

A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts

A. Burton,¹ H. J. Fowler,¹ C. G. Kilsby,¹ and P. E. O'Connell¹

Received 9 November 2009; revised 3 June 2010; accepted 15 June 2010; published 2 November 2010.

[1] The nonhomogeneous spatial activation of raincells (NSAR) model is presented which provides a continuous spatial-temporal stochastic simulation of rainfall exhibiting spatial nonstationarity in both amounts and occurrence. Spatial nonstationarity of simulated rainfall is important for hydrological modeling of mountainous catchments where orographic effects on precipitation are strong. Such simulated rainfall fields support the current trend toward distributed hydrological modeling. The NSAR model extends the Spatial Temporal Neyman-Scott Rectangular Pulses (STNSRP) model, which has a homogeneous occurrence process, by generating raincells with a spatially nonhomogeneous Poisson process. An algorithm to simulate nonhomogeneous raincell occurrence is devised. This utilizes a new efficient and accurate algorithm to simulate raincells from an infinite 2-D Poisson process, in which only raincells relevant to the application are simulated. A 4009 km² Pyrenean catchment exhibiting extreme orographic effects provides a suitable case study comprising seven daily rain gauge records with hourly properties estimated using regional downscaling relationships. Both the NSAR and the STNSRP models are fitted to five calibration rain gauges. Simulated hourly fields are validated using the remaining two rain gauges providing the first validation of time series sampled from STNSRP or NSAR processes at locations not used in model fitting. The NSAR model exhibits considerable improvement over the STNSRP model particularly with respect to nonhomogeneous rainfall occurrence at both daily and hourly resolutions. Further, the NSAR simulation provides an excellent match to the spatially nonhomogeneous observed daily mean, proportion dry, variance, coefficient of variation, autocorrelation, skewness coefficient, cross correlation and extremes, and to the hourly proportion dry and variance properties.

Citation: Burton, A., H. J. Fowler, C. G. Kilsby, and P. E. O'Connell (2010), A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts, *Water Resour. Res.*, 46, W11501, doi:10.1029/2009WR008884.

1. Introduction

[2] As catchment size increases, both rainfall amounts and occurrence become more spatially variable until modeling assumptions of spatial homogeneity of rainfall become unsuitable for hydrological applications [e.g., Fowler *et al.*, 2005]. This increased variability is often due to the interaction of mountainous terrain with the atmosphere, creating orographic effects, which typically include increased rainfall amounts and occurrence on the windward side of mountains and reduced amounts and occurrence downwind [e.g., see Hill *et al.*, 1981; Wheeler *et al.*, 2000; Barros and Lettenmaier, 1993, 1994; Roe, 2005].

[3] Stochastic models of rainfall in mountainous catchments are needed for hydrology, e.g., Barros and Lettenmaier [1993] report that 70% of annual runoff in the western

United States is disproportionately controlled by the duration and distribution of high elevation precipitation. In addition to amounts, the simulation of realistic rainfall occurrence is important because runoff generation is nonlinear (i.e., the same rainfall depth in fewer days produces higher runoff) and when conditioning a stochastic weather generator [e.g., Wilks and Wilby, 1999; Kilsby *et al.*, 2007] occurrence is crucial as wet days are cooler and have lower potential evapotranspiration than dry days.

[4] A number of multisite methodologies exist that can simultaneously simulate differing expected amounts and occurrence at a finite number of rain gauges: multisite Markov chains [e.g., Wilks, 1998; Mehrotra and Sharma, 2007; Srikanthan and Pegram, 2009], multivariate autoregressive approaches [e.g., Bárdossy and Plate, 1992; Stehlik and Bárdossy, 2002], generalized linear models (GLMs) [e.g., Chandler and Wheeler, 2002; Segond *et al.*, 2006], resampling from historical records using *K*-nearest neighbors (KNN) [Buishand and Brandsma, 2001], GLM plus KNN [Mezghani and Hingray, 2009] and best match to rainfall probability and mean rainfall amount estimated using regression of atmospheric state variables [Wilby *et al.*, 2003]. However, Markov chain and autoregressive models may have a large number of parameters and the resampling

¹Water Resource Systems Research Laboratory, School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne, UK.

schemes cannot generate spatial rainfall patterns that have not been observed in the historical record. Further, these methodologies typically generate time series with daily time steps and at a finite number of locations (rather than generating spatial fields). Exceptions include the potential to spatially interpolate the GLM methodologies (with a potential loss of spatial heterogeneity) and the *Mezghani and Hingray* [2009] resampling of daily sets of hourly data, albeit with an underestimate of autocorrelation (at 6 h daily aggregation levels) and a limitation of repeating sub-daily multisite patterns seen in the observed record.

