A spatial-temporal point process model with a continuous distribution of storm types

Paul S. P. Cowpertwait[1]

Received 6 July 2010; revised 29 August 2010; accepted 8 September 2010; published 2 December 2010.

[1] A point process rainfall model is further developed that has storm origins occurring in space-time according to a Poisson process, where each storm origin has a random radius so that storms occur as circular regions in two-dimensional space, where the storm radii are taken to be independent exponential random variables. Each storm origin is of random type z, where z follows a continuous probability distribution. Cell origins occur in a further spatial Poisson process and have arrival times that follow a Neyman-Scott point process. Each cell origin has a radius so that cells form discs in two-dimensional space, where the cell radii are independent exponential random variables. Each cell has a random lifetime and an intensity that remains constant over both the cell lifetime and cell disk area. Statistical properties up to third order are given for the model. Using these properties, the model is fitted to 10 min series taken from 23 sites across the Rome region, Italy. Distributional properties of the observed annual maxima are compared to equivalent values sampled from series that are simulated using the fitted model. The results indicate that the model will be of use in urban drainage projects for the Rome region.

Citation: Cowpertwait, P. S. P. (2010), A spatial-temporal point process model with a continuous distribution of storm types, *Water Resour. Res.*, *46*, W12507, doi:10.1029/2010WR009728.

1. Introduction

[2] Spatial-temporal point process models of rainfall are proving to be of value in hydrological catchment studies and in the analysis of urban drainage systems. For example, in the Thames Tideway Tunnels Project a Neyman-Scott spatial temporal model was used to stochastically generate multisite rainfall series at sites lacking data [*Cowpertwait*, 2006]. In the Auckland City Integrated Catchment Study, future rainfall scenarios were simulated using a Neyman-Scott model to allow for a possible change in climate under global warming [*Sharman et al.*, 2006; *Cowpertwait et al.*, 2009]. In both cases, a point process model of rainfall was used to simulate data to provide input into a hydrological catchment model, thus enabling the performance of a drainage system, under a range of scenarios, to be quantified. Examples of other hydrological applications of point process rainfall models are given by *Burton et al.* [2008] and *Wheater et al.* [2005]; while other studies of the spatial temporal rainfall process include *Mellor* [1996], *Northrop* [1998], *Cox and Isham* [1988], and *Waymire et al.* [1984]. An advantage of the spatial temporal model described herein is the availability of a third moment function for the aggregated process, which has been found useful in modeling extreme values [*Cowpertwait*, 1998].

[3] Rainfall variables, such as the intensity and the duration of an event, are correlated, which has been allowed for in point process model formulations in several previous studies. For example, rain cell intensity and duration have been treated as dependent random variables within the same storm [*Cowpertwait*, 1995; *Kakou*, 1997; *Kim and Kavvas*, 2006; *Evin and Favre*, 2008]. In addition, correlation has been introduced using independent superposed point processes to represent different storms types [*Cowpertwait et al.*, 2007]. The latter approach has the advantage in that statistical properties (up to third order) of the superposed process can be obtained by aggregation of the properties of each independent process, thus providing a flexible model structure that does not require extensive further derivations such as rederiving the third moment function. Different independent marked processes can also be superposed: the Poisson rectangular pulses model and the Neyman-Scott rectangular pulses model can be superposed to allow isolated single cell events to occur in the same series as longer-duration multicell storms.

[4] Point process rainfall models are fitted to observed sample properties using derived moments and model functions [*Wheater et al.*, 2005; *Cowpertwait et al.*, 2007; *Leonard et al.*, 2008; *Evin and Favre*, 2008; *Burton et al.*, 2008]. However, superposing independent point processes will result in too many model parameters when the number of parameters exceeds the number of independent sample properties used to fit the model. The sample properties—the mean, variance, third moment, and autocorrelation, at various levels of aggregation—used in previous studies are not independent. However, using more sample properties than model parameters in the optimization procedure can ensure a good overall fit [*Wheater et al.*, 2005; *Cowpertwait et al.*, 2007; *Leonard et al.*, 2008; *Evin and Favre*, 2008]. Nevertheless, the addition of further sample properties and model

[1]School of Computing and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand.

0043-1397/10/2010WR009728

parameters through superposition should not continue ad infinitum; there will come a point when insufficient additional information is contained in further sample properties to enable a robust estimation of the model parameters. Likelihood functions and, hence, information criteria, such as AIC [*Akaike*, 1974], are not usually available for point process rainfall models and, it has been argued, they may not provide a suitable estimation and assessment procedure because of the idealized and unrealistic model structure [*Wheater et al.*, 2005; *Rodriguez-Iturbe et al.*, 1988]. Work in progress suggests that a principle component analysis of the observed sample properties can be used to estimate the number of independent components needed to represent the data (J. Xie, Ph.D. thesis, submitted to Institute of Information and Mathematical Sciences, Massey University, Auckland, New Zealand, 2010). The total model parameters can then be kept close to (or below) that number of components.

[5] In previous studies, goodness of fit has been assessed by comparing observed properties that were not used to estimate the model parameters with equivalent properties simulated using the fitted model [*Entekhabi et al.*, 1989; *Cowpertwait et al.*, 2007; *Leonard et al.*, 2008; *Evin and Favre*, 2008]. Although this approach lacks statistical rigor, it is of practical value since properties that are important in hydrological applications, such as annual maxima at various aggregation levels, can be chosen. If simulated extreme values compare favorably with the equivalent observed values, then the fitted model can be used with confidence because the distributions of the extremes are sensitive to any departures in the goodness of fit of the depth distributions.

[6] Recently, a more parsimonious temporal superposed Neyman-Scott process was developed by taking a storm type to be a continuous random variable z and the model parameters to be functions of z [*Cowpertwait*, 2010]. In section 2, the temporal process is extended into two dimensional space by applying the same methodology to the spatial temporal Neyman-Scott model [*Cowpertwait*, 1995; *Leonard et al.*, 2008]. The spatial temporal model is then fitted to fine resolution series taken from 23 sites across the Rome region in Italy (section 4) and compared, via simulation, with properties not used in the fitting procedure (section 5). The fitted model described herein is intended for use in urban drainage projects in the Rome region. However, the flexible and parsimonious model parameterization structure implies that the model should be of practical value in other hydrological studies.

2. Model Definition and Intensity Process

[7] Let storm origins occur in a spatial temporal Poisson process with rate ζ_s (per unit area per unit time) and let $dN_s(r, \theta, t)$ be the number of storm origins in a small interval $(dr, d\theta, dt)$ at the location (r, θ), given in polar coordinates, and at time t, so that $E[dN_s(r, \theta, t)] = \zeta_s r\, dr\, d\theta\, dt$. Let each storm origin be of random type z, where z is independent and has a continuous probability density function f_z so that type z storm origins occur in a spatial temporal Poisson process with rate $\zeta_s f_z(z)dz$ and expected number $E[dN_{z,s}(r, \theta, t)] = \zeta_s f_z(z) r\, dr\, d\theta\, dt\, dz$ in $(dr, d\theta, dt)$. In the $\mathbb{R}^2$ plane, let a radius $R_s(z)$ be associated with each type z storm origin so that storms are discs in $\mathbb{R}^2$, where $R_s(z)$ is taken to be an independent exponential random variable with parameter $\gamma_s(z)$.

[8] The number of type z storm discs that cover an arbitrary point in $\mathbb{R}^2$ during a time interval $(0,T)$ is a Poisson random variable with mean $2\pi T\zeta_s f_z(z)\gamma_s^{-2}(z)dz$. Hence, storm discs overlap an arbitrary point in $\mathbb{R}^2$ at times that follow a Poisson process with rate λ given by

$$\lambda = 2\pi\zeta_s \int_z \frac{f_z(z)}{\gamma_s^2(z)} dz. \tag{1}$$

[9] Associated with a type z storm origin is a spatial Poisson process with rate $\zeta_c(z)$ (per unit area) of cell origins that may occur at any location in $\mathbb{R}^2$. It is assumed that the waiting time of a cell origin after a type z storm origin is an independent exponential random variable with parameter $\beta(z)$, so that the arrival times of cell origins occur in a Neyman-Scott point process. Hence, if $dN_{z,c,u}(r, \theta, v)$ is the number of cell origins in a small interval $(dr, d\theta, dv)$, at time v and location (r, θ), due to a type z storm origin at time u, then

$$E\left[dN_{z,c,u}(r,\theta,v)\right] = \zeta_c(z)\beta(z)e^{-\beta(z)(v-u)} r\, dr\, d\theta\, dv \quad (u < v). \tag{2}$$

[10] Each cell origin has a radius $R_c(z)$ that is an independent exponential random variable with parameter $\gamma_c(z)$ so that cells are discs in $\mathbb{R}^2$; the number of cell discs, associated with type z storms, overlapping an arbitrary point in $\mathbb{R}^2$ is a Poisson random variable with mean $\nu(z) = 2\pi\zeta_c(z)\gamma_c^{-2}(z)$ [*Cowpertwait*, 1995, section 2b]. Associated with each cell origin is a lifetime $L(z)$ and intensity $X(z)$, taken to be independent exponential random variables with parameters $\eta(z)$ and $\xi(z)$, respectively, where $X(z)$ remains constant throughout the cell lifetime and over the area of the cell disc. To allow for orographical effects (at least partially), the intensity $X(z)$ is scaled by a factor $\varphi_{\mathbf{p}}$ at point $\mathbf{p}\in\mathbb{R}^2$.

