Table 1). These samples consisted of simulated 1 h rainfall depths and each had record lengths (60 years) equal to the record length of the historical data. For the simulated series, annual maximum values at the 1 h aggregation level were extracted. These were ordered and the median of the ordered values found. In addition, approximate 95% confidence intervals of the ordered maximum values were found from the simulated samples. Together with the ordered 1 h annual maxima taken from the historical record, these were plotted against the reduced Gumbel variate and are shown in Figure 1. This procedure was repeated at the 24 h (daily) aggregation level and the result shown in Figure 2.

There is some evidence of underestimation at the 1 h level at return periods in excess of five years (Figure 1). For the 24 h level the observed values fall within the confidence bands indicating a good fit to daily extreme values (Figure 2). Further empirical work is needed to compare these results with those obtained for other parameterizations of the model.

4 Conclusions

A continuous distribution of storm types was built into the parameterization of the NSRP model. A uniform distribution of storm types and linear functions for expected cell numbers and mean cell intensity was used in fitting the model to the Kelburn series. This allowed for different types of storms whilst retaining a parsimonious number of parameters. For most of year, there was

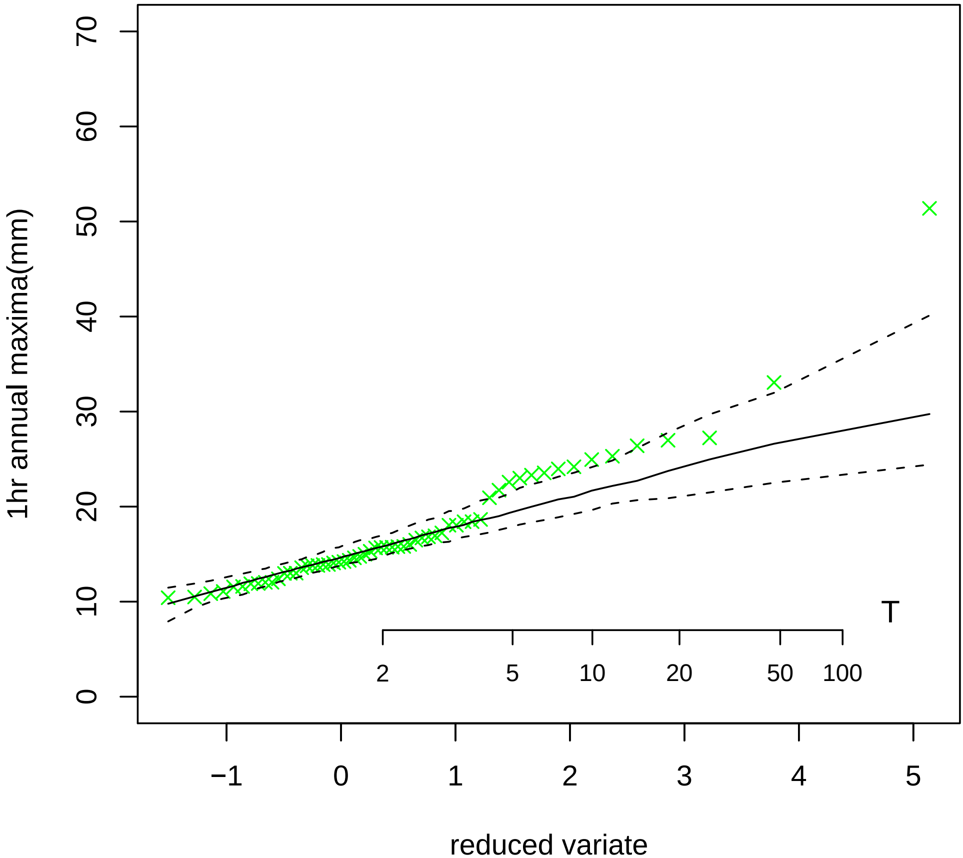

Figure 1: Annual maximum 1 h rainfall. The dotted lines are the 95% confidence intervals and the solid line the median value based on a simulated sample of 200 years. The crosses are the observed values for Kelburn.

