A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes

P. S. P. Cowpertwait

Institute of Information and Mathematical Sciences, Massey University, Auckland, New Zealand

C. G. Kilsby and P. E. O'Connell

Water Resource Systems Research Laboratory, Department of Civil Engineering, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Received 5 June 2001; revised 7 February 2002; accepted 7 February 2002; published 3 August 2002.

[1] A spatial-temporal model of rainfall, based on a Neyman-Scott stochastic point process, is fitted to hourly data taken from nine sites in the Arno Basin, Italy. The stochastic model is an extension of the temporal Neyman-Scott rectangular pulses model into two-dimensional space and introduces a further parameter into the model. In the model, storms arrive in a Poisson process, where each storm consists of discs representing rain cells, with centers distributed over an area according to a spatial Poisson process. The cells have a random radius, lifetime, and intensity, with the intensity remaining constant over the area of the disc and cell lifetime. A fitting procedure is proposed which couples the results obtained in two preceding papers: the second-order properties of the spatial-temporal model and the third moment function of the single site model [*Cowpertwait*, 1995, 1998]. The model is validated by comparing extreme historical hourly data and equivalent data simulated using the fitted spatial-temporal model. These comparisons are made using a regional frequency analysis, based on L moments, and log-log plots of the upper distribution tail. The results indicate that the model is able to preserve regional extremes and support the use of the model in hydrological applications. *INDEX TERMS:* 1854 Hydrology: Precipitation (3354); 3354 Meteorology and Atmospheric Dynamics: Precipitation (1854); 1869 Hydrology: Stochastic processes; 1821 Hydrology: Floods; *KEYWORDS:* time series, point process, Poisson cluster model, L moments, regional frequency analysis

1. Introduction

[2] During the past decade, considerable research on the modeling of rainfall has been undertaken using stochastic point processes of the Neyman-Scott and Bartlett-Lewis types [see, e.g., *Islam et al.*, 1990; *Cowpertwait*, 1991, 1994; *Burlando and Rosso*, 1991; *Puente et al.*, 1993; *Onof and Wheater*, 1994; *Cowpertwait et al.*, 1996; *Cowpertwait and O'Connell*, 1997; *Verhoest et al.*, 1997]. Briefly, these models are based on a Poisson process of storm origins which have associated with them a random number of rectangular pulses ("rain cells"), with heights corresponding to rain intensity and widths to cell duration. Different cells and storms may overlap so that the total rain intensity at any time is the sum of the intensities of all cells active at that time (Figure 1).

[3] The Neyman-Scott and Bartlett-Lewis models differ in the displacement of cell origins relative to storm origins. In the Neyman-Scott model, the times between a storm origin and its associated cell origins are independent random variables, while in the Bartlett-Lewis model the times between adjacent cell origins are independent random variables. In the context of rainfall, it is usual to adopt an exponential distribution for the cell displacements for both models. There have been several empirical studies comparing the two models [e.g., *Islam et al.,* 1990; *Burlando and Rosso*, 1991; *Velghe et al.*, 1994]. In addition, *Cowpertwait* [1998] showed analytically that the two rainfall models were equivalent up to second-order properties.

[4] There has also been some work on extending the Neyman-Scott and Bartlett-Lewis models into two-dimensional space, to provide a framework for modeling multisite and spatial-temporal rainfall data. For example, *Cowpertwait* [1995] has derived properties for a spatial-temporal model based on the Neyman-Scott rectangular pulses (NSRP) model, while *Northrop* [1998] has developed a spatial-temporal model based on the Bartlett-Lewis process.

[5] The model presented here is the simplest stochastic extension of the NSRP model into two-dimensional space, which has not yet been assessed in previous work. Our objective is to propose a suitable fitting procedure and to assess model performance with respect to regional extreme values. The fitting procedure combines results in two previous papers [*Cowpertwait*, 1995, 1998] to produce an appropriate method when data are taken from more than one site across a region. A special case of the generalized spatial-temporal model [see *Cowpertwait* 1994, 1995] is used with the third moment function for the single site model derived by *Cowpertwait* [1998].

[6] The paper is organized as follows. In section 2, the spatial-temporal model is formulated and statistical properties of the model given. In section 3, the fitting procedure is given, which relates the statistical properties of the model to

Copyright 2002 by the American Geophysical Union.
0043-1397/02/2001WR000709

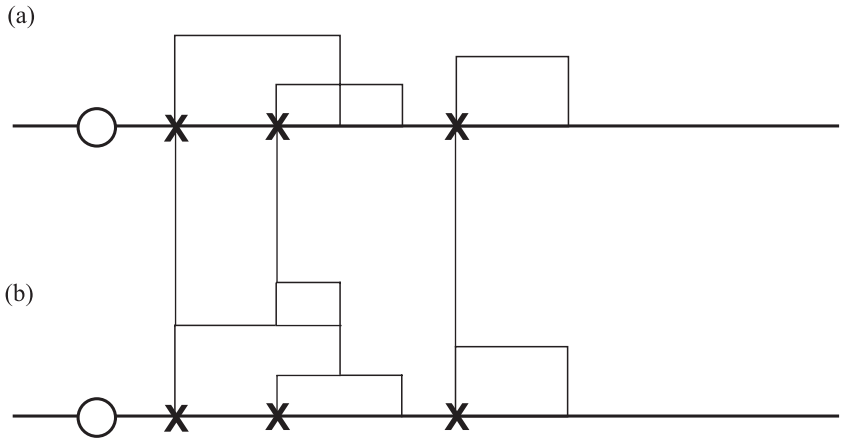

Figure 1. Temporal Neyman Scott rectangular pulses model: (a) a storm origin (circles) with a cluster of cells origins (crosses) and rectangular pulses (rain cells); (b) total rain intensity due to active rain cells.

the sample properties of the historical data. In section 4, the model is fitted to hourly data from nine sites in the Arno Basin, Italy, and an assessment of fit made with respect to regional extremes. Some overall conclusions and directions for further research are given in section 5.

2. Stochastic Model and Properties

2.1. Spatial-Temporal NSRP Model

[7] In the following formulation, the stochastic rainfall model is stationary in time and space, so that the resulting model parameters are constant. Consequently in fitting the model, the data has to be transformed to a stationary series or the model fitted to stationary subsets of the data (e.g., in the fitting procedure of section 3 the data are scaled by site means to achieve approximate spatial stationarity, and the model fitted to data from each calendar month, resulting in twelve estimates of each parameter, to account for the seasonal variation in rainfall).

[8] Let the arrival times $\{T_i\}$ of storms ('storm origins') occur in a Poisson process with rate λ (per hour), so that the times between adjacent storm origins are independent exponential random variables with mean λ^{-1}. A 'storm' consists of a stochastic process of rain cells $\{(U_{ij}, V_{ij}), S_{ij}, L_{ij}, X_{ij}, R_{ij}\}$ where for each storm i: (1) $\{(U_{ij}, V_{ij})\}$ forms a two-dimensional Poisson process with rate φ (per km^2). (2) (U_{ij}, V_{ij}) and R_{ij} form discs in two-dimensional space, where (U_{ij}, V_{ij}) is the disc centre and R_{ij} is a random variable representing the disc radius. (3) S_{ij} is the arrival time of cell j in storm i, where $S_{ij}-T_i$ are independent random variables, so that the arrival times $\{ S_{ij} \}$ form a Neyman-Scott point process. (4) L_{ij} is a random variable representing the cell lifetime, so that cell j in storm i terminates at time $S_{ij} + L_{ij}$. (5) The intensity X_{ij} of cell j in storm i is a random variable that remains constant throughout the cell lifetime and over the area of the disc (cells can thus be thought of as cylinders in three-dimensional space with heights given by the intensity X). (6) The total intensity at time t and location $\mathbf{x} = (x_1, x_2) \in R^2$, $Y(\mathbf{x}, t)$, is the sum of the intensities of all cells alive at time t and overlapping $\mathbf{x}$.

[9] Condition 6 can be expressed as follows. If $dN(u, t)$ is the number of cell centers at time t a distance u from an arbitrary point $\mathbf{x}$, and $Z_{u,s}(\mathbf{x}; t-s)$ is the intensity at location $\mathbf{x}$, time t, due to a cell with center a distance u from $\mathbf{x}$ and arrival time s, then the total intensity at location $\mathbf{x}$ and time t is given by:

$$Y(\mathbf{x};t) = \int_{u=0}^{\infty} \int_{s=0}^{\infty} Z_{u,s}(\mathbf{x};t-s)dN(u;t-s) \tag{1}$$

As rainfall data are usually sampled over discrete time intervals, it is necessary to consider the aggregated time series $\{Y_k^{(h)}(\mathbf{x})\}$, where

$$Y_k^{(h)}(\mathbf{x}) = \int_{(k-1)h}^{kh} Y(\mathbf{x};t)dt \tag{2}$$

so that $Y_k^{(h)}(\mathbf{x})$ is the rainfall depth in the kth time interval of duration h at location $\mathbf{x}$.