[5] However, the current trend toward distributed hydrological modeling requires continuous spatial rainfall fields to capture heterogeneities and subdaily time steps for rapid processes. Spatial extensions to point process stochastic rainfall models such as the BLRP or NSRP models [e.g., *Cowpertwait*, 1995; *Onof et al.*, 2000] provide the requisite continuous temporal and spatial simulation processes that can be sampled at arbitrary temporal aggregations (e.g., hourly), spatial integrations (e.g., for the grid squares of a distributed model), or spatial locations (e.g., at rain gauge network locations or grid nodes). Additionally, such approaches can be easily reparameterized and used to downscale climate change scenarios [e.g., *Kilsby et al.*, 2007].

[6] The Spatial Temporal Neyman-Scott Rectangular Pulses (STNSRP) model [e.g., *Burton et al.*, 2008] was formulated analytically by *Cowpertwait* [1995], extending the single-site NSRP process [*Rodriguez-Iturbe et al.*, 1987]. This model has a spatially heterogeneous amounts process to account for orographic enhancement but is limited to a homogeneous occurrence process. The STNSRP model has been demonstrated in a number of practical applications [e.g., *Cowpertwait et al.*, 2002; *Fowler et al.*, 2005; *Burton et al.*, 2008]. However, for a downscaling application to a 15,000 km² region in Yorkshire, UK, *Fowler et al.* [2005] found it necessary to develop two separate subregional rainfall models to reduce the modeling errors arising from the assumption of spatial homogeneity in rainfall occurrence.

[7] Orographic effects lead to an increase in the number and duration of precipitation events [*Barros and Lettenmaier*, 1994] and rainfall in mountains is associated with the passage of preexisting precipitation systems, some of which may be too slight to be recorded by upwind rain gauges [*Hill et al.*, 1981]. These observations suggest two possible modifications of the spatial BLRP or spatial NSRP models to account for nonhomogeneous orographic effects: (1) modifying the storm incidence process; (2) modifying the number, properties or behavior of the raincells. The former may be more appropriate where the type of event giving rise to rainfall varies spatially, as may happen over a large region, and the latter when large scale events trigger local rainfall in accordance with the physical geographical setting.

[8] Following the latter alternative *Cowpertwait* [1995] suggested a possible analytical extension to the STNSRP model to represent spatially heterogeneous rainfall occurrence for a multisite application. Each sample location was considered to have an associated survival probability which modulated a homogeneous STNSRP raincell generation process. A raincell may therefore be observed at some locations and not at others across its disc. While such an approach is analytically appealing, no application of such a model has been demonstrated and it is not clear if the pro-

cedure can be extended to a spatial process. An approach based on the single-site NSRP process was however demonstrated by *Favre et al.* [2002]. Simulated storms followed a master Poisson process, but raincells were generated at two rain gauges using bivariate distributions of number, intensity, and duration. Fitting yielded models with different expected numbers of raincells and the ability to simulate different precipitation occurrence probabilities at each rain gauge. Generalization of this approach to more sample locations may not be straightforward. *Wheater et al.* [2000] also provided an extension to the spatial BLRP model where the mean duration of raincells was adjusted spatially. Non-homogeneous amounts and hourly occurrence were well fitted. However, the fit to daily and 6 hourly occurrences was typically biased low and the spatial nonstationarity of the daily and 6 hourly occurrence was not well modeled.

[9] Here the nonhomogeneous spatial activation of raincells (NSAR) model is developed and demonstrated to provide a single process which addresses the need for stochastic models able to simulate continuous spatial-temporal rainfall with spatially heterogeneous occurrence and amounts properties. This is achieved by extending the STNSRP model so that raincell occurrence within storms follows a spatially nonhomogeneous Poisson process.

2. The NSAR Model Structure

[10] The spatial-temporal NSAR model provides a stochastic conceptualization of rainfall in which a sequence of “storm events,” each of which may represent for example frontal rainfall or a mesoscale convective system, provide spatially large scale initiation of localized rainfall. Such localized rainfall exhibits both high variability and short range correlation in both space and time. Orographic enhancement may also significantly affect both the occurrence and amounts of rainfall at specific locations. For hourly and daily temporal scales and for spatial scales from point to ~10,000 km² the localized rainfall is conceptualized as occurring as circular raincells with uniform intensity throughout their lifetime and extent. These are distributed unevenly across the region of interest to account for geographic location effects on rainfall occurrence and clustered in time following each storm event. A final scaling of raincell intensity further accounts for location specific effects on rainfall amounts.