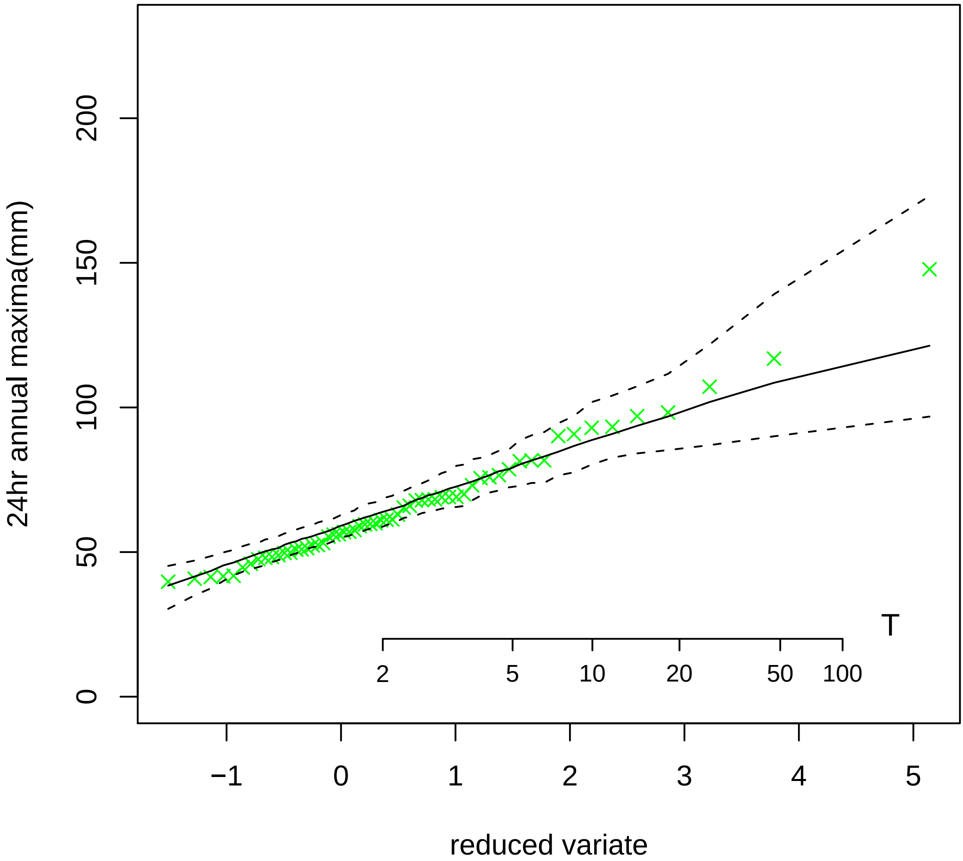

Figure 2: Annual maximum 24 h rainfall. The dotted lines are the 95% confidence intervals and the solid line the median value based on a simulated sample of 200 years. The crosses are the observed values for Kelburn.

a positive correlation between expected cell numbers and mean cell intensity. Further work is required to investigate alternative parameterizations.

Acknowledgements The New Zealand National Institute of Water and Atmospheric research (NIWA) are gratefully acknowledged for supplying the data. John Xie is thanked for producing the extreme value plots in R.

References

Burton, A., Kilsby, C., Fowler, H., Cowpertwait, P., and O'Connell, P. (2008). RainSim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling & Software*, 23:1356–1369. C98

Cowpertwait, P. (1998). A Poisson-cluster model of rainfall: High order moments and extreme values. *Proceedings of the Royal Society of London, Series A*, 454:885–898. C101, C102

Cowpertwait, P. (2004). Mixed rectangular pulses models of rainfall. *Hydrology and Earth System Sciences*, 8:993–1000. C98, C102

Cowpertwait, P. (2009). Spatial-temporal poisson cluster models of rainfall: Applications and further developments. 9th Engineering Mathematics and Applications Conference, Adelaide 6–9 December. C98

Cox, D. and Isham, V. (1980). *Point Processes.* Chapman & Hall. C98, C99

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. C102

Rodriguez-Iturbe, I., Cox, D., and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A*, 410:269–288. C98, C99, C101

Wheater, H., Chandler, R., Onof, C., Isham, V., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M.-L. (2005). Spatial-temporal rainfall modelling for flood risk estimation. *Stochastic Environmental Research and Risk Assessment*, 19:403–416. C98

Author address

1. **Paul S. P. Cowpertwait**, School of Computing & Mathematical Sciences, Auckland University of Technology, New Zealand.
mailto:paul.cowpertwait@aut.ac.nz