[10] To derive properties of $Y_k^{(h)}(\mathbf{x})$, and for model fitting and data simulation, it is necessary to make further assumptions. We will take R_{ij}, L_{ij}, and $S_{ij}-T_i$ to be independent exponential random variables with means ϕ^{-1}, η^{-1} and β^{-1} respectively. Under these assumptions, we find that, for each storm, the number of cells C that overlap a point in R^2 is a Poisson random variable with mean:

$$\mu_C = 2\pi\varphi/\phi^2 \tag{3}$$

[11] Without loss of generality, the subscripts i, j may be omitted, so that $L \equiv L_{ij}$, $X \equiv X_{ij}$, and $R \equiv R_{ij}$. Taking expectations in equation (1) gives

$$\begin{aligned} E\{Y(\mathbf{x};t)\} &= \int_{u=0}^{\infty} \int_{s=0}^{\infty} E\{Z_{u,s}(\mathbf{x};t-s)\}E\{dN(u;t-s)\} \\ &= \int_{u=0}^{\infty} \int_{s=0}^{\infty} E\{Z_{u,s}(\mathbf{x};t-s)\}2\pi\varphi du\lambda ds \\ &= \int_{u=0}^{\infty} \int_{s=0}^{\infty} E(X)e^{-\phi u}e^{-\eta s}2\pi\varphi du\lambda ds \\ &= 2\pi\varphi\mu_X\lambda/\phi^2\eta = \lambda\mu_X\mu_C/\eta \end{aligned} \tag{4}$$

which is the expected intensity for the single-site NSRP model. It can also be shown that other single-site NSRP properties are recovered [*Cowpertwait*, 1995, section 2].

[12] The cell intensity X will be taken as an independent Weibull random variable, with survivor function: $P(X > x) = \exp(-x^{1/\alpha}/\theta)$; the exponential distribution is then the special case $\alpha = 1$, which we make use of in the fitting procedure (section 3). Moments of this distribution are given by: $E(X^r) = \theta^{r\alpha}\,\Gamma(1 + r\alpha)$, so that $\mu_X = \theta^{\alpha}\,\Gamma(1 + \alpha)$ in equation (4).

[13] The model therefore has the following parameters: λ^{-1}, β^{-1}, η^{-1}, α, θ, μ_c, ϕ^{-1} (see also the notation). It is not necessary to include the spatial rate parameter φ because it is a function of μ_c and ϕ (equation (3)). The model is a special case (one "cell type") of the generalized model discussed by *Cowpertwait* [1994, 1995], and adds one further parameter ϕ into the NSRP model to account for spatial variability (over a homogeneous region). The model is therefore the simplest stochastic extension of the NSRP model into two-dimensional space.

2.2. Statistical Properties

[14] Mathematical properties of the spatial-temporal NSRP model have been derived by *Cowpertwait* [1995] and are given below:

$$\mu_h = E\{Y_k^{(h)}(\mathbf{x})\} = \lambda\mu_C\mu_X h/\eta, \tag{5}$$

$$\begin{aligned}\gamma_{x,y,h,l} &= Cov\{Y_k^{(h)}(\mathbf{x}), Y_{k+l}^{(h)}(\mathbf{y})\}\\ &= \gamma_{x,x,h,l} - \lambda\{1 - P(\phi, d)\}\mu_C E\big(X^2\big)A(h,l)/\eta^3,\end{aligned} \tag{6}$$

where:

$$\begin{aligned}\gamma_{x,x,h,l} = \gamma_{y,y,h,l} &= \lambda\eta^{-3}A(h,l)\\ &\cdot\{2\mu_C E\big(X^2\big) + \mu_X^2\beta^2 E\big(C^2 - C\big)/\big(\beta^2 - \eta^2\big)\}\\ &- \lambda\mu_X^2 B(h,l)E\big(C^2 - C\big)/\{\beta\big(\beta^2 - \eta^2\big)\};\end{aligned} \tag{7}$$

$$A(h,0) = \big(h\eta + e^{-\eta h} - 1\big); B(h,0) = \big(h\beta + e^{-\beta h} - 1\big);$$

$$A(h,l) = \frac{1}{2}\big(1 - e^{-\eta h}\big)^2 e^{-\eta h(l-1)};$$

$$B(h,l) = \frac{1}{2}\big(1 - e^{-\beta h}\big)^2 e^{-\beta h(l-1)} \text{ (for } l \text{ a positive integer)};$$

$d = \|\mathbf{x} - \mathbf{y}\|$, the distance between the points $\mathbf{x}$ and $\mathbf{y}$; and $P(\phi, d)$ is the probability that a cell overlaps a point $\mathbf{x}$ given that it overlapped a point $\mathbf{y}$ a distance d from $\mathbf{x}$. *Cowpertwait* [1995] gave an expression for $P(\phi, d)$ in terms of a Bessel function. However, an easier way to evaluate $P(\phi, d)$ is using the integral expression:

$$P(\phi, d) = \frac{2}{\pi}\int_0^{\pi/2}\left(\frac{\phi d}{2\cos y} + 1\right)\exp\left(\frac{-\phi d}{2\cos y}\right)dy \tag{8}$$

A good approximation to equation (8) can be obtained using Simpson's Rule with five ordinates and setting the last function value to zero, i.e., $P(\phi, d) \approx \frac{2}{\pi}\int_0^{\pi/2} f(y)dy$, where f is given by $f(y) = \left(\frac{\phi d}{2\cos y} + 1\right)\exp\left(\frac{-\phi d}{2\cos y}\right), 0 \le y < \pi/2, f(\pi/2) = 0$. Hence, from Simpson's rule,

$$P(\phi, d) \approx \frac{1}{30}\sum_{i=1}^{4}\left\{2f\left(\frac{2\pi i}{20}\right) + 4f\left(\frac{2\pi i + \pi}{20}\right)\right\} - \frac{1}{30}f(0). \tag{9}$$

Using equation (9) enables the cross correlation function equation (6) to be easily computed.

[15] The third moment function was derived by *Cowpertwait* [1998] and is given by

$$\begin{aligned}E\{Y_k^{(h)}(\mathbf{x}) - \mu_h\}^3 &= 6\lambda\mu_C E\big(X^3\big)\big(\eta h - 2 + \eta h e^{-\eta h} + 2e^{-\eta h}\big)/\eta^4\\ &+ 3\lambda\mu_X E\big(X^2\big)E\{C(C-1)\}f(\eta,\beta,h)\\ &/\{2\eta^4\beta\big(\beta^2 - \eta^2\big)^2\} + \lambda\mu_X^3\\ &\cdot E\{C(C-1)(C-2)\}g(\eta,\beta,h)\\ &/\{2\eta^4\beta\big(\eta^2 - \beta^2\big)(\eta - \beta)(2\beta + \eta)(\beta + 2\eta)\}\end{aligned} \tag{10}$$

where the functions $f(\eta, \beta, h)$ and $g(\eta, \beta, h)$ are given by:

$$\begin{aligned}f(\eta,\beta,h) = &-2\eta^3\beta^2 e^{-\eta h} - 2\eta^3\beta^2 e^{-\beta h} + \eta^2\beta^3 e^{-2\eta h} + 2\eta^4\beta e^{-\eta h}\\ &+ 2\eta^4\beta e^{-\beta h} + 2\eta^3\beta^2 e^{-(\eta+\beta)h} - 2\eta^4\beta e^{-(\eta+\beta)h} - 8\eta^3\beta^3 h\\ &+ 11\eta^2\beta^3 - 2\eta^4\beta + 2\eta^3\beta^2 + 4\eta\beta^5 h + 4\eta^5\beta h - 7\beta^5\\ &- 4\eta^5 + 8\beta^5 e^{-\eta h} - \beta^5 e^{-2\eta h} - 2h\eta^3\beta^3 e^{-\eta h}\\ &- 12\eta^2\beta^3 e^{-\eta h} + 2h\eta\beta^5 e^{-\eta h} + 4\eta^5 e^{-\beta h}\end{aligned} \tag{11}$$

$$\begin{aligned}g(\eta,\beta,h) = &\,12\eta^5\beta e^{-\beta h} + 9\eta^4\beta^2 + 12\eta\beta^5 e^{-\eta h} + 9\eta^2\beta^4\\ &+ 12\eta^3\beta^3 e^{-(\eta+\beta)h} - \eta^2\beta^4 e^{-2\eta h} - 12\eta^3\beta^3 e^{-\beta h} - 9\eta^5\beta\\ &- 9\eta\beta^5 - 3\eta\beta^5 e^{-2\eta h} - \eta^4\beta^2 e^{-2\beta h} - 12\eta^3\beta^3 e^{-\eta h}\\ &+ 6\eta^5\beta^2 h - 10\beta^4\eta^3 h + 6\beta^5\eta^2 h - 10\beta^3\eta^4 h + 4\beta^6\eta h\\ &- 8\beta^2\eta^4 e^{-\beta h} + 4\beta\eta^6 h + 12\beta^3\eta^3 - 8\beta^4\eta^2 e^{-\eta h} - 6\eta^6\\ &- 6\beta^6 - 2\eta^6 e^{-2\beta h} - 2\beta^6 e^{-2\eta h} + 8\eta^6 e^{-\beta h}\\ &+ 8\beta^6 e^{-\eta h} - 3\beta\eta^5 e^{-2\beta h}\end{aligned} \tag{12}$$

In the spatial-temporal NSRP model, C is a Poisson random variable, so that $E\{C(C-1)\} = \mu_C^2$ and $E\{C(C-1)(C-2)\} = \mu_C^3$ in equations (7) and (10).