[11] For a type z storm origin occurring at time u, let $X_{\mathbf{p},z,u,v}(r_c, t-v)$ be the cell intensity at point $\mathbf{p}\in\mathbb{R}^2$ and time t due to a cell with center located at a distance r_c from $\mathbf{p}$ and with an arrival time v. Let $I_{\mathbf{p},z,u}(r_s)$ be an indicator function which takes the value 1 if, and only if, a storm origin arriving at time u with center a distance r_s from $\mathbf{p}$, overlaps $\mathbf{p}$ (or zero otherwise). It is assumed that only cells that overlap their associated storm disc in $\mathbb{R}^2$ contribute rain to any point $\mathbf{p}$ within a storm disc. The overall intensity process $Y_{\mathbf{p}}(t)$ is the superposition of the independent intensity processes $Y_{\mathbf{p},z}(t)$ due to type z storms, and is therefore given by

$$Y_{\mathbf{p}}(t) = \varphi_{\mathbf{p}} \int_z Y_{\mathbf{p},z}(t), \tag{3}$$

where

$$Y_{\mathbf{p},\mathbf{z}}(t) = \int_{\theta_s}\int_{r_s}\int_{\theta_c}\int_{r_c}\int_{u=-\infty}^{t}\int_{v=u}^{t} X_{\mathbf{p},z,u,v}(r_c, t-v) \cdot I_{\mathbf{p},z,u}(r_s) dN_{z,c,u}(\theta_c, r_c, v) dN_{z,s}(\theta_s, r_s, u). \tag{4}$$

[12] The variables $I_{\mathbf{p},z,u}(r_s)$ and $X_{\mathbf{p},z,u,v}(r_c, t-v)$ in (4) are independent of the counting processes (and each other) and

have expectations $e^{-\gamma_s(z)r_s}$ and $e^{-\eta(z)(t-v)}$ $e^{-\gamma_c(z)r_c}\xi^{-1}(z)$, respectively. The variables $dN_{z,c,u}(\theta_c, r_c, v)$ and $dN_{z,s}(\theta_s, r_s, u)$ are not independent, but the expectation of their product is independent of both $I_{\mathbf{p},z,u}(r_s)$ and $X_{\mathbf{p},z,u,v}(r_c, t - v)$ and is given by

$$E\left[dN_{z,c,u}(\theta_c, r_c, v)dN_{z,s}(\theta_s, r_s, u)\right] = \zeta_c(z)\beta(z)e^{-\beta(z)(v-u)}r_c dr_c d\theta_c \zeta_s f_z(z) r_s dr_s d\theta_s dv\, du\, dz \quad (5)$$

Hence,

$$E\left[Y_{\mathbf{p}}(t)\right] = 2\pi\zeta_s\varphi_{\mathbf{p}}\int_z \frac{\nu(z)f_z(z)}{\xi(z)\eta(z)\gamma_s^2(z)}dz. \quad (6)$$

[13] The mean of the intensity process of the original Neyman-Scott rectangular pulses model [*Rodriguez-Iturbe et al.*, 1987] is the special case obtained by setting f_z equal to the Dirac delta function, $\varphi_{\mathbf{p}} = 1$, and the parameters to constants in (6) and is $\lambda\nu\xi^{-1}\eta^{-1}$, where $\lambda = 2\pi\zeta_s\gamma_s^{-2}$ is the temporal rate of storms overlapping an arbitrary point $\mathbf{p} \in \mathbb{R}^2$ and $\nu = 2\pi\zeta_c\gamma_c^{-2}$ is the mean number of cells per storm that overlap $\mathbf{p}$.

[14] The cross covariance of the intensities $C_{\mathbf{p},\mathbf{q}}(\tau) = \mathrm{cov}[Y_{\mathbf{p}}(t), Y_{\mathbf{q}}(t+\tau)]$ at a pair of points $\mathbf{p}, \mathbf{q} \in \mathbb{R}^2$ is dependent on the probability that a cell overlaps both points. For a spatial process of discs in $\mathbb{R}^2$, with independent exponential (γ) radii and centers occurring in a spatial Poisson process, the probability that a disc overlaps $\mathbf{p}$ given that it overlaps $\mathbf{q}$ is $\kappa_{\mathbf{p},\mathbf{q}}[\gamma] = 2\pi^{-1}\int_0^{\pi/2}(1 + \frac{1}{2}\gamma d/\cos y)\, e^{-\frac{1}{2}\gamma d/\cos y}dy$, where $d = \|\mathbf{p} - \mathbf{q}\|$ is the distance between the points [*Cowpertwait*, 1995, section 2d; *Leonard et al.*, 2008]. Hence, the rate at which type z storm discs overlap both $\mathbf{p}$ and $\mathbf{q}$ is $2\pi\kappa_{\mathbf{p},\mathbf{q}}[\gamma_s(z)]\zeta_s\gamma_s^{-2}(z)f_z(z)dz$. This can be substituted for λ in the derived cross covariance of the generalized spatial temporal model (see equation 2.23 with $i = j$ of *Cowpertwait* [1995] and the remaining parameters replaced with functions of z), to give the cross covariance of the intensities cov $[Y_{\mathbf{p},z}(t), Y_{\mathbf{q},z}(t + \tau)]$, because the intensities from different storms are independent and therefore make no contribution to the cross-covariance function. The lag τ cross covariance, due to all storm types, thus follows as

$$C_{\mathbf{p},\mathbf{q}}(\tau) = 2\pi\varphi_{\mathbf{p}}\varphi_{\mathbf{q}} \cdot \int_z \frac{\left\{C_z(\tau) + 2\kappa_{\mathbf{p},\mathbf{q}}[\gamma_c(z)]\kappa_{\mathbf{p},\mathbf{q}}[\gamma_s(z)]e^{-\eta(z)\tau}\right\}\nu(z)\zeta_s f_z(z)}{\gamma_s^2(z)\eta(z)\xi^2(z)}dz, \quad (7)$$

where $C_z(\tau) = \beta(z)\nu(z)\{\beta(z)e^{-\eta(z)\tau} - \eta(z)e^{-\beta(z)\tau}\}\ \{\beta^2(z) - \eta^2(z)\}^{-1}$. By integrating over storm types, equation (7) also follows from *Leonard et al.* [2008, equation 13] (with the target region probability $\psi = 1$).

3. Time Series Properties

[15] Let $\{S_{\mathbf{p},i}(h)\}$ be a rainfall time series at $\mathbf{p} \in \mathbb{R}^2$ aggregated over discrete time intervals of width h, so that $S_{\mathbf{p},i}(h) = \int_{(i=1)h}^{ih} Y_{\mathbf{p}}(t)dt$, and let $S_{\mathbf{p},z,i}(h)$ be the aggregated total due to type z storms. From *Rodriguez-Iturbe et al.* [1987, equation 2.14] the single-site second-order temporal properties for $S_{\mathbf{p},i}(h)$ can be derived by integration of the autocovariance of the intensity process (7) with $\mathbf{p} = \mathbf{q}$, and the third moment function derived by following *Cowpertwait* [1998, section 2c]. However, the moments also follow by noting that $E\left[\{S_{\mathbf{p},i}(h) - \mu_{\mathbf{p}}(h)\}^k\right] = \int_z E\left[\{S_{\mathbf{p},z,i}(h) - \mu_{\mathbf{p},z}(h)\}^k\right](k \le 3)$ where $\mu_{\mathbf{p}}(h) = E[S_{\mathbf{p},i}(h)]$ and $\mu_{\mathbf{p},z}(h) = E[S_{\mathbf{p},z,i}(h)]$ [*Cowpertwait*, 2010, equation 4]. These are given in equations (8)–(10) below, with the autocovariance of the aggregated process (9), and are essentially the same as those given by *Cowpertwait* [2010, equations 5–7] with the exception that storms may have radii that depend on z and, hence, include terms with $\gamma_s(z)$:

$$\mu_{\mathbf{p}}(h) = E\left[S_{\mathbf{p},i}(h)\right] = 2\pi\varphi_{\mathbf{p}}h\int_z \frac{\nu(z)\zeta_s f_z(z)}{\gamma_s^2(z)\eta(z)\xi(z)}dz, \quad (8)$$

$$\Phi_{\mathbf{p}}(h,l) = \mathrm{cov}\left[S_{\mathbf{p},i}(h), S_{\mathbf{p},i+l}(h)\right] = 2\pi\varphi_{\mathbf{p}}^2\int_z \frac{\nu(z)\zeta_s f_z(z)}{\gamma_s^2(z)\xi^2(z)} \cdot \left[A(z,h,l)\eta(z)^{-3}\left\{4 + \beta(z)^2\nu(z)\left\{\beta(z)^2 - \eta(z)^2\right\}^{-1}\right\} - B(z,h,l)\nu(z)\left\{\beta(z)\left(\beta(z)^2 - \eta(z)^2\right)\right\}^{-1}\right]dz, \quad (9)$$

$$\Psi_{\mathbf{p}}(h) = E\left[\left\{S_{\mathbf{p},i}(h) - \mu_{\mathbf{p}}(h)\right\}^3\right] = 2\pi\varphi_{\mathbf{p}}^3\int_z \frac{\nu(z)\zeta_s f_z(z)}{\gamma_s^2(z)\eta^4(z)\xi^3(z)}\left[36h\eta(z) - 72 + 72he^{-\eta(z)} + 36h\eta(z)e^{-\eta(z)h} + 3\nu(z)Q_1(\eta(z),\beta(z),h)\beta(z)^{-1} \cdot \left(\beta(z)^2 - \eta(z)^2\right)^{-2} + \nu(z)^2 Q_2(\eta(z),\beta(z),h) \cdot \left\{2\beta(z)\left(\eta(z)^2 - \beta(z)^2\right)(\eta(z) - \beta(z)) \cdot (2\beta(z) + \eta(z))(\beta(z) + 2\eta(z))\right\}\right]dz. \quad (10)$$

In (9), $A(z, h, 0) = h\eta(z) + e^{-\eta(z)h} - 1$, $B(z, h, 0) = h\beta(z) + e^{-\beta(z)h} - 1$; for l a positive integer, $A(z, h, l) = \frac{1}{2}(1 - e^{-\eta(z)h})^2 e^{-\eta(z)h(l-1)}$, $B(z, h, l) = \frac{1}{2}(1 - e^{-\beta(z)h})^2 e^{-\beta(z)h(l-1)}$. In (10), Q_1 and Q_2 are high-order polynomials [*Cowpertwait*, 1998, equations 2.10 and 2.11].

[16] The rate at which type z storm discs overlap two points $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, $2\pi\kappa_{\mathbf{a},\mathbf{b}}[\gamma_s(z)]\zeta_s\gamma_s^{-2}(z)f_z(z)dz$, can be substituted for λ in the derived cross covariance of the aggregated process for the generalized spatial temporal model [*Cowpertwait*, 1995, equation 2.25] (with $i = j$) to give

$$\Phi_{\mathbf{a},\mathbf{b}}(h,l) = \mathrm{cov}\left[S_{\mathbf{a},i}(h), S_{\mathbf{b},i+l}(h)\right] = \varphi_{\mathbf{a}}\varphi_{\mathbf{b}}\varphi_{\mathbf{p}}^{-2}\Phi_{\mathbf{p}}(h,l) - 4\pi\varphi_{\mathbf{a}}\varphi_{\mathbf{b}} \cdot \int_z \frac{\nu(z)\kappa_{\mathbf{a},\mathbf{b}}[\gamma_s(z)]\zeta_s f_z(z)\left\{1 - \kappa_{\mathbf{a},\mathbf{b}}[\gamma_c(z)]\right\}A(z,h,l)}{\gamma_s^2(z)\eta(z)\xi^2(z)}dz. \quad (11)$$

[17] To obtain the probability that an arbitrary time interval $[(i-1)h, ih]$ is dry at an arbitrary point $\mathbf{p} \in \mathbb{R}^2$, note that $P[S_{\mathbf{p},i}(h) = 0]$ if, and only if, $P[S_{\mathbf{p},z,i}(h) = 0]$ for all storm types z. Storm types are independent, hence

$$P\left[S_{\mathbf{p},i}(h) = 0\right] = \exp\left[-2\pi\int_z \frac{\zeta_s f_z(z)}{\gamma_s^2(z)}\left\{\int_0^\infty \{1 - p(z,h,t)\}dt + \int_0^h \{1 - p(z,t,0)\}dt\right\}dz\right], \quad (12)$$

Table 1. Spatial Locations of Sites With 10 min Rainfall Gauges

Site j	Easting x_1 (m)	Northing x_2 (m)	Altitude x_3 (m)	Missing (%)
1	283506	4653164	144	1.6
2	289249	4652683	85	0.9
3	292813	4648564	18	1.4
4	285418	4647858	134	0.8
5	305408	4645738	68	1.4
6	292041	4645808	21	1.4
7	288748	4644362	130	1.8
8	298668	4643765	28	1.5
9	294271	4643919	9	1.5
10	290320	4642154	35	1.4
11	293962	4640504	76	1.6
12	289759	4638775	86	0.8
13	312566	4637521	82	1.4
14	290681	4638748	27	1.1
15	267095	4637319	3	1.8
16	296223	4638896	38	1.7
17	291911	4633464	53	2.4
18	279241	4632980	64	1.5
19	285636	4632106	10	1.9
20	270015	4627652	4	1.5
21	281157	4629155	4	0.9
22	297230	4625593	108	2.1
23	274498	4623187	6	2.0

where $p(z, h, t) = \exp[-\nu(z) + \nu(z)e^{-\beta(z)(t+h)} + \omega(z, t)\nu(z)\{1 - e^{-\beta(z)t}\}]$, and $\omega(z,t) = 1 - \beta(z)\{e^{-\beta(z)t} - e^{-\eta(z)t}\}/[\{\eta(z) - \beta(z)\}\{1 - e^{-\beta(z)t}\}]$ [*Cowpertwait*, 1995, equations 2.17 and 2.19].

4. Empirical Analysis

4.1. Data and Fitting Procedure

[18] In the following, 23 spatial points (sites) are used where each point corresponds to the location of a rain gauge from which observed data are available at the 10 min (1/6 h) aggregation level: $\{S_{\mathbf{p}_j,i}(h): j = 1,\ldots,23;\ h = 1/6\}$. The sites are located in the Rome region, Italy, and have data in the period 1992–2009 (Table 1). In the work presented herein sites are identified by their first two coordinates $\mathbf{p}(x_1, x_2) \in \mathbb{R}^2$. The percentage of missing values are also shown in Table 1 (missing values are excluded when estimating the sample properties used to fit the model).

[19] In order to fit the model to data a distribution needs to be selected for z and functions chosen for the model parameters. A simple choice, that also enables the properties up to third order to be evaluated in closed form, is to take z to be a uniform random variable $f_z(z) = 1$ $(0 < z < 1)$, $\nu(z) = (1 - z)\nu_0 + z\nu_1$, and $\xi(z) = 1/z$ with the remaining parameters set to constants [*Cowpertwait*, 2010]. The cell intensity $X(z)$ is scaled by $\varphi_{\mathbf{p}}$ so that, under this parameterization, the mean cell intensity is $\varphi_{\mathbf{p}}/2$. The relationship between expected cell numbers and expected cell intensity is linear and therefore makes an allowance for the correlation between these variables; the special case $\nu_0 = \nu_1$, which corresponds to the original Neyman-Scott rectangular pulses model, implies there is no correlation between these variables. The spatial storm radii parameter $\gamma_s(z)$ is a constant γ_s, which implies the temporal rate of storm arrivals at an arbitrary point $\mathbf{p} \in \mathbb{R}^2$ is also a constant $\lambda = 2\pi\zeta_s/\gamma_s^2$.