[11] The stochastic model structure is illustrated in Figure 1 and its detailed construction follows.

[12] 1. Storm events are modeled as storm origins, instants at which spatially large scale triggers of raincell events occur. Storm origins occur following a stationary Poisson process in time with rate λ (1/h).

[13] 2. Each storm origin generates a set of immobile circular raincells whose centers follow a nonhomogeneous spatial Poisson process, with density $\rho(\mathbf{x})$ (1/km²) over an infinite planar simulation region $\mathbf{x} \in \mathcal{R}^2$, and whose radii are exponentially distributed, with parameter γ (1/km).

[14] 3. Each raincell begins producing rainfall at its origin time, which follows the storm origin after a waiting time interval. Waiting time is exponentially distributed with parameter β (1/h).

[15] 4. Each raincell produces a uniform rainfall rate across its disc and throughout its lifetime. The duration and

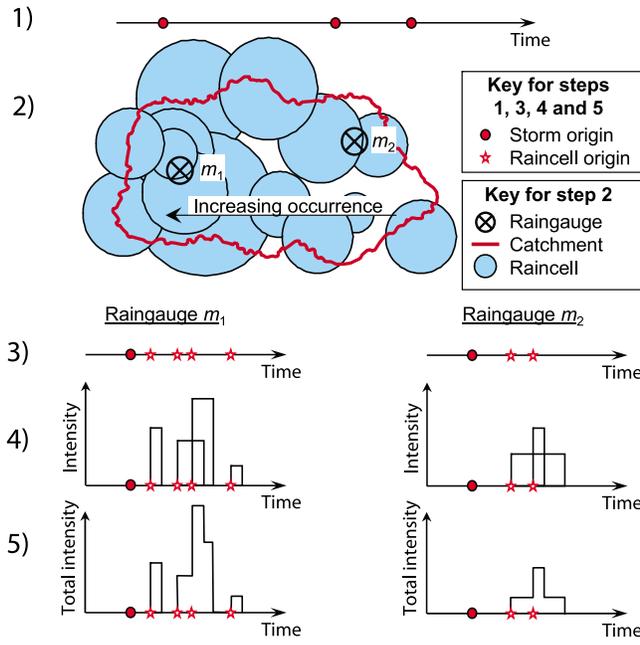


Figure 1. The conceptual structure of the stochastic NSAR model shown by means of a possible realization which is sampled at two rain gauges m_1 and m_2 . Steps are labeled (1–5) (1) a time series of storm origins; (2) the spatially non-homogeneous distribution of raincells generated by one such storm origin over a hypothetical catchment; (3) time series of raincell origins relevant to each rain gauge for this storm; (4) intensity and duration properties of these raincells; (5) scaled intensity time series generated at each rain gauge.

the intensity of the raincell are exponentially distributed with parameters η (1/h) and ξ (h/mm), respectively.

[16] 5. The rainfall intensity field at any instant is the sum of the intensities of all active raincells scaled by a spatially nonuniform intensity scaling field, $\psi(\mathbf{x})$. This field models geographically varying raincell intensities.

[17] Time series of spatially distributed fields of accumulated rainfall depths may be obtained numerically by integrating the intensity field over regular time steps for a grid of sampling locations. Similarly, multisite time series may be obtained by sampling the process at several locations, e.g., corresponding to rain gauges, and integrating the field over regular time steps. The NSAR model's parameters are summarized in Table 1 and consist of five parameters and two fields $\{\lambda, \beta, \rho(\mathbf{x}), \gamma, \eta, \xi, \psi(\mathbf{x})\}$ all of which are non-negative. Different parameterizations for each calendar month provide an annual cycle of rainfall properties.