[16] The following notation will be used for the variance, coefficient of variation, autocorrelation (at lag 1), cross correlation and skewness functions, respectively, where the dependent parameters are shown in brackets:

$$\begin{aligned}\sigma_h^2(\lambda,\beta,\eta,\mu_C,\alpha,\theta) &= \gamma_{x,x,h,0},\\ \upsilon_h(\lambda,\beta,\eta,\mu_C,\alpha) &= \sigma_h/\mu_h,\\ \rho_h(\beta,\eta,\mu_C,\alpha) &= \gamma_{x,x,h,1}/\sigma_h^2,\\ \rho_{x,y,h,l}(\beta,\eta,\mu_C,\alpha,\phi) &= \gamma_{x,y,h,l}/\sigma_h^2,\\ \kappa_h(\lambda,\beta,\eta,\mu_C,\alpha) &= E\Big(\{Y_k^{(h)}(\mathbf{x}) - \mu_h\}^3\Big)/\sigma_h^3.\end{aligned} \tag{13}$$

[17] Note that the autocorrelation, coefficient of variation, cross correlation, and skewness are dimensionless functions and do not depend on the scale parameter θ. Thus the parameters λ, β, η, μ_c, α, and ϕ can be estimated by fitting these functions to their equivalent dimensionless sample values which can be extracted using data from more than one site. The scale parameter θ can then be estimated for each site from the sample mean and equation (5), with the parameters in equation (5) replaced by their estimated values (see step 6 below equation (25)). The use of dimensionless functions allows the parameters λ, β, η, μ_c, α, and ϕ to be estimated for a region with the scale parameter θ determining site differences within the region.

3. Fitting Procedure

[18] The fitting procedure introduced here includes the following criteria, which can be applied when fitting similar stochastic spatial-temporal models to multisite data.

Table 1. Sites Used in Study

Site i	Name	x Coordinate, km	y Coordinate, km	Altitude, m	Annual Mean, mm	Proportion of Dry Days[a]	Hourly Maximum, mm
1	Arezzo	53	−16	277	806	0.66	38.7
2	Castelfiorentino	−23	−3	75	701	0.71	42.2
3	Empoli	−23	10	33	757	0.68	47.5
4	Firenze	1	18	51	790	0.70	52.0
5	Foiano	48	−40	315	714	0.68	42.2
6	Larderello	−27	−44	400	917	0.70	94.6
7	Livorno	−75	−10	3	778	0.74	120.2
8	Pisa	−67	7	6	896	0.70	150.4
9	Siena	9	−34	348	724	0.70	36.7

[a] Not significantly different using an ANOVA F test.

1. Regional dimensionless statistics are used to enable the model to be fitted to pooled data over a homogeneous region (where homogeneity within the region is with respect to the proportion of dry intervals).

2. Site differences are modeled using a scale parameter fitted to the sample means from each site.

3. Moments up to third order are used to fit the model to ensure a good representation of the distribution tail and extreme values.

4. Seasonal variation is accounted for by fitting the model to pooled data from each calendar month.

5. The fitting procedure is carried out in steps, which make use of special cases of the model. For example, the single-site model and the exponential distribution for cell intensity are special cases (see step 1 below), which can be fitted separately. In addition, carrying out the fitting procedure in steps (rather than using a single objective function) is more efficient as a smaller parameter space is searched in each step, which tends to results in fewer computations overall.

[19] The data from all available sites can be pooled to enable the dimensionless statistics (criterion 1 above) to be evaluated and used to fit the functions described in section 2. Fitting the special cases (criterion 5) helps to ensure a global optimum is found (because of the reduction of parameters in the parameter space). The steps in criterion 5 can also be tailored for the application. In this paper, an emphasis was made on getting an exact fit to the 1h sample moments (up to third order), which are important in hydrological catchment modeling.

[20] Now suppose that we have N years of hourly time series data at M sites. Let $x_{ijkl}^{(h)}$ be the observed rainfall depth sampled in a discrete time interval of width h (hours), at site i, year j, month k, and interval l, where: $i = 1, \ldots, M$; $j = 1, \ldots, N$; $k = 1, \ldots, 12$; $l = 1, \ldots, n(h, k)$, and $n(h, k)$ is the number of intervals of width h in month k.

[21] The model is stationary in both time and space, while multisite rainfall data are non-stationary in both domains. Seasonal variation is allowed for by fitting the model to each calendar month, as in previous work [e.g., *Cowpertwait et al.*, 1996]. The between-site variation is modeled by scaling the data by the mean rainfall, which is appropriate when the sites are contained in a homogeneous rainfall region (see section 4.1). Thus the sample mean is first calculated for each site-month:

$$\bar{x}_{ik}^{(1)} = \sum_{j=1}^{N} \sum_{l=1}^{n(l,k)} x_{ijkl}^{(1)} / Nn(1,k), \quad i = 1, \ldots, M; k = 1, \ldots, 12; \quad (14)$$

The following dimensionless (mean scaled) statistics are then found for each month k by pooling all available data (across sites and years):

$$\hat{\sigma}_{h,k}^2 = \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{l=1}^{n(h,k)} \left(x_{ijkl}^{(h)} / \bar{x}_{ik}^{(1)} - h \right)^2 / MNn(h,k) \quad (15)$$

$$\hat{\gamma}_{h,k} = \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{l=1}^{n(h,k)-1} \left(x_{ijkl}^{(h)} / \bar{x}_{ik}^{(1)} - h \right) \cdot \left(x_{ijkl+1}^{(1)} / \bar{x}_{ik}^{(1)} - h \right) / MN\{n(h,k) - 1\} \quad (16)$$

$$\hat{\upsilon}_{h,k} = \hat{\sigma}_{h,k} / h \quad (17)$$

$$\hat{\rho}_{h,k} = \hat{\gamma}_{h,k} / \hat{\sigma}_{h,k}^2 \quad (18)$$

$$\hat{\kappa}_{h,k} = \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{l=1}^{n(n,k)} \left(x_{ijkl}^{(h)} / \bar{x}_{jk}^{(1)} - h \right)^3 / \hat{\sigma}_{h,k}^3 MNn(h,k) \quad (19)$$

$$\hat{\rho}_{x,y,h,k} = \frac{\sum_{j=1}^{N} \sum_{l=1}^{n(h,k)} \left(x_{xjkl}^{(h)} - \bar{x}_{xk}^{(h)} \right) \left(x_{yjkl}^{(h)} - \bar{x}_{yk}^{(h)} \right)}{\sqrt{\sum_{j=1}^{N} \sum_{l=1}^{n(h,k)} \left(x_{xjkl}^{(h)} - \bar{x}_{xk}^{(h)} \right)^2 \sum_{j=1}^{N} \sum_{l=1}^{n(h,k)} \left(x_{yjkl}^{(h)} - \bar{x}_{yk}^{(h)} \right)^2}} \quad (20)$$

[22] The model is fitted using moments up to third order and the hourly cross-correlation function, using steps 1–6 outlined below. As a main interest in this paper was assessing goodness of fit to hourly extremes, it seemed essential to have exact fits to the sample 1 h coefficient of variation and skewness, because the tail of the distribution of 1 h rainfall is sensitive to small changes in these statistics. Consequently, steps 4 and 6 are designed to give exact fits to the 1 h sample moments up to third order; the remaining steps give close fits to other important time series properties. The steps are repeated for each calendar month (k), which gives twelve estimates of each parameter.

1. The single-site NSRP parameters (λ, β, η, μ_c) using exponential cell intensities ($\alpha = 1$) are estimated from hourly and daily coefficients of variation (υ) and autocorrelation (ρ) functions by minimizing the following sum of squares:

$$F_A = \left(1 - \upsilon_1 / \hat{\upsilon}_{1,k}\right)^2 + \left(1 - \rho_1 / \hat{\rho}_{1,k}\right)^2 + \left(1 - \upsilon_{24} / \hat{\upsilon}_{24,k}\right)^2 + \left(1 - \rho_{24} / \hat{\rho}_{24,k}\right)^2 \quad (21)$$

subject to λ, β, η, $\mu_c > 0$. This amounts to solving four simultaneous equations with four unknown parameters; an

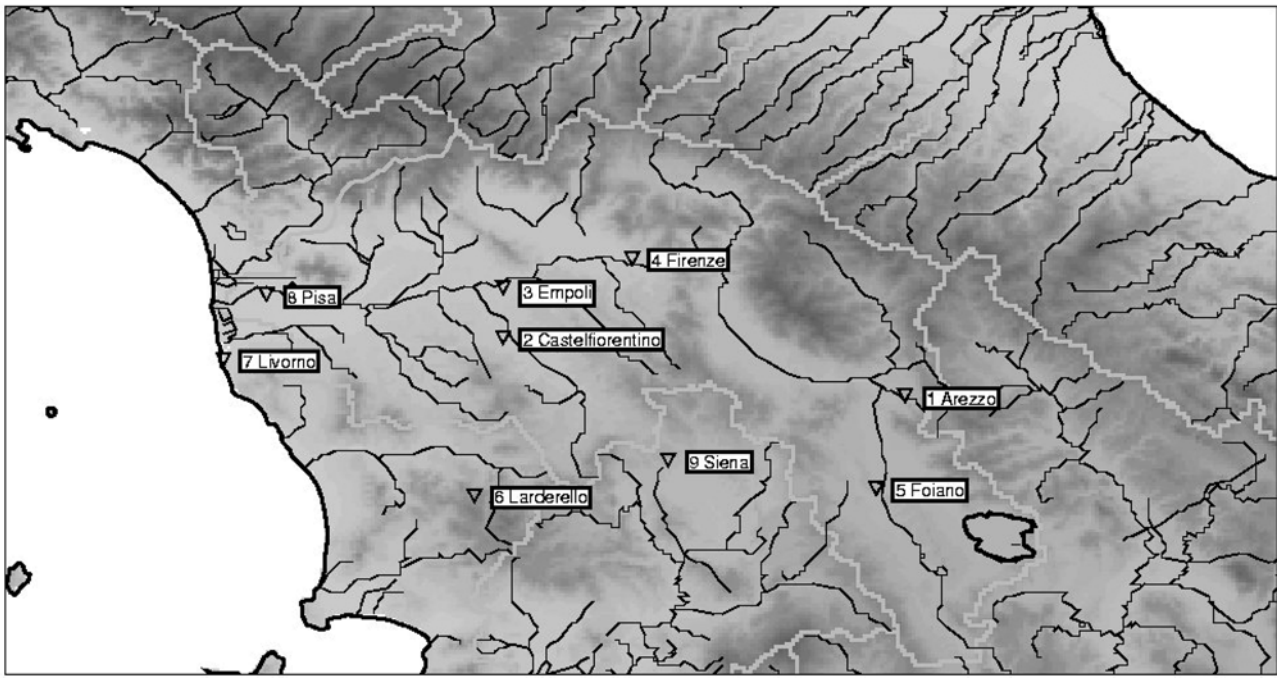

Figure 2. Map of Arno Basin and sites used in study.

exact fit, $F_A = 0$, is expected at the solution. The motivation behind this step is to obtain good starting values for the parameters: λ, β, and μ_c, and to estimate the cell lifetime parameter η, which is largely determined by the hourly autocorrelation at lag 1.