[20] For each site-month, the 1 h mean rainfall is divided into the 10 min values to produce standardized records with unit mean at the 1 h level of aggregation. For each month ($i = 1,\ldots,12$), the site data are pooled and the sample coefficient of variation $\mathrm{cv}_i(h)$, the lag 1 autocorrelation $\mathrm{ac}_i(h)$, and the sample skewness $\mathrm{sk}_i(h)$ evaluated at aggregation levels $h = 1/6$, 1, and 24 h. These sample properties $\hat{g}(h)$ have equivalent model functions $g(h) \in G = \{\Phi_{\mathbf{p}}^{1/2}(h, 0)h^{-1}, \Phi_{\mathbf{p}}(h, 1)\Phi_{\mathbf{p}}^{-1}(h, 0), \Psi_{\mathbf{p}}(h)\Phi_{\mathbf{p}}^{-3/2}(h, 0): h = 1/6, 1, 24\}$ (from (8)–(10)), where each member of G is a dimensionless function of λ, ν_0, ν_1, β, η and takes the same value for all sites $\mathbf{p} \in \mathbb{R}^2$.

[21] Following *Cowpertwait et al.* [2007, equation 6.4], the parameters λ, ν_0, ν_1, β, η are estimated for each month i by minimizing $\Sigma\{(1 - g/\hat{g})^2 + (1 - \hat{g}/g)^2\}$ where the summation is over G (Table 2). Using the estimates of λ, ν_0, ν_1, η, the scale parameter $\varphi_{\mathbf{p}_j,i}$ is estimated for each site $\mathbf{p}_j$ and each month i using the equation

$$\hat{\varphi}_{\mathbf{p}_j,i} = \frac{6\hat{\mu}_{\mathbf{p}_j,i}(1)\hat{\eta}_i}{\hat{\lambda}_i\left(\hat{\nu}_{0,i} + 2\hat{\nu}_{1,i}\right)}, \tag{13}$$

where $\hat{\mu}_{\mathbf{p}_j,i}(1)$ is the 1 h sample mean (corresponding to equation (8)) for site $\mathbf{p}_j$ month i ($i = 1, \ldots, 12$, $j = 1,\ldots,23$). The resulting scale parameter estimates are given in Table 3.

[22] The storm and cell radii parameters, γ_s and γ_c, respectively, are estimated by minimizing $\Sigma_{g\in F}\{(1 - g/\hat{g})^2 + (1 - \hat{g}/g)^2\}$, where

$$F = \left\{\Phi_{\mathbf{p}_j,\mathbf{p}_k}(h, 0)\Phi_{\mathbf{p}_j}^{-1/2}(h, 0)\Phi_{\mathbf{p}_k}^{-1/2}(h, 0) : h = 1/6, 1, 24;\ j, k = 1, \ldots, 23; j \neq k\right\}, \tag{14}$$

that is, the summation is taken over the cross correlations for each pair of sites at aggregation levels 1/6, 1, and 24 h. The estimates for γ_s and γ_c are given in Table 2.

Table 2. Parameter Estimates for the Rome Region

Month i	$\hat{\lambda}_i$ (h^{-1})	$\hat{\nu}_{0,i}$ (cells per storm)	$\hat{\nu}_{1,i}$ (cells per storm)	$\hat{\beta}_i$ (h^{-1})	$\hat{\eta}_i$ (h^{-1})	$\hat{\gamma}_{c,i}$ (km^{-1})	$\hat{\gamma}_{s,i}$ (km^{-1})
1	0.00290	67.3	43.5	0.141	7.94	0.303	0.0145
2	0.00360	62.4	46.1	0.138	7.63	0.365	0.0150
3	0.00596	41.2	54.7	0.125	7.56	0.311	0.0152
4	0.00419	44.4	41.0	0.121	7.70	0.292	0.0144
5	0.00175	63.9	0.00657	0.130	7.77	0.457	0.0174
6	0.00130	31.4	0.00130	0.157	7.59	0.565	0.0419
7	0.00177	14.8	0.000768	0.205	7.11	0.452	0.0288
8	0.00255	12.9	0.000622	0.250	6.63	0.325	0.0175
9	0.00319	25.4	0.00133	0.215	6.46	0.284	≈ 0
10	0.00532	42.9	14.0	0.166	6.81	0.258	0.0100
11	0.00775	39.5	41.2	0.139	7.58	0.343	0.0158
12	0.00429	55.0	47.8	0.136	8.05	0.235	0.00722

Table 3. Scale Parameter Estimates $\hat{\varphi}_{\mathbf{p}j,i}$ (mm) for Each Month i at Each Site j[a]

	Month i											
Site j	1	2	3	4	5	6	7	8	9	10	11	12
1	9.36	7.57	4.37	9.16	28.0	40.2	26.8	34.3	54.0	17.4	7.99	10.6
2	8.51	6.80	4.29	8.39	28.8	46.2	35.6	41.1	50.2	15.8	6.91	9.95
3	9.47	7.49	4.24	9.37	32.1	28.5	30.8	34.5	49.1	15.0	7.35	10.5
4	9.82	6.99	4.13	9.18	24.9	37.4	38.3	33.3	52.0	15.9	7.33	10.9
5	9.31	7.51	4.45	10.5	34.4	43.1	54.3	45.8	51.5	14.7	7.40	10.4
6	9.97	7.33	4.32	9.16	30.2	35.3	40.7	33.0	48.1	16.3	6.89	10.5
7	9.71	6.83	4.05	8.84	29.2	31.2	32.6	35.0	49.0	16.1	5.96	9.56
8	9.17	7.61	4.05	9.05	29.0	39.0	49.9	36.9	48.6	15.3	7.08	9.81
9	9.12	7.38	4.22	9.39	27.2	36.7	45.9	37.8	54.4	17.1	7.28	10.3
10	9.23	6.84	3.82	8.74	28.6	32.7	21.9	37.8	49.8	15.7	6.87	10.1
11	9.06	7.32	4.14	8.40	25.8	36.4	33.4	38.4	47.8	16.0	7.14	10.1
12	9.20	6.63	3.76	8.68	27.5	37.2	21.8	33.5	49.5	15.4	7.18	10.2
13	9.05	7.66	4.84	9.76	36.0	44.6	49.6	54.3	48.0	16.0	7.90	12.1
14	9.14	7.10	3.92	9.26	27.6	43.9	25.8	36.8	48.7	17.0	7.15	10.7
15	7.41	4.92	2.63	6.17	15.1	26.9	14.8	20.9	40.7	13.8	5.35	7.82
16	8.07	7.17	3.84	8.61	26.7	28.5	36.4	45.9	55.6	15.6	7.14	9.87
17	9.35	6.33	3.72	8.29	28.4	33.7	31.4	44.1	47.0	15.6	6.89	9.07
18	9.43	5.90	3.53	8.47	21.3	32.5	19.6	32.3	56.8	16.5	7.89	9.77
19	8.94	6.84	3.91	8.04	25.0	34.7	16.3	31.5	51.1	16.6	7.34	10.5
20	8.40	5.60	2.99	7.80	18.2	29.3	13.8	22.3	51.3	16.6	6.81	9.34
21	8.09	6.06	3.72	7.76	20.9	43.4	22.3	24.7	46.3	14.7	7.12	10.0
22	9.52	6.82	3.24	7.62	25.7	27.2	26.4	55.2	48.9	15.1	8.55	10.5
23	8.72	5.76	3.12	7.14	17.2	28.9	26.7	25.5	44.6	16.2	7.28	8.63

[a]The element in the jth row and the ith column corresponds to the estimate for the ith month at the jth site (j = 1,…,23; i = 1,…,12).

4.2. Parameter Estimates

[23] In Table 2, there is clear evidence of seasonal variation. The rate of storm arrivals $\hat{\lambda}_i$ and expected number of cells $\frac{1}{2}(\hat{\nu}_0 + \hat{\nu}_1)$ both decrease over the summer months. In addition, both cells and storms tend to be more localized in the summer, because both $\hat{\gamma}_{c,i}$ and $\hat{\gamma}_{s,i}$ both increase, and this is expected because there would be more convective rain during summer months. The estimate of $\hat{\gamma}_s$ for September is small ($< 10^{-8}$; the lower bound in the optimization algorithm), which implies that storms occurring in that month will tend to cover the whole catchment. This is a special case of the model, similar to that given by *Cowpertwait* [2006] but containing a continuous distribution of storm types. The low values of $\hat{\nu}_{1,i}$ for the summer months imply a negative correlation between cells numbers and cell intensity, a characteristic of summer convective thunderstorms. When the estimates for ν_0 and ν_1 are approximately equal, the original NS model is probably satisfactory, and this happens for April and November (Table 2).