2. The hourly cross-correlation function is added to F_A to estimate the cell radius parameter ϕ. The cell dispersion parameter β is reestimated to allow for low correlations at sites separated by large spatial distances.

$$F_B = F_A + \sum_{x=1}^{M} \sum_{y=x+1}^{M} \left(1 - \rho_{x,y,1,k}/\hat{\rho}_{x,y,1,k}\right)^2 \tag{22}$$

An exact fit is not expected because there are more equations than unknown parameters. However, a near fit $F_B \approx 0$ is expected.

3. To compensate for lack of fit due to reestimating β in step 2, the parameters (λ, η, μ_c), with exponential cell intensities ($\alpha = 1$), are reestimated by minimizing F_A. In this step, there are three unknown parameters and four equations, so a close fit, $F_A \approx 0$, is expected at the solution.

4. The parameter α is estimated and μ_c reestimated from the hourly coefficient of variation (υ_1) and skewness (κ_1) by minimizing:

$$F_D = \left(1 - \upsilon_1/\hat{\upsilon}_{1,k}\right)^2 + \left(1 - \kappa_1/\hat{\kappa}_{1,k}\right)^2 \tag{23}$$

As there are two unknown parameters and two sample values, an exact fit, $F_D = 0$, is expected at the solution.

5. The cell radius parameter ϕ is reestimated by minimizing:

$$F_E = \sum_{x=1}^{M} \sum_{y=x+1}^{M} \left(1 - \rho_{x,y,1,k}/\hat{\rho}_{x,y,1,k}\right)^2 \tag{24}$$

A close fit, $F_E \approx 0$, is expected. The fitted moments obtained in the previous steps remain invariant in step 5. After this step, the following parameters estimates are obtained: $\hat{\lambda}_k, \hat{\beta}_k, \hat{\eta}_k, \hat{\alpha}_k, \hat{\mu}_{C_k}, \hat{\phi}_k$.

6. The scale parameter θ is estimated from:

$$\hat{\theta}_{ik} = \{\bar{x}_{ik}^{(1)} \hat{\eta}_k / \hat{\lambda}_k \Gamma(1 + \hat{\alpha}_k) \hat{\mu}_{Ck}\}^{1/\hat{\alpha}_k}, i = 1, \ldots, M, \tag{25}$$

using the estimates obtained in the previous steps and the hourly sample mean for each site. The fitted moments obtained in the previous steps remain invariant throughout step 6, because dimensionless functions are used in those steps. Step 6 provides an exact fit to the sample mean.

[23] The above steps are repeated for each calendar month, resulting in twelve estimates of each parameter: $\hat{\lambda}_k, \hat{\beta}_k, \hat{\eta}_k, \hat{\alpha}_k, \hat{\mu}_{C_k}, \hat{\phi}_k$, $k = 1, \ldots, 12$. In addition, this results in $12M$ estimates of the scale parameter θ: $\hat{\theta}_{ik}$, $i = 1, \ldots, M$; $k = 1, \ldots, 12$.

4. Empirical Data Analysis

4.1. Data Set

[24] Hourly data were available at nine sites in the Arno Basin, Italy, for the period: January 1962 to December 1985

Table 2. Parameter Estimates for Each Month

Month k	$\hat{\lambda}_k$, h^{-1}	$\hat{\mu}_{C_k}$	$\hat{\beta}_k$, h^{-1}	$\hat{\eta}_k$, h^{-1}	$\hat{\alpha}_k$	$\hat{\phi}_k$, km^{-1}
1	0.00252	24.7	0.00446	0.738	1.31	0.0446
2	0.00908	9.04	0.0510	0.929	1.30	0.0585
3	0.0101	14.0	0.0610	1.09	1.75	0.0603
4	0.00490	13.0	0.0140	1.20	1.26	0.0504
5	0.00884	7.75	0.0513	1.55	1.49	0.0639
6	0.00239	68.2	0.00888	2.10	2.87	0.105
7	0.00332	7.93	0.0185	2.63	1.86	0.0983
8	0.00722	5.12	0.0189	2.62	1.80	0.117
9	0.00428	10.6	0.0195	2.05	1.50	0.0902
10	0.00260	33.2	0.00906	1.42	1.85	0.0481
11	0.00359	31.4	0.00973	1.05	1.80	0.0511
12	0.00748	15.3	0.0439	0.871	1.74	0.0597

Table 3. Scale Parameter Estimate $\hat{\theta}_{ik}$ (mm) for Each Site-Month

Month k	Site i								
	1	2	3	4	5	6	7	8	9
1	0.0864	0.0781	0.0806	0.0818	0.0766	0.111	0.0902	0.106	0.0769
2	0.104	0.0904	0.0886	0.0933	0.0941	0.124	0.0920	0.103	0.0941
3	0.0917	0.0846	0.0893	0.102	0.0741	0.124	0.0966	0.106	0.0889
4	0.0874	0.0760	0.0927	0.0790	0.0748	0.0955	0.0673	0.0846	0.0853
5	0.0887	0.0685	0.0706	0.0864	0.0741	0.0881	0.0721	0.0824	0.0779
6	0.0743	0.0603	0.0652	0.0667	0.0603	0.0682	0.0620	0.0575	0.0646
7	0.0493	0.0354	0.0485	0.0431	0.0469	0.0436	0.0220	0.0294	0.0316
8	0.0813	0.0656	0.0745	0.0914	0.0721	0.0775	0.0641	0.100	0.0520
9	0.0998	0.0762	0.0966	0.0952	0.0954	0.0941	0.112	0.125	0.0860
10	0.105	0.0989	0.102	0.0944	0.101	0.139	0.146	0.150	0.106
11	0.125	0.123	0.137	0.150	0.123	0.172	0.144	0.166	0.145
12	0.112	0.106	0.0932	0.101	0.0882	0.122	0.100	0.118	0.0869

(Table 1 and Figure 2). An ANOVA F test was used on the proportion of dry days at each site (Table 1) to determine if the rainfall data scaled by the mean could be treated as approximately homogeneous over the region. The P value was 0.43, which is not significant, so the sites were kept in one region.

[25] Some of the sites had unusually high maximum rainfalls. For example, 150 mm of rain fell at Pisa on 20 August 1966 over a 1 h sampling interval (Table 1). It was tempting to discard these "outlying" observations, on the basis that they seemed almost physically impossible. However, they were checked against the record source, and also against neighboring daily data, and found to be reliable. Thus, they were retained in the fitting procedure, and we were left with the challenge of testing whether the stochastic spatial-temporal model could generate such unusually high values, while preserving essential statistical properties of the historical data.

4.2. Parameter Estimates and Fitted Values

[26] The parameters were estimated by following the fitting procedure in section 3, and are given in Tables 2 and 3. It can be seen that the estimates follow a seasonal variation characteristic of well-known rainfall patterns. For example, there are fewer storms (smaller λ), which are more isolated (lower ϕ^{-1}) and more intense (higher α), over the summer months, compared with winter months, which is characteristic of summer convective rainfall. Conversely, the smaller α and η over the winter months represent low-intensity long-duration rain cells, which are characteristic of frontal weather systems.

[27] The estimates of α and μ_c for June were notably higher than those of the other months (Table 2). This is likely to be due to June having the largest coefficient of skewness (Table 4) from which these parameters are estimated (in step 4 of the fitting procedure, equation (23)), and so the estimates were retained.

[28] The observed and fitted values are shown in Table 4, from which it is clear that a good fit is obtained to the sample values, with exact fits to the 1h sample moments. A slight underestimation of the daily coefficient of variation is obtained across the months, which is likely to lead to an underestimation of daily extremes.

[29] The fitted and sample cross correlations were plotted against distance and are shown in Figure 3. Figure 3a gives the mean cross correlations taken over the 12 months, while Figures 3b and 3c give the cross correlations for a representative winter and summer month (January and July respectively). In Figures 3b and 3c it can be seen that the summer cross correlations are generally smaller than the winter cross correlations, which is due to the more isolated nature of summer convective storms. The summer correlations

Table 4. Observed and Fitted Values[a]

Month k	υ_1		κ_1		ρ_1		υ_{24}		ρ_{24}	
	Observed	Fitted	Observed	Fitted	Observed	Fitted	Observed	Fitted	Observed	Fitted
1	5.19	5.19	10.0	10.0	0.600	0.643	2.48	2.09	0.183	0.207
2	5.09	5.09	9.64	9.64	0.588	0.608	2.32	2.14	0.205	0.218
3	5.26	5.26	12.3	12.3	0.546	0.564	2.31	2.12	0.185	0.197
4	5.99	5.99	11.8	11.8	0.488	0.507	2.44	2.14	0.215	0.237
5	7.13	7.13	16.0	16.0	0.430	0.439	2.62	2.44	0.161	0.170
6	11.1	11.1	54.4	54.4	0.319	0.325	3.62	3.21	0.140	0.154
7	15.9	15.9	47.9	47.9	0.259	0.261	4.54	4.27	0.101	0.107
8	12.9	12. 9	37.9	37.9	0.255	0.260	3.98	3.42	0.071	0.080
9	9.38	9.38	22.4	22.4	0.331	0.339	3.20	2.84	0.171	0.189
10	7.48	7.48	20.1	20.1	0.437	0.452	2.91	2.52	0.209	0.234
11	5.79	5.79	14.6	14.6	0.521	0.548	2.53	2.16	0.219	0.247
12	5.34	5. 34	12.2	12.2	0.605	0.623	2.44	2.28	0.220	0.231

[a] Values are dimensionless.

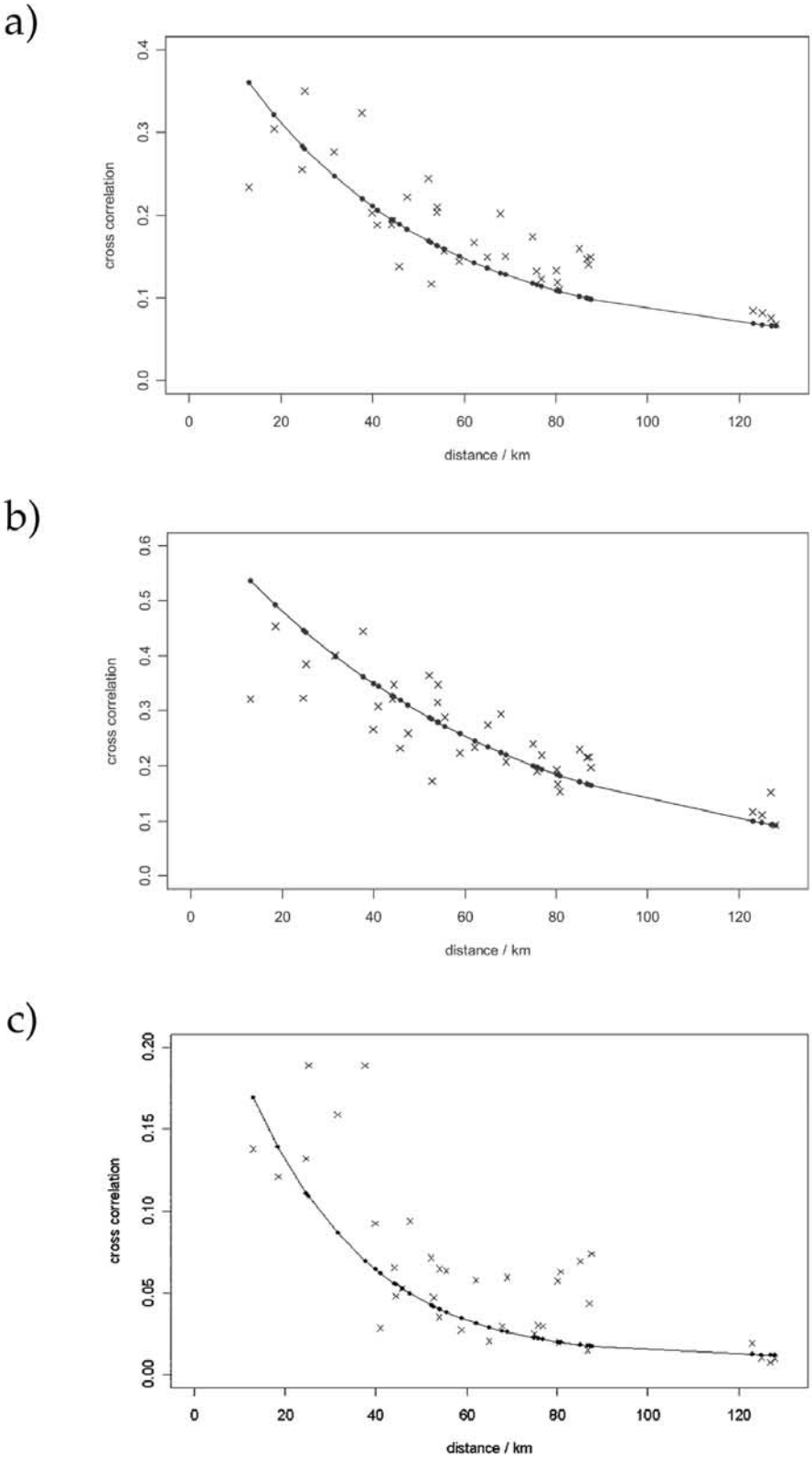

Figure 3. Hourly cross correlations plotted against spatial separation of sites (in km). The curved line is the fitted value under the model and the points (crosses) are the sample historical values. (a) Mean value taken over the months; (b) values for January; (c) values for July.

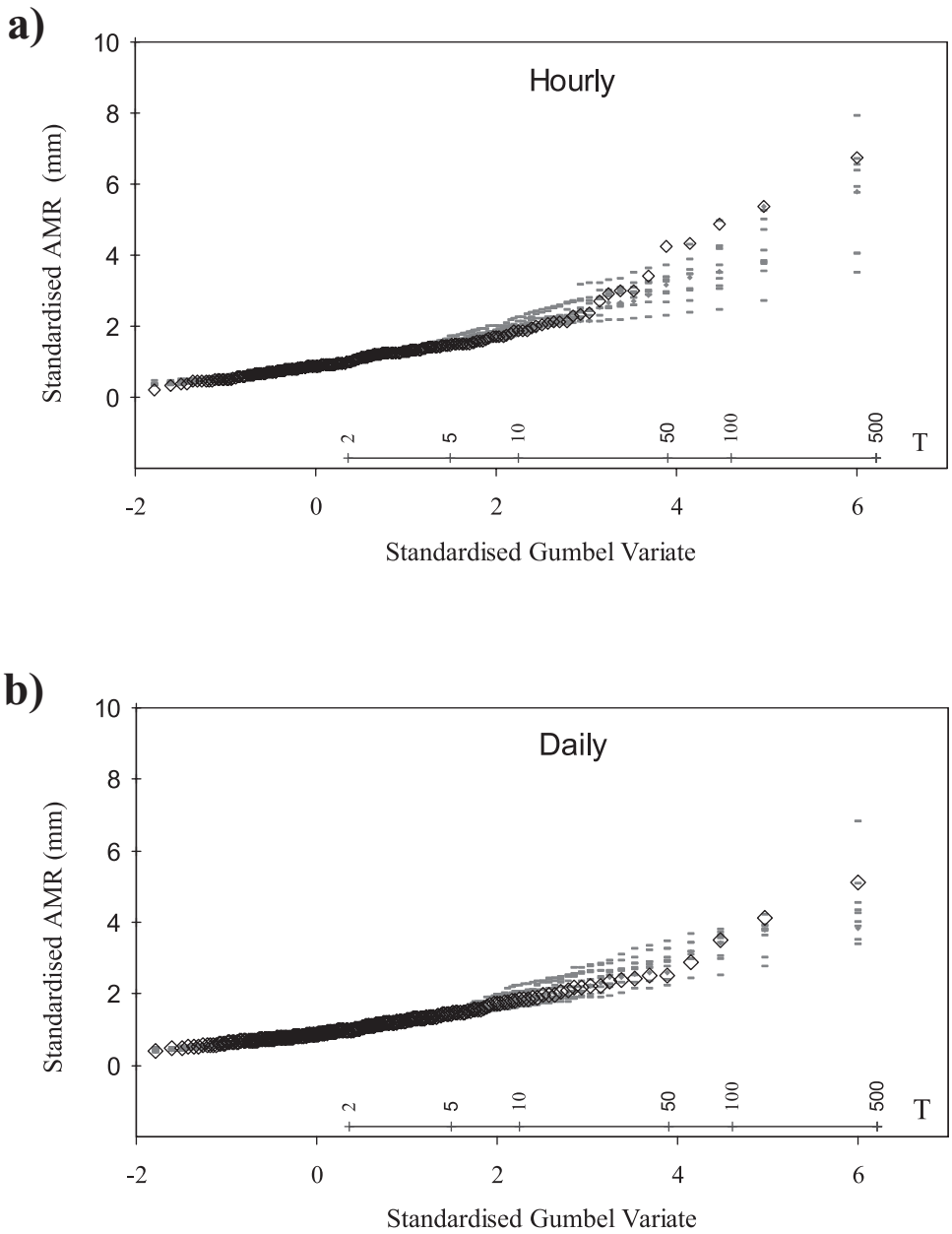

Figure 4. Standardized annual maximum rainfall frequency curves for the historical (diamonds) and simulated (dashes) values for (a) hourly and (b) daily time aggregation.

also tend to have greater variability, and there is a slightly poorer fit to these values (Figure 3c). However, overall the model has a good fit to the sample cross correlations (Figure 3a).

4.3. Regional Frequency Analysis

[30] To assess the model performance with respect to extreme values, twenty-five years of hourly data were simulated at the nine sites using the fitted spatial-temporal model. The simulation was repeated 10 times, from different starting points, so that bootstrap errors could be constructed for the model values.

[31] An analysis was performed according to the regional frequency analysis procedure developed by *Hosking and Wallis* [1997] and implemented in the Flood Estimation Handbook [*Institute of Hydrology*, 1999]. An identical procedure is followed for both the daily and hourly time aggregations. Because of the relatively short observed record (25 years), a regional analysis is preferred to at site analysis in order to provide robust estimates of the high return period rainfall quantiles.