[24] The scale parameter estimates in Table 3 vary both seasonally and with location, and are easier to interpret through fitted regression models (Table 4). The estimated scale parameter $\hat{\varphi}_{\mathbf{p}_j,i}$ each month (i = 1, … ,12) is regressed on the site coordinates using the linear model

$$E\left[\hat{\varphi}_{\mathbf{p}_j i}\right] = c_{i,0} + c_{i,1}x'_{j,1} + c_{i,2}x'_{j,2} + c_{i,3}x'_{j,3}, \qquad (15)$$

where $x'_{j,k}$ is the adjusted $x_{j,k}$ coordinate (k = 1,2,3) of the jth site $\mathbf{p}_j$ $(x_{j,1}, x_{j,2}, x_{j,3}) \in \mathbb{R}^3$ obtained from: $x'_{j,k} = 10^{-3}(x_{j,k} - \bar{x}_k)$, where $\bar{x}_k$ is the sample mean of the kth coordinates. The coefficients are given in Table 4, from which it is clear that the scale parameter is higher in summer months, corresponding to an increased frequency in the occurrence of convective rain cells. In addition, there is a tendency for the estimates of the scale parameter to increase at sites further inland, as evident by the positive values of $\hat{c}_{i,1}$, and is due to the effects of orography. Most of the coefficients ($\hat{c}_{i,3}$) for altitude are not significant, essentially because the coordinates are correlated and higher altitudes correspond to sites further inland (Table 4).

4.3. Fitted Values

[25] The fitted temporal properties are shown in Figure 1. The model has a very good fit to the observed sample properties at each level of aggregation, following the strong seasonal variation closely (Figure 1). There is a slight underestimation of the 10 min skewness and a slight overestimation of the 24 h coefficient of variation (Figure 1).

[26] Selected cross correlations, to represent summer and winter, are shown in Figures 2–4. A good fit is obtained to the sample cross correlations with the fitted values (shown as lines) falling within the sample estimates (Figure 2–4). There is some overestimation for pairs of sites separated by spatial distances of approximately 40 km for January and

Table 4. Coefficients of Fitted Linear Models for $E[\hat{\varphi}_{\mathbf{p}_j,i}]$ and Adjusted r^{2}[a]

Month i	$\hat{c}_{i,0}$	$\hat{c}_{i,1}$	$\hat{c}_{i,2}$	$\hat{c}_{i,3}$	r^2
1	9.3	0	0	5.2*	0.24
2	7.1	0.050	0.034	0	0.80
3	4.1	0.030	0.030	0	0.76
4	8.9	0.054	0.048	0	0.66
5	29	0.39	0.20	0	0.89
6			0	0	0.14
7	35	0.80	0.41*	0	0.64
8	41	0.74	−0.26**	58	0.82
9	51	0	0	0	0
10	16	0	0	0	0
11	7.4	0.029**	−0.032*	0	0.26
12	10	0.052	0	0	0.47

[a]A coefficient of 0 has a p value $p > 0.1$. All other coefficients have $p < 0.01$, except those marked with one asterisk ($0.05 \le p \le 0.1$) and two asterisks ($0.01 \le p < 0.05$).

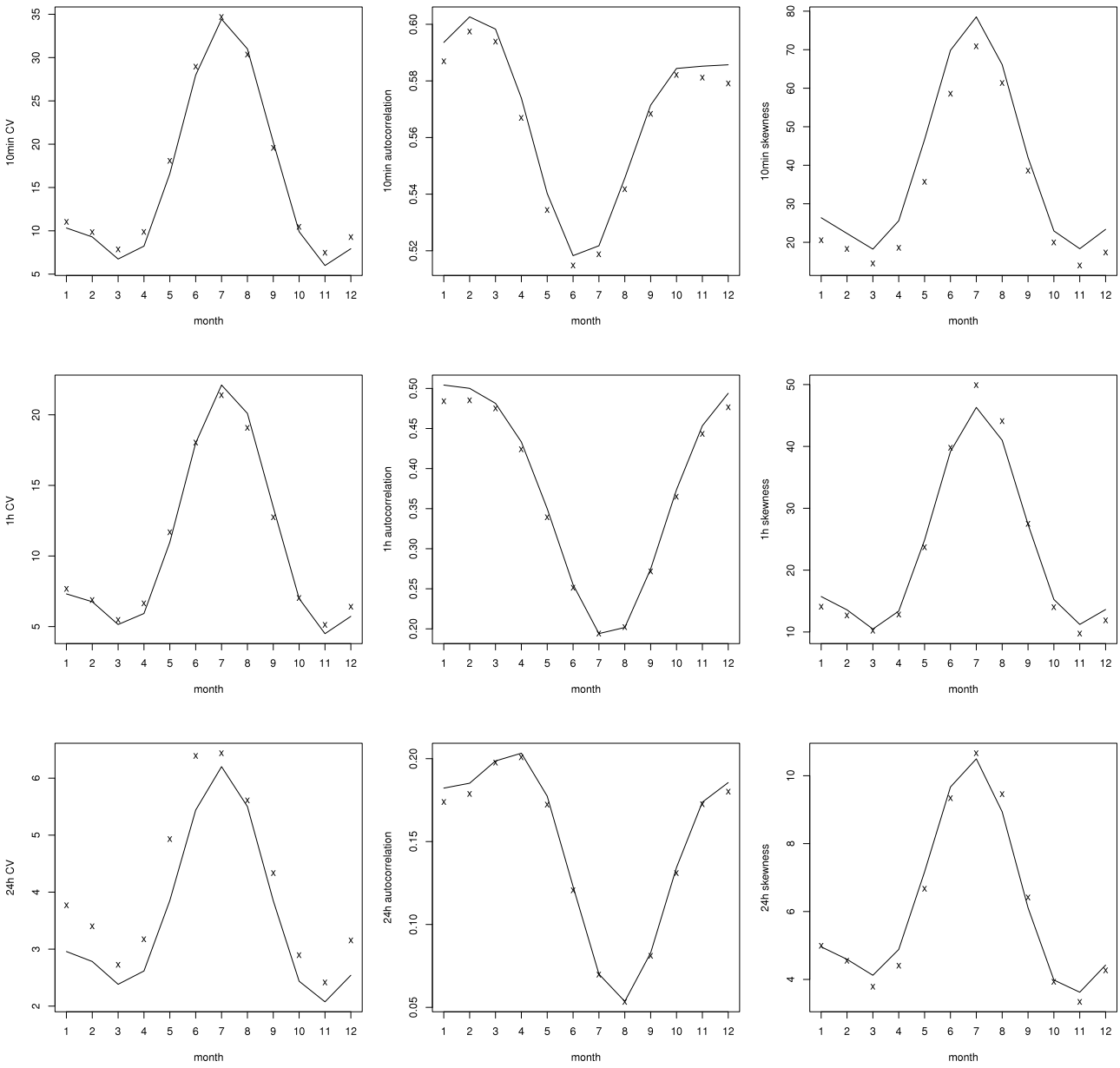

Figure 1. Sample (lines) and fitted properties (crosses): aggregation levels (top) $h = 1/6$ h, (middle) $h = 1$ h, and (bottom) $h = 24$ h and the properties (left) $\mathrm{cv}_i(h)$, (middle) $\mathrm{ac}_i(h)$, and (right) $\mathrm{sk}_i(h)$, plotted against month i.

February at the 1 h aggregation level (Figure 3). The cross correlations decay more rapidly with distance in the summer months, corresponding to an increased frequency of localized convective rain during those months (Figures 2–4). Conversely, the decay rate is slower for the winter months, especially at the daily level of aggregation (Figure 4), which corresponds to an increased frequency of stratiform or frontal rainfall (Figure 2–4). For all cases, the fitted values decay at approximately the same rate as the observed values (Figures 2–4).

5. Model Validation

[27] Five hundred years of 10 min data were simulated at the 23 sites across the region (Table 1) using the fitted model (Tables 2 and 3). To validate the model, properties not used in the fitting procedure were extracted from the simulated series and compared with equivalent properties sampled from the historical data.

[28] Figure 5 gives a box plot of the annual mean daily rainfall for each of the 23 sites. (Annual means are used instead of annual totals to avoid bias due to missing values in the historical records.) For each year and each site the observed annual mean daily rainfall is plotted over the box plot for the simulated series, where it is evident that the observed values fall within the simulated range, indicating a satisfactory fit (Figure 5).