[32] Firstly, for each of the nine 25-year historic series the annual maximum rainfall (AMR) events were extracted, ranked and standardized by dividing by the median AMR. The standardized AMR series were then pooled and ranked to produce a series of 225 values. This was then repeated for each of the 10 simulation runs, each containing 25-year series for each of the nine sites. This produced 10 pooled AMR series of 225 values. These historic and simulated values are plotted against a Gumbel reduced variate in Figure 4.

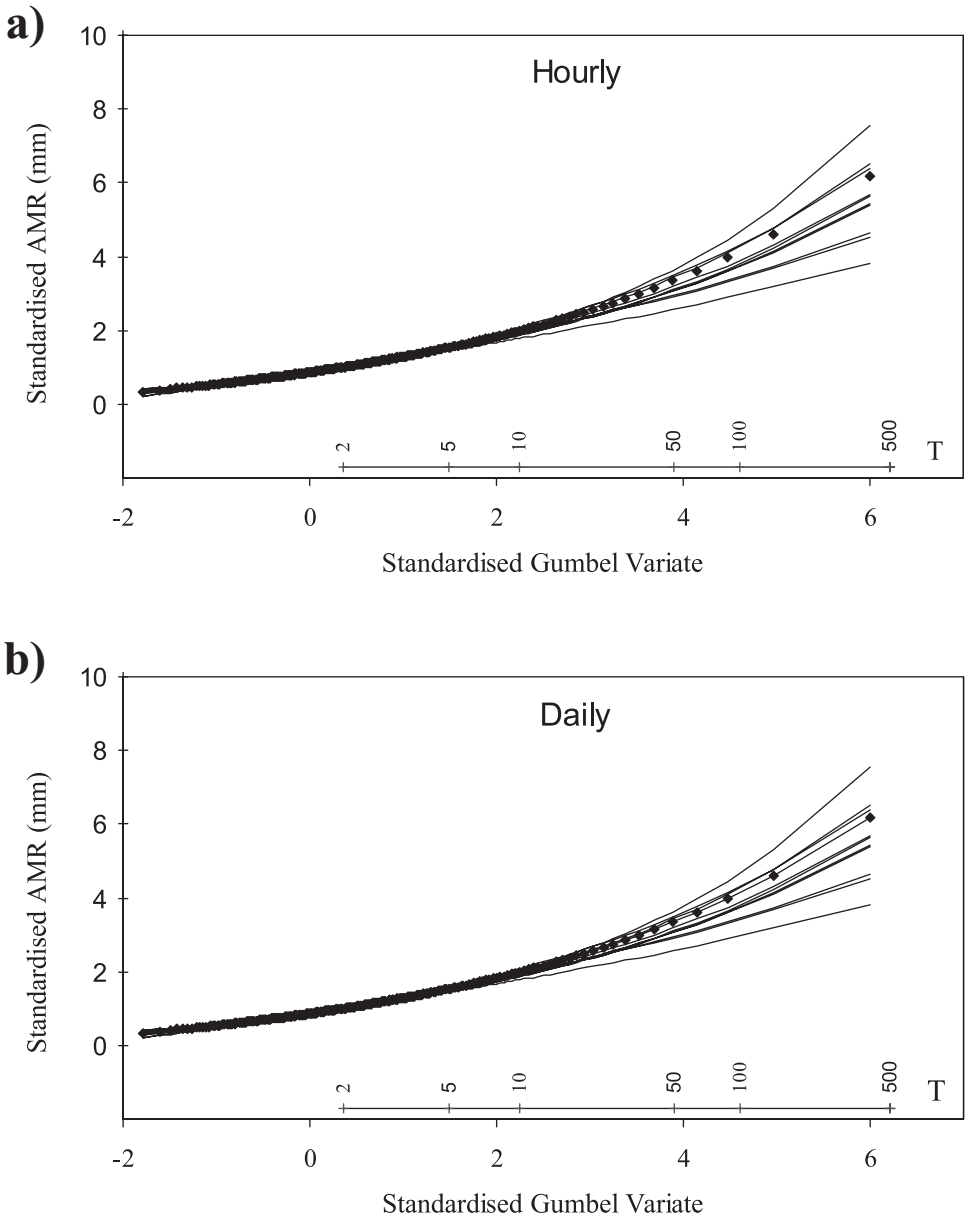

Figure 5. Fitted GEV curves to the standardized annual maximum rainfall values for the historical (solid diamonds) and simulated (lines) cases for (a) hourly and (b) daily time aggregation.

[33] A generalized extreme value (GEV) distribution was then fitted to each of the series, using the method of L moments [*Hosking and Wallis*, 1997]. The L moment fitting method has been shown to provide unbiased fitting for short sample records, and is affected to a lesser degree by the high return period values than other methods. These fitted curves are plotted in Figure 5.

[34] Figure 4 shows good agreement between the observed and simulated data points, except that the highest few standardized data points are at the limit of the ensemble for the hourly data. Figure 5 shows that the observed distribution falls well within the simulated range. This methodology using the NSRP has been successfully demonstrated for single site applications [*Hashemi et al.*, 2000]. The regional frequency analysis thus provides a good validation of the model and demonstrates that the NSRP model can be successfully extended to a multisite application where the fitting of regional extremes is important.

4.4. Log-Log Plots

[35] In the previous analysis results are derived based on annual maxima, which exclude a large number of observations. Thus it is appropriate to consider other approaches which use a larger proportion of the distribution tail. In the geophysical literature, there are numerous examples of physical processes having distribution tails that follow hyperbolic power laws (which are a feature of the ubiquitous concept of self-organized criticality [e.g., see *Andrade et al.*, 1998; *Bak*, 1996; *Bak and Tang*, 1989]). These models have: $N(x) \sim x^{-a}$, where $N(x)$ is the number of events with magnitude greater than (or equal) to x in a record of data, and a is a scaling exponent (which can be estimated by least squares regression after taking logs). It

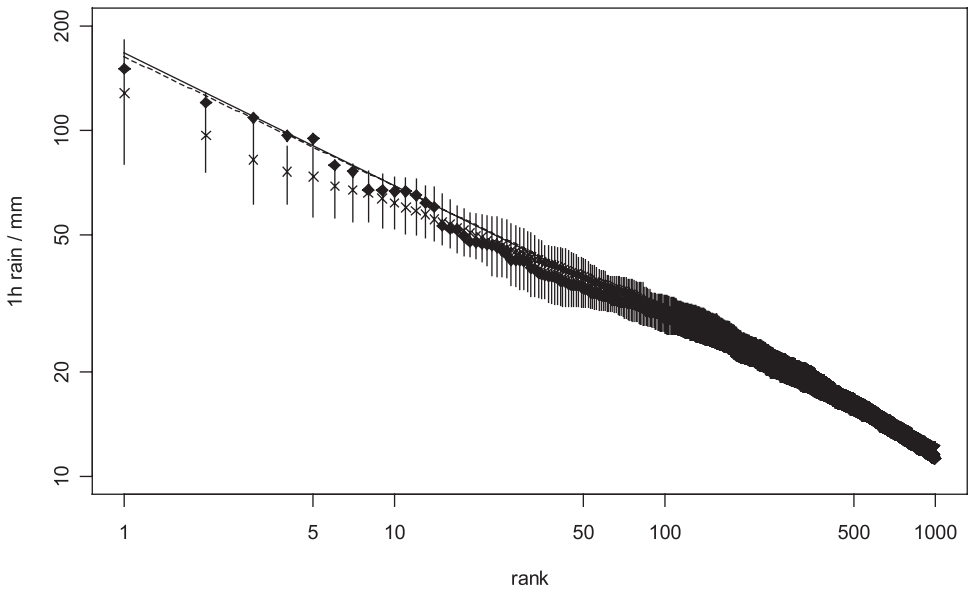

Figure 6. Log-log plot for the ordered hourly series. The error bar is the range of simulated values; the sample historical values are the points (solid diamonds) and the mean simulated values are the points (crosses). The straight lines (solid for historical and dotted for simulated) represent the hyperbolic distributions fitted using regression (see text).

thus seemed appropriate to consider log-log plots of the data.

[36] The historical hourly data for the nine sites were ordered and the largest one thousand values plotted on logarithmic scales (Figure 6). An approximate straight line is evident (shown) which implies a hyperbolic-type distribution for the upper tail. The largest one thousand values for each of the 10 simulation runs were also ranked, and the mean, maximum and minimum of the 10 simulated values found for each rank. The means were plotted and a line drawn between the maximum and minimums to represent the range of simulated values. Figure 6 shows that the simulated data are in good agreement with the historical values, except for some small departures around rank 5. The simulated data also show slightly more curvature than the historical values in the far tail of the distribution (Figure 6). The range of 10 simulations is generally less than two standard deviations, and the outliers and fitted lines lie

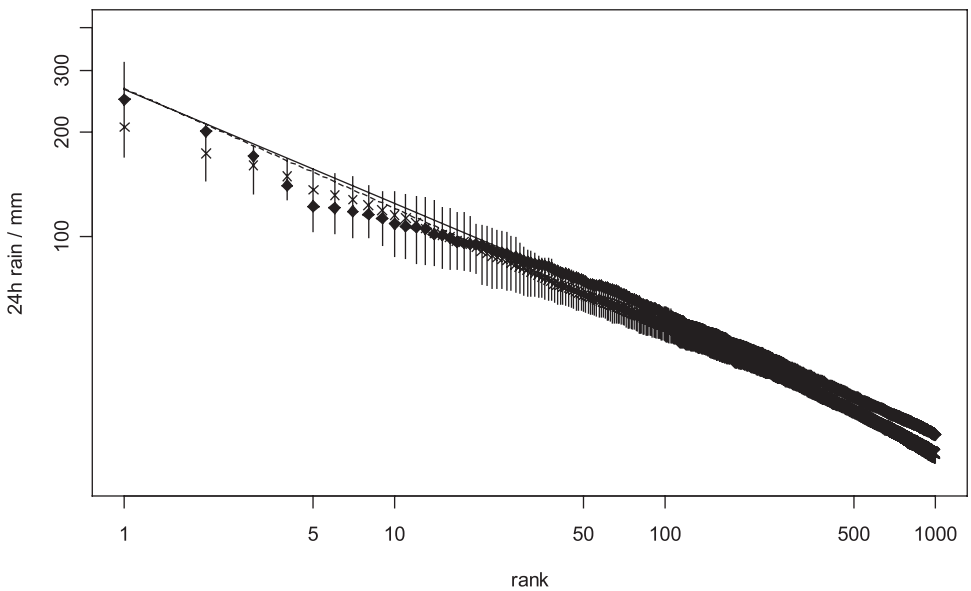

Figure 7. Log-log plot for the ordered daily series. The error bar is range of simulated values; the sample historical values are the points (solid diamonds) and the mean simulated values are the points (crosses). The straight lines (solid for historical and dotted for simulated) represent the hyperbolic distributions fitted using regression (see text).

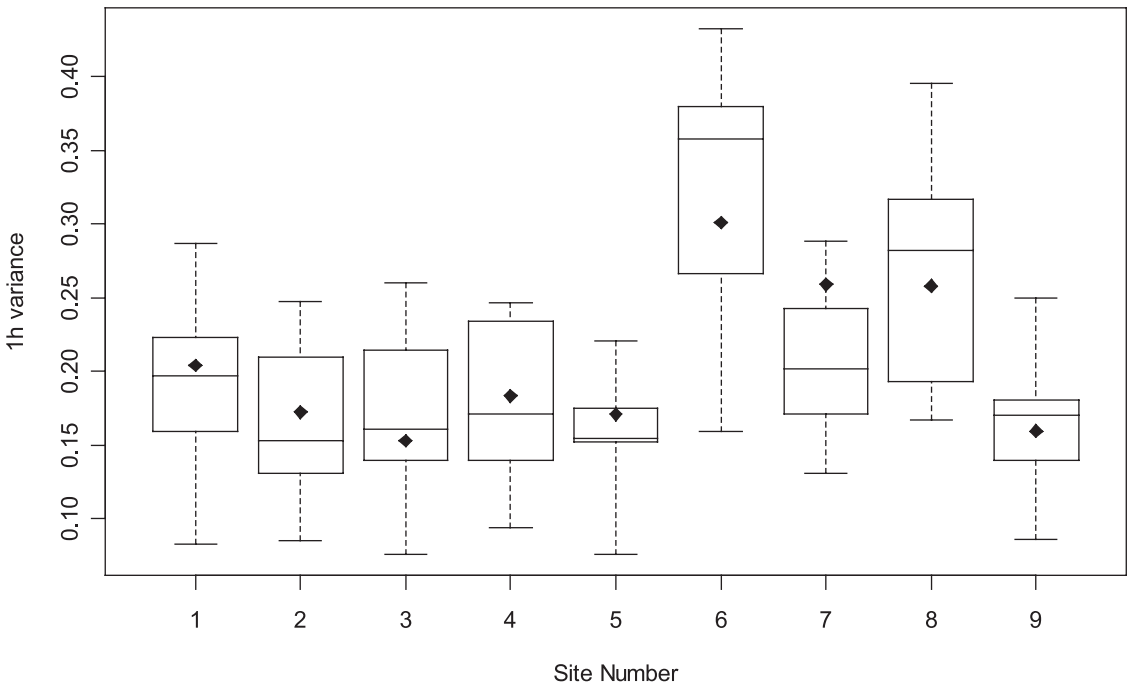

Figure 8. Box plot of the 10 simulated 1h variances (mm^2) for January. The historical value is shown as solid diamonds.

within two standard deviation confidence intervals (not shown).

[37] For both the simulated and historical series, weighted least squares regression was used to estimate the scaling exponent a, where the weights were equal to the rank of the observation (which is proportional to the variance of the observation). Both of the fitted regression models had a percentage of explained variance R^2 greater than 99%, which indicates that the linear regression model is a good approximation over the range of values fitted, and provides good evidence that the distribution tail follows a power law. Thus for the historical series, $N(x) \sim x^{-2.35}$, while for the simulated series $N(x) \sim x^{-2.42}$, so that both estimates of a are equal to two significant figures.

[38] Equivalent log-log plots were also found for the aggregated daily series, where a good fit is evident between

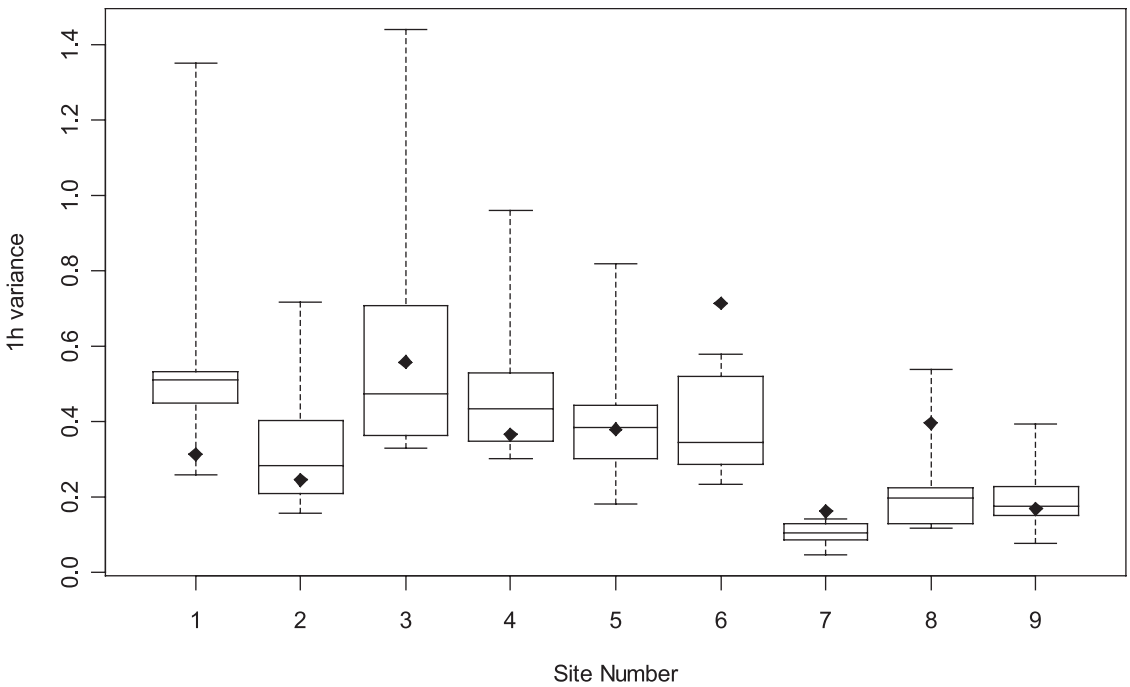

Figure 9. Box plot of the 10 simulated 1h variances (mm^2) for July. The historical value is shown as solid diamonds.

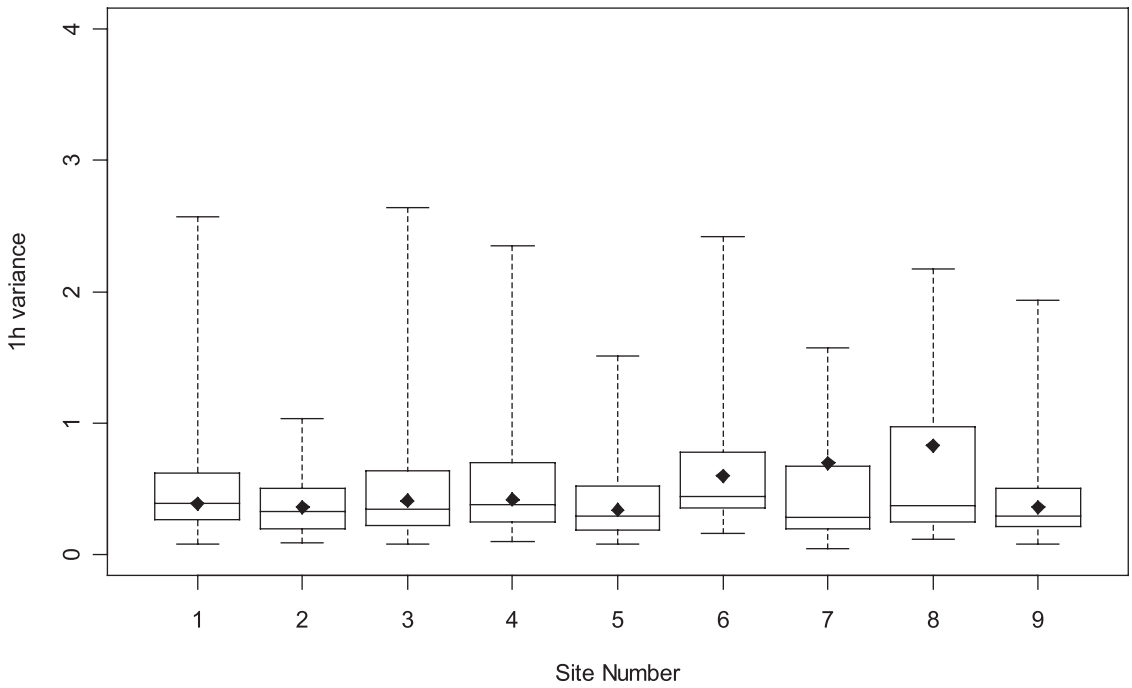

Figure 10. Box plot of the 10 simulated 1h variances (mm^2) for all months. The historical value is the mean over all 12 months and is shown as solid diamonds.

the historical and simulated values (Figure 7). Weighted least squares regression was used to estimate the scaling exponent a for the historical and simulated daily series. Both of the fitted regression models had a percentage of explained variance R^2 greater than 99%, which indicates that the linear regression is a good approximation over this range of values, and that the distribution tail follows a power law. Furthermore, the fitted lines shown in Figures 6 and 7 generally lie within the range of simulated values in the upper distribution tail. For the historical daily series, $N(x) \sim x^{-2.85}$, while for the simulated series $N(x) \sim x^{-2.66}$, so that the estimates of a are in reasonable agreement and generated daily extremes behave in a similar way to the observed values. Some discrepancy in the magnitudes of the smaller extremes can be seen in the Figure 7, which is likely to be due to the underestimation of the daily coefficient of variation (Table 4).