[29] To compare extreme values the following procedure is used. The historical record is for the period July 1992 to August 2009, so the first and last year are pooled

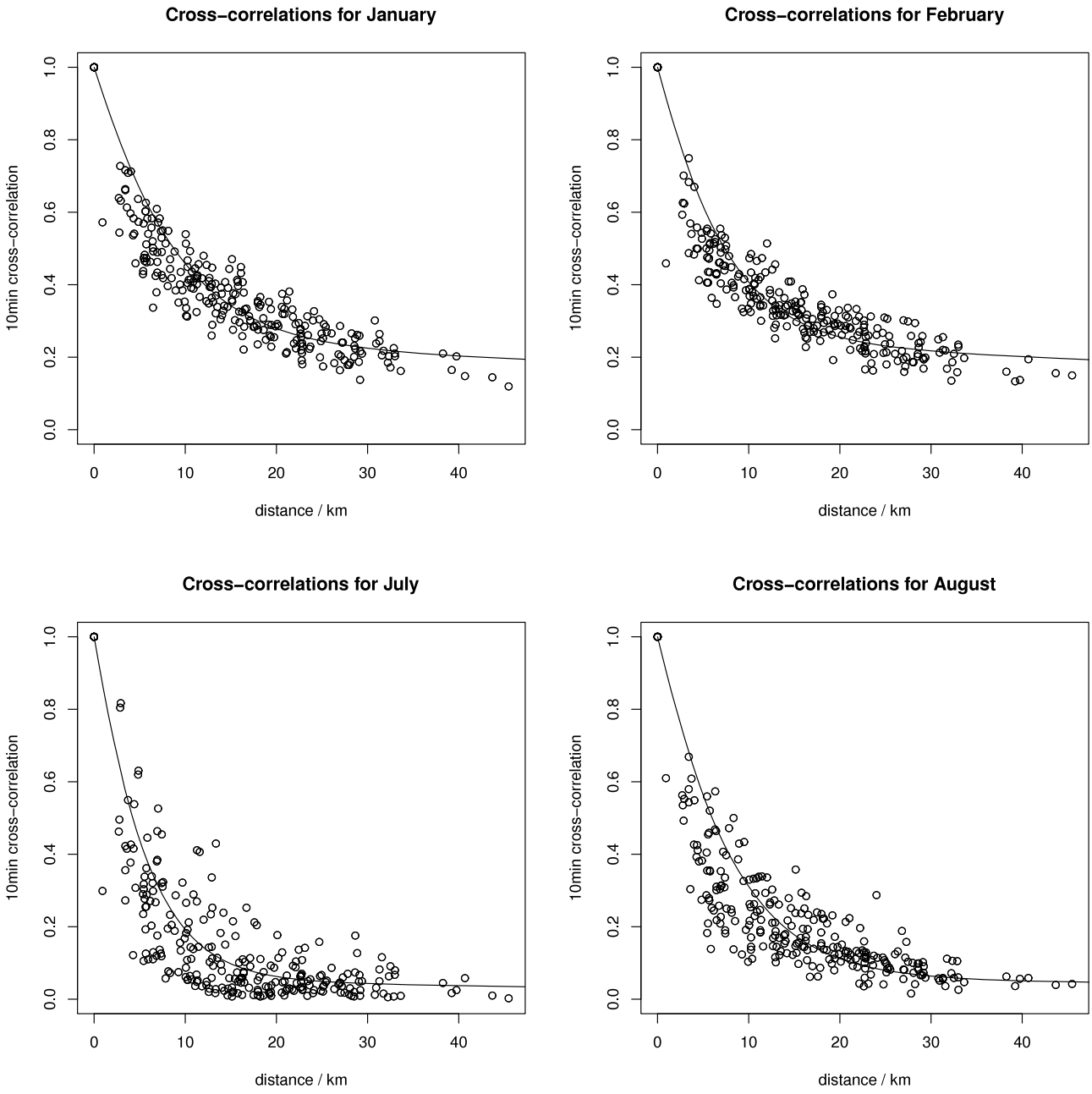

Figure 2. Fitted (lines) and observed (circles) cross correlations at the 10 min aggregation level between pairs of sites plotted against the distance between the sites.

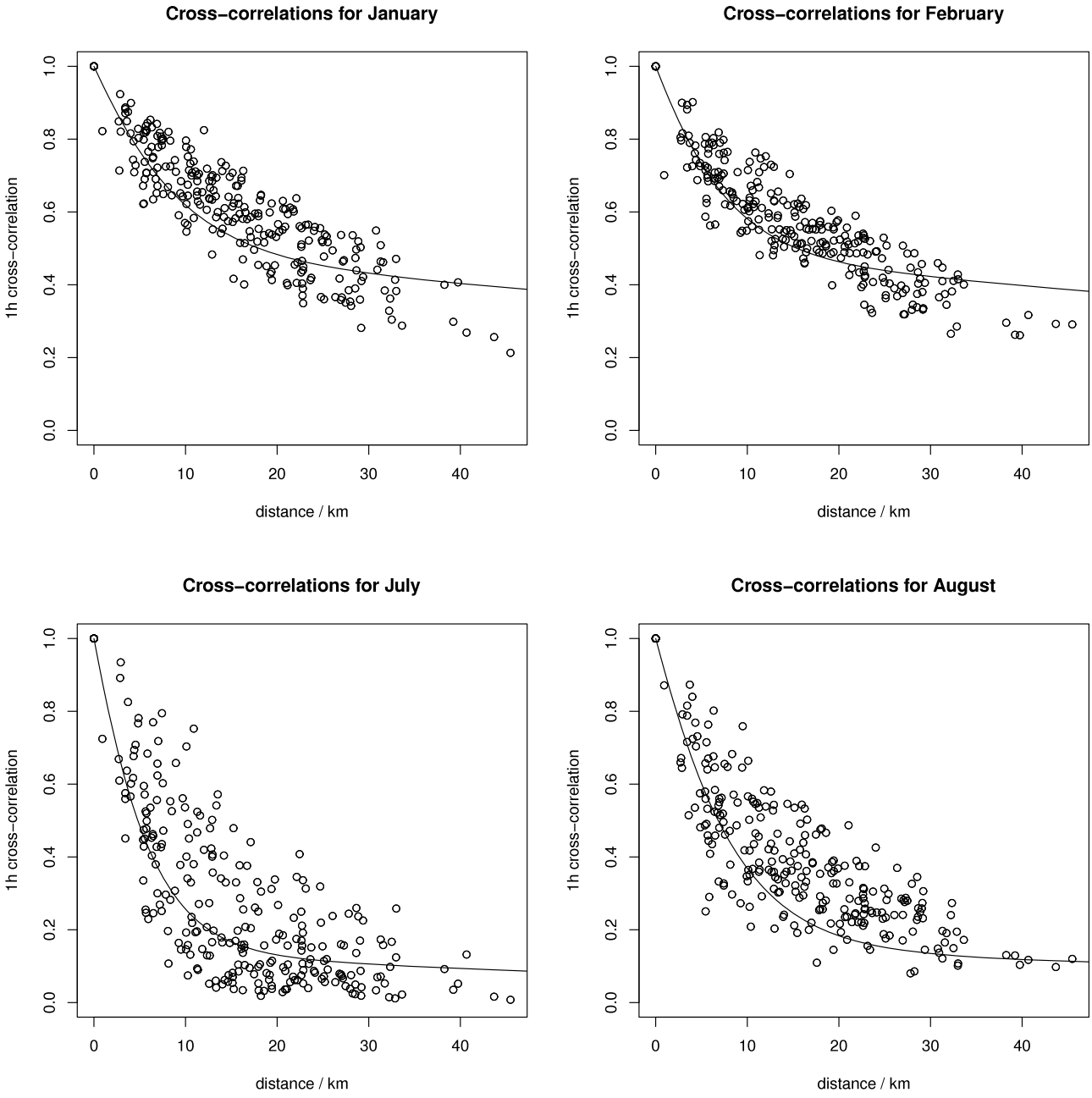

Figure 3. Fitted (lines) and observed (circles) cross correlations at the 1 h aggregation level between pairs of sites plotted against the distance between the sites.

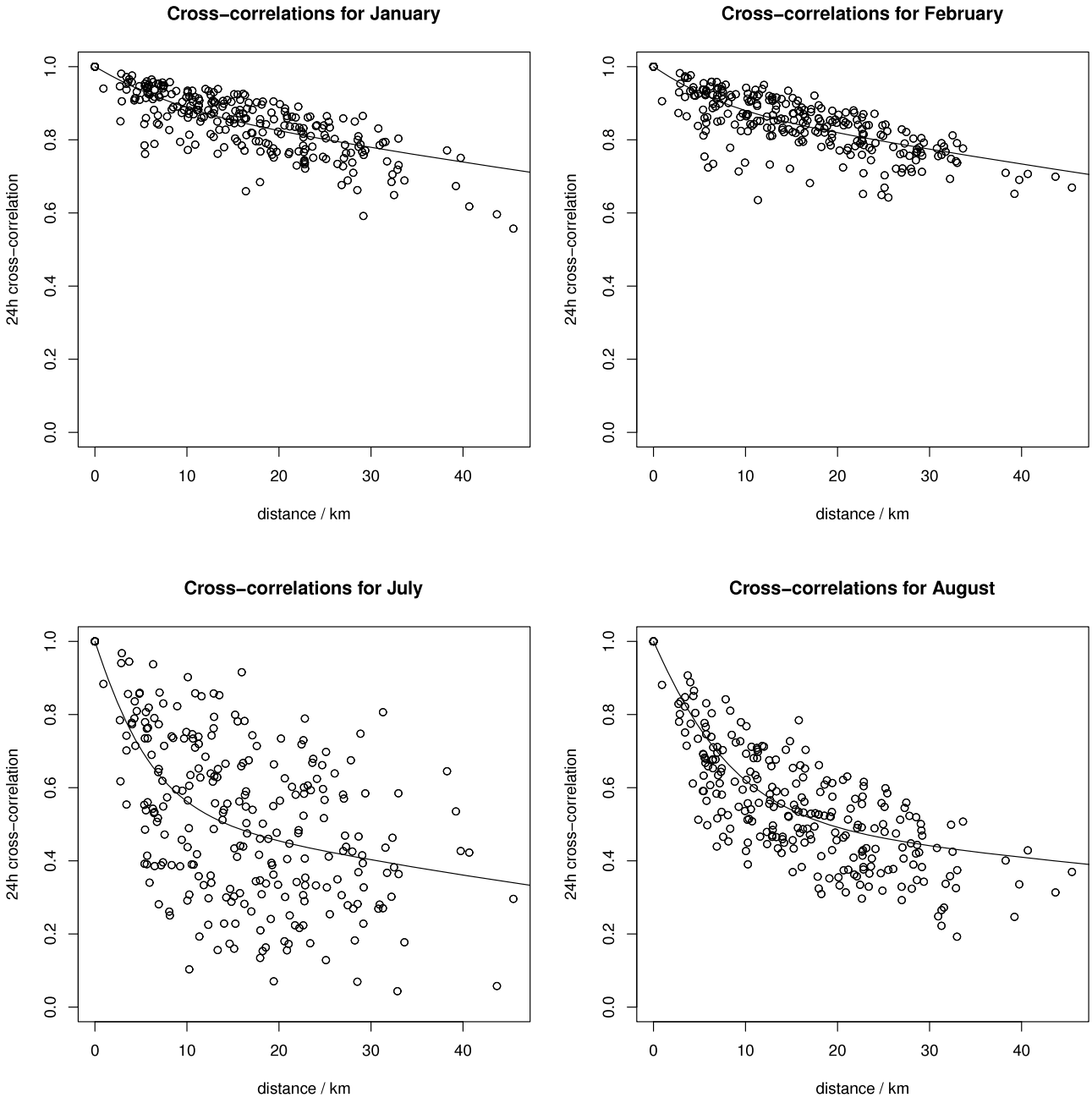

Figure 4. Fitted (lines) and observed (circles) cross correlations at the 24 h aggregation level between pairs of sites plotted against the distance between the sites.

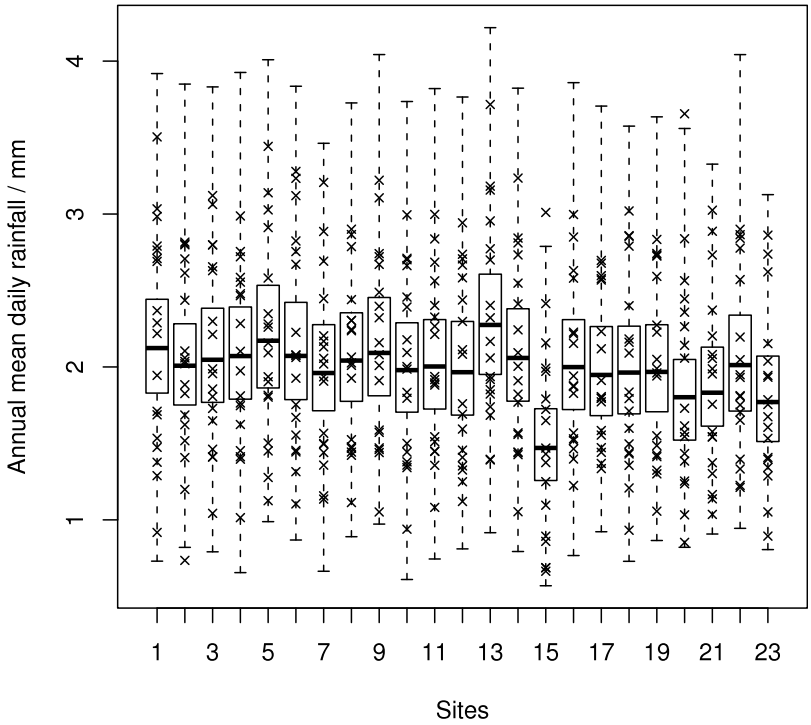

Figure 5. Annual mean daily rainfall: Box plots are for the 500 years of simulated data. The crosses are observed values which are expected to fall with the range of simulated values.

to give 17 years of data. The annual maximum at aggregation level h is found for each site-year in the historical series. For each year, the median, upper quartile, lower quartile, maximum, minimum, upper 5th percentile, and lower 5th percentile are found from the site values. These are then ordered, in order of lowest to highest value of the median, and plotted against the reduced Gumbel variate (Figure 6). The error bars in Figure 6 represent the interquartile range. For the simulated series, there are 29 samples of equivalent length over the 500 years from which to calculate these properties. So overall estimates for the simulated series are found in the following way.

[30] Let x_{ijk} be the annual maximum taken at aggregation level h in the ith year jth sample at site k ($i = 1,\ldots,17$; $j = 1\ldots,29$; $k = 1,\ldots,23$), and let y_{ij} be the median value for the i year, jth sample, i.e., $y_{ij} = P_{0.5,k}(x_{ijk})$, where $P_{\alpha,k}$ is a percentile function that returns the 100αth percentile across index k. Assume the index for the year corresponds to the order of the medians, i.e., $y_{ij} \leq y_{i+1,j}$ ($i = 1,\ldots,16$), and that this also holds for the year index in x_{ijk}. In Figure 6, the solid line is the overall median $P_{0.5,j}(y_{ij})$ plotted against the ith-ordered reduced variate. The ith-ordered estimate of the maximum, minimum, and quartiles are obtained by taking the median of the sample values obtained across the sites. For example, the estimate of the 100αth percentile is given by: $P_{0.5,j}(P_{\alpha,k}\ (x_{ijk}))$ ($\alpha = 0,0.25,0.75,1$). The maximum ($\alpha = 1$) and minimum ($\alpha = 0$) are shown as dashed lines in Figure 6. The dotted lines are the upper and lower quartiles, corresponding to the data used to calculate the median at a particular Gumbel ordinate. This procedure gives key properties of the simulated distribution of maximum values across the sites, enabling a "like with like" comparison to be made with the extreme value distributional properties taken from the historical records (Figure 6).

[31] Overall there is good agreement between the observed and simulated distributions of annual maxima across the sites; the distributions overlap and take a similar range of values (Figure 6). The lower return period extremes are an exception, as these tend have smaller values for the observed data (Figure 6). This is due to missing values in the historical records, which bias the observed estimates at the low return periods (74% of the missing values occurred in two calendar years). There is a slight tendency for the lower aggregation level (10 and 30 min) maxima to be overerestimated by the model at high return periods (Figure 6). This is due to the underestimation of skewness at the 10 min aggregation level (Figure 1). Therefore, if deemed necessary from a practical perspective, this could be improved by refitting the model and weighting this property in the optimization

Figure 6. Ordered site medians of annual maxima plotted against the reduced Gumbel variate. The observed values are shown as quartile error bars, with maximum and minimum and upper and lower 5th percentiles shown as crosses and the medians shown as dots. The solid line is an overall estimate of the ordered median for the simulated series. The dotted lines are estimates of the upper and lower quartiles, and the dashed lines are estimates of the maximum and minimum values of data used to calculate the ordered medians (see text for full explanation).