4.5. Box Plots to Further Assess the Spatial Homogeneity Assumption

[39] Spatial homogeneity (or stationarity) is assumed in the formulation of the stochastic spatial-temporal model, with site differences being taking into account using the scale parameter θ. An ANOVA F test showed that the proportion of dry days was not significantly different between the sites, providing evidence of approximate spatial homogeneity with respect to dry periods. The assumption of spatial homogeneity can be further assessed using the simulated data, and comparing variances to assess spatial homogeneity to second-order level.

[40] Thus the hourly variances were found for each site for each of the 10 simulations. The variances were found for all months combined (Figure 10) and also for January and July separately (Figures 8 and 9). Box plots were constructed for each site based on the simulations and are shown in Figures 8–10, together with the equivalent statistic taken from the historical records (in the case of Figure 10 the historical value for each site was the mean taken over all months). In these box plots the "whiskers" extend from the lower and upper quartiles to the minimum and maximum values, respectively, and the line near the middle of the box shows the median of the 10 simulations.

[41] Figure 8 indicates that for January spatial homogeneity can largely be taken into account via the scale parameter θ. Figure 9 shows that some discrepancies can be found for July. For example, the variances at sites 6 and 7 are underestimated in all 10 simulations, so that a more detailed at-site analysis (rather than a regional analysis) might show some differences in the site extremes. Figure 10 gives a broad indication that the method of incorporating site differences via the scale parameter is reasonable for this region, as all historical values fall within the interquartile range of the simulated values (except the historical value for site 7, which fell marginally above the upper quartile). Overall, Figures 8–10 support the use of the fitting methodology, especially as a relatively simple model has been fitted to a large catchment with highly variable topography (containing both coastal and high altitude sites).

5. Discussion and Conclusions

[42] The temporal NSRP model readily extends into two-dimensional space to form a spatial-temporal rainfall model with an underlying temporal cluster process. The spatial development effectively introduces one further parameter into the NSRP model, which represents cell radius, and is therefore the simplest stochastic extension of the NSRP model into two-dimensional space.

[43] Statistical properties for the spatial-temporal model [*Cowpertwait*, 1995, 1998] were combined into a fitting

procedure, which used moments up to third order and the cross correlation function. The fitting procedure was carried out in steps, which involved fitting dimensionless properties to pooled sample estimates and applying special cases of the model to reduce the parameter space at each step. In addition, the fitting procedure gave some emphasis to fitting the distribution tail using the 1 h skewness. Such a procedure could be adopted by other hydrologists when fitting similar models over a region, although this would require first evaluating the third moment function (as described by *Cowpertwait* [1998]).

[44] The fitting procedure was applied to data from nine sites in the Arno Basin, Italy. An exact fit was obtained to sample moments up to third order at the hourly level of aggregation, and a good fit obtained to other sample properties used in the fitting procedure.

[45] The regional frequency analysis provided good evidence that the spatial-temporal model is able to represent extreme values over a region. There was also evidence that both the historical data and data simulated by the model had a distribution tail that followed a power law (which is consistent with the concept of self-organized criticality). Overall the fit to the extremes was very good given that they had not been used in the model fitting procedure.

[46] There are many other possible comparisons that could be made, and it is inevitable that such an idealized model will break down, or reveal lack of fit, at certain time or spatial scales, or for particular applications. For example, the model could fail to represent rainfall for larger catchments (or smaller, less homogeneous, catchments), in which case it may be better to fit the model to data from subcatchments within the region of interest. Alternatively, the generalized Neyman-Scott model [*Cowpertwait*, 1995] could be used with different cell types for different homogeneous subcatchments.

[47] Some further tests, relevant to the intended application, could be made before the model is applied in practice. As the model was able to provide a good fit to the combined extreme values across the Arno catchment, it would be reasonable to hypothesise the use of the model for basin flood frequency analysis for this catchment and others. In this case, it would be appropriate to compare historical flow data with flows generated by routing rainfall data simulated by the model through a rainfall-runoff model. An empirical analysis of the flows should reveal whether the model is then appropriate for use in flood frequency analysis across a catchment.

[48] The model could be used for the simulation of rainfall at ungauged sites or over a whole region. This could be achieved by developing regression equations for predicting the scale parameter from site geographic variables (e.g., altitude). The scale parameter could then be predicted at any node in a grid of points over a region, and the fitted model used to simulate rainfall over the whole region.

[49] Overall, the results in this paper support the use of the spatial-temporal model for applications in hydrology.

Notation

λ^{-1} mean time (in units of hours) between two adjacent storm origins.

β^{-1} mean time (h) between storm origins and cell starting times.

η^{-1} mean lifetime (h) of a cell.

α power parameter for cell intensity X.

θ scale parameter (mm/h) for cell intensity X.

μ_c mean number of cells per storm per site.

ϕ^{-1} mean cell radius (km).

φ^{-1} mean number of cell origins per unit area.

σ_h^2 variance of rainfall depths taken over time intervals of width h.

υ_h coefficient of variation taken over time intervals of width h.

ρ_h temporal autocorrelation at lag 1 taken over time intervals of width h.

κ_h coefficient of skewness taken over time intervals of width h.

$\rho_{x,y,h,l}$ Lag l correlation of rainfall depths between sites at locations x and y taken over time intervals of width h.

[50] **Acknowledgments.** The authors gratefully acknowledge the Laboratory of Scientific and Applied Hydrology and Meteorology of the Politecnico di Milano for providing the data. The referees and associate editor are gratefully acknowledged for their comments on a previous draft. Paul Cowpertwait is grateful to the Department of Civil Engineering, University of Newcastle upon Tyne, for funding an overseas visit which made this work possible.

References

Andrade, R. F. S., H. J. Schellnhuber, and M. Claussen, Analysis of rainfall records: Possible relation to self-organised criticality, *Physica A*, *254*, 557–568, 1998.

Bak, P., *How Nature Works*, Springer-Verlag, New York, 1996.

Bak, P., and C. Tang, Earthquakes as a self-organized critical phenomenon, *J. Geophys. Res.*, *94*, 15,635–15,637, 1989.

Burlando, P., and R. Rosso, Comment on "Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall" by S. Islam et al., *J. Geophys. Res.*, *96*, 9391–9395, 1991.

Cowpertwait, P. S. P., Further developments of the Neyman-Scott clustered point process for modeling rainfall, *Water Resour. Res.*, *27*, 1431–1438, 1991.

Cowpertwait, P. S. P., A generalized point process model for rainfall, *Proc. R. Soc. London, Ser. A*, *447*, 23–37, 1994.

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

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

Cowpertwait, P. S. P., and P. E. O'Connell, A regionalised Neyman-Scott model of rainfall with convective and stratiform cells, *Hydrol. Earth Syst. Sci.*, *1*, 71–80, 1997.

Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metcalfe, and J. Mawdsley, Stochastic point process modelling of rainfall, *J. Hydrol.*, *175*, 17–65, 1996.

Hashemi, A. M., M. Franchini, and P. E. O'Connell, Climatic and basin factors affecting the flood frequency curve, part 1, A simple sensitivity analysis based on the continuous simulation approach, *Hydrol. Earth Syst. Sci.*, *4*, 463–482, 2000.

Hosking, J. R. M., and J. R. Wallis, *Regional Frequency Analysis: An Approach Based on L-Moments*, 224 pp., Cambridge Univ. Press, New York, 1997.

Institute of Hydrology, *Flood Estimation Handbook*, Wallingford, UK, 1999.

Islam, S., D. Entekhabi, R. L. Bras, and I. Rodriguez-Iturbe, Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall, *J. Geophys. Res.*, *95*, 2093–2100, 1990.

Onof, C., and H. Wheater, Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model, *J. Hydrol.*, *157*, 177–195, 1994.

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

Puente, C. E., M. F. P. Bierkens, M. A. Diaz-Granados, P. E. Dik, and M. M. Lopez, Practical use of analytically derived runoff models based on rainfall point processes, *Water Resour. Res.*, *29*, 3551–3560, 1993.

Velghe, T., P. A. Troch, F. P. De Troch, and J. Van de Velde, Evaluation of cluster-based rectangular pulses point process models for rainfall, *Water Resour. Res.*, *30*, 2847–2857, 1994.

Verhoest, N., P. Troch, and F. P. De Troch, On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point, *J. Hydrol.*, *202*, 108–120, 1997.

P. S. P. Cowpertwait, Institute of Information and Mathematical Sciences, Massey University, Private Bag 102 904, North Shore MSC, Auckland, New Zealand. (p.s.cowpertwait@massey.ac.nz)

C. G. Kilsby and P. E. O'Connell, Water Resource Systems Research Laboratory, Department of Civil Engineering, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.