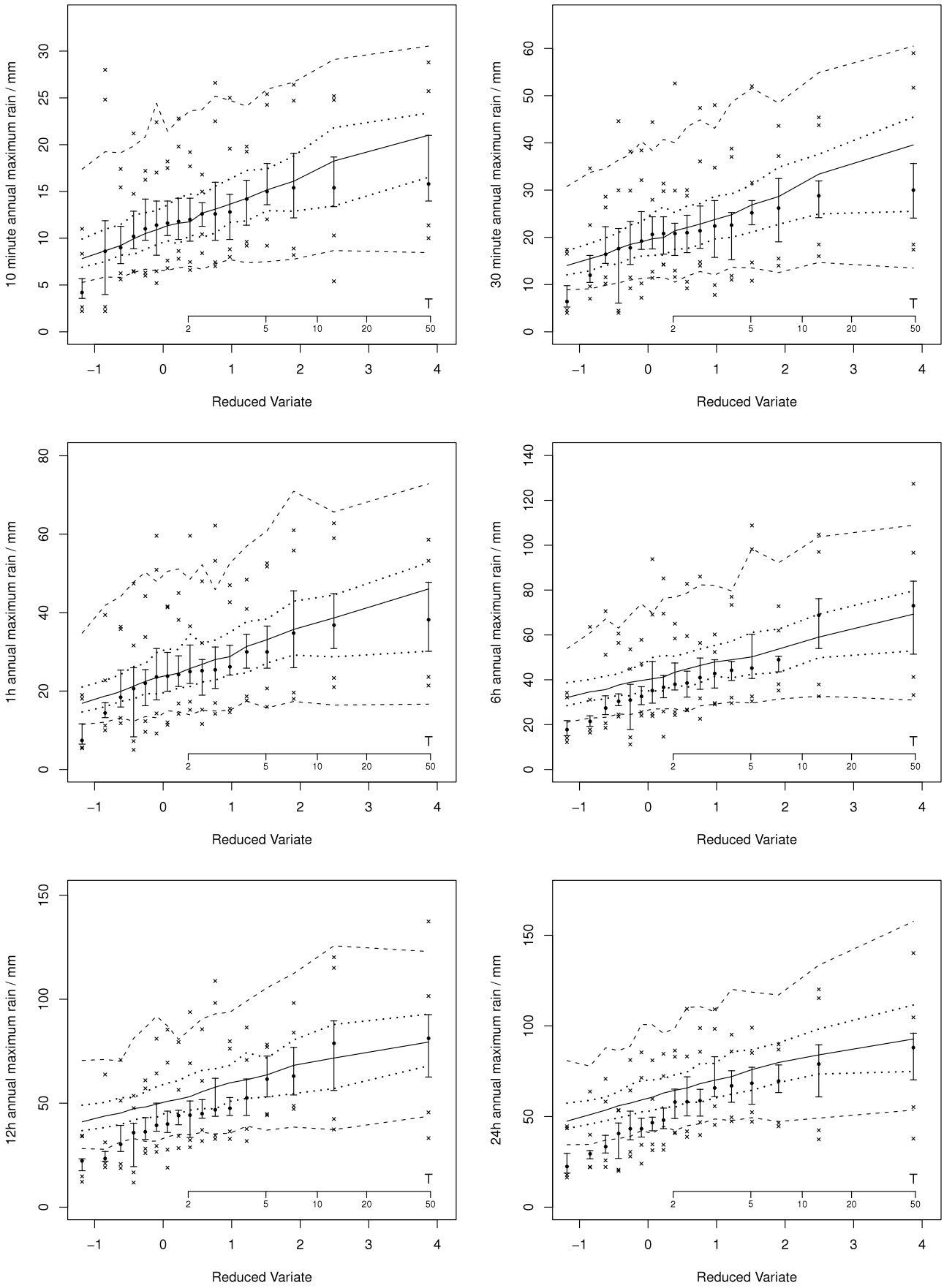
10 minute annual maximum rain / mm
30 minute annual maximum rain / mm
1h annual maximum rain / mm
6h annual maximum rain / mm
12h annual maximum rain / mm
24h annual maximum rain / mm
Reduced Variate
T
2
5
10
20
50

Figure 6

algorithm. Overall, Figure 6 indicates that the model provides a satisfactory fit to the observed extreme value distributions.

6. Summary and Conclusions

[32] A spatial temporal model with a continuous distribution of storm types was proposed and statistical properties up to third order given. The model has a very flexibly parameter structure and therefore the potential to provide a good fit to a large range of historical rainfall records.

[33] A special case of the model was fitted to statistical properties of the multisite data from the Rome region. These properties included the coefficient of variation, skewness, autocorrelation, and cross correlation over a range of aggregation levels from 10 min upward, and were generally well fitted by the model. Validation tests, based on properties not used to fit the model, indicated a satisfactory model performance.

[34] The fitted model given in this paper, is intended for use in projects aimed at upgrading drainage systems in the Rome region. The results herein imply that the model will be of practical value in these forthcoming projects.

[35] **Acknowledgments.** Acea Spa are gratefully acknowledged for supplying the data that made this research possible. The Water Research Centre (WRc, United Kingdom) are thanked for coordinating the project. The associate editor and the reviewers are gratefully acknowledged for their helpful comments.

References

Akaike, H. (1974), A new look at statistical model indentification, *IEEE Trans. Autom. Control*, *19*, 716–722.

Burton, A., C. G. Kilsby, H. J. Fowler, P. S. P. Cowpertwait, and P. E. O'Connell (2008), RainSim: A spatial-temporal stochastic rainfall modelling system, *Environ. Modell. Software*, *23*, 1356–1369.

Cowpertwait, P. S. P. (1995), A generalized spatial-temporal model of rainfall based on a clustered point process, *Proc. R. Soc. London, Ser. A*, *450*, 163–175.

Cowpertwait, P. S. P. (1998), A Poisson-cluster model of rainfall: High order moments and extreme values, *Proc. R. Soc. London, Ser. A*, *454*, 885–898.

Cowpertwait, P. S. P. (2006), A spatial-temporal point process model for the Thames catchment, London, *J. Hydrol.*, *330*, 586–595.

Cowpertwait, P. S. P. (2010), A Neyman–Scott model with continuous distributions of storm types, *Aust. N. Z. Ind. Appl. Math. J.*, *51*, C97–C108.

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

Cowpertwait, P. S. P., J. Salinger, and B. Mullen (2009), A spatial-temporal stochastic rainfall model for Auckland City: Scenarios for current and future climates, *N. Z. J. Hydrol.*, *48*, 95–109.

Cox, D. R., and V. Isham (1988), A simple spatial temporal model of rainfall, *Proc. R. Soc. London, Ser. A*, *415*, 317–328.

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

Evin, G., and A.-C. Favre (2008), A new rainfall model based on the Neyman-Scott process using cubic copulas, *Water Resour. Res.*, *44*, W03433, doi:10.1029/2007WR006054.

Kakou, A. (1997), Point process based models for rainfall, Ph.D. thesis, Dep. of Stat. Sci., Univ. Coll. London, London.

Kim, S., and M. L. Kavvas (2006), Stochastic point rainfall modelling for correlated rain cell intensity and duration, *J. Hydrol. Eng.*, *11*, 29–36.

Leonard, M., M. F. Lambert, A. V. Metcalfe, and P. S. P. Cowpertwait (2008), A space-time Neyman-Scott rainfall model with defined storm extent, *Water Resour. Res.*, *44*, W09402, doi:10.1029/2007WR006110.

Mellor, D. (1996), The modified turning bands (MTB) model for space-time rainfall. I. Model definition and properties, *J. Hydrol.*, *175*, 113–127.

Northrop, P. (1998), A clustered spatial-temporal model of rainfall, *Proc. R. Soc. London, Ser. A*, *454*, 1875–1888.

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

Rodriguez-Iturbe, I., D. Cox, and V. Isham (1988), A point process model for rainfall: Further developments, *Proc. R. Soc. London, Ser. A*, *417*, 283–298.

Sharman, B., G. G. Paterson, R. Brown, M. Davis, P. Kinley, and S. Ranchhod (2006), The Integrated Catchment Study of Auckland City (New Zealand): Overview, paper presented at World Environmental and Water Resources Congress 2006, Am. Soc. of Civ. Eng., Omaha, Neb.

Waymire, E., V. Gupta, and I. Rodriguez-Iturbe (1984), A spectral theory of rainfall intensity at the meso-β scale, *Water Resour. Res.*, *20*, 1453–1465.

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

P. S. P. Cowpertwait, School of Computing and Mathematical Sciences, Auckland University of Technology, Wakefield Street, Auckland 1010, New Zealand. (paul.cowpertwait@aut.ac.nz)