[18] The new NSAR model is developed from the STNSRP stochastic rainfall model described by *Cowpertwait* [1995] and *Burton et al.* [2008]. Both models utilize a nonuniform intensity scaling field which models spatially varying rainfall amounts. However, the models differ in that the STNSRP model uses a homogeneous Poisson process to generate raincells in space with a uniform parameter ρ in step 2. Consequently rainfall occurrence is simulated as a spatially homogeneous process which takes no account of geographic effects and for each month the STNSRP model is parameterized by six parameters and one field $\{\lambda, \beta, \rho, \gamma, \eta, \xi, \psi(\mathbf{x})\}$ (see Table 1). In contrast, the new

methodology has the advantage that varying incidence of raincells at different geographic locations can additionally model spatially varying rainfall occurrence and storm duration. Both the NSAR and the STNSRP models provide a simplistic stochastic representation of the rainfall process: for example, raincells do not move and have a simplistic geometry, and storms are all considered to arise from the same process. However, the STNSRP process has been found useful at the spatial and temporal scales of particular relevance to hydrological investigations of catchments of up to $\sim 10,000\text{km}^2$ [see *Burton et al.*, 2008].

[19] In the single-site NSRP process [e.g., *Cowpertwait*, 1991], the sampling of spatial properties, step 2, is omitted altogether and instead the number of raincells that affect the single rain gauge may be sampled directly for each storm (a Poisson random variable with mean ν). Typically step 5 is also omitted as it is redundant. So for each month the NSRP model has five parameters $\{\lambda, \beta, \nu, \eta, \xi\}$. Here, however, it is convenient to consider an NSRP process at a location \mathbf{x}_m with intensity field scaling (i.e., atypically including step 5) which has the six parameters $\{\lambda, \beta, \nu, \eta, \xi, \psi_m\}$ where $\psi_m \equiv \psi(\mathbf{x}_m)$ (see Table 1). A useful property of this model [e.g., *Cowpertwait*, 1995] is that it is equivalent to an STNSRP process sampled at location \mathbf{x}_m provided common parameters are equal and

$$\nu = \frac{2\pi\rho}{\gamma^2}. \quad (1)$$

3. Fitting the NSAR Model

3.1. Fitting the STNSRP Model

[20] The fitting scheme of the NSAR model uses many properties of the STNSRP model and so the STNSRP fitting procedure is briefly summarized. First the intensity scaling field, $\psi(\mathbf{x})$, is estimated at rain gauge locations $\{\mathbf{x}_m\}$ in proportion to the mean daily rainfall, producing the vector $\Psi \equiv [\psi_m] \equiv [\psi(\mathbf{x}_m)]$. During a spatial simulation the intensity scaling field may be estimated by interpolation of these values so that the full STNSRP process is parameterized by $\{\lambda, \beta, \rho, \gamma, \eta, \xi, \Psi\}$, $(6 + M)$ parameters where M is the number of rain gauges, for each calendar month in turn. The same parameter set is used for multisite simulations, where time series are sampled at rain gauge record locations, but interpolation of the field is unnecessary [e.g., *Cowpertwait*, 1995; *Cowpertwait et al.*, 2002].

[21] Second, a numerical optimization scheme is used to find the best choice of the remaining parameters to minimize an objective function, equation (2) [see *Burton et al.*, 2008], which describes the degree to which a simulation is expected to correspond to observed rainfall statistics for a given month,

$$D(\lambda, \beta, \rho, \gamma, \eta, \xi | \Psi) = \sum_{g \in \Omega} \frac{w_g^2}{g_s} (\hat{g} - g(\lambda, \beta, \rho, \gamma, \eta, \xi, \Psi))^2, \quad (2)$$

where Ω is the set of rainfall statistics each with an aggregation period and a location (such as 24 h variance at rain gauge m_2), \hat{g} is the observed sample estimate of each statistic, and $g(\cdot)$ is the corresponding expected statistic from the simulation process expressed analytically in terms of the model's parameters. The weight applied to each statistic w_g is set by the user to control the relative accuracy with which

Table 1. The Parameters of the NSAR, STNSRP, and NSRP Models as Described in the Text^a

Parameter or Field	Model Usage			Description	Unit
	NSAR	STNSRP	NSRP		
λ	•	•	•	Storm origin arrival rate	(1/h)
β	•	•	•	1/(mean waiting time)	(1/h)
$\rho(\mathbf{x})$	•			Spatially varying raincell density field	(1/km ²)
ρ		•		Uniform raincell density field	(1/km ²)
ν			•	Number of raincells affecting a rain gauge	(-)
γ	•	•		1/(mean raincell radius)	(1/km)
η	•	•	•	1/(mean raincell duration)	(1/h)
ξ	•	•	•	1/(mean raincell intensity)	(h/mm)
$\psi(\mathbf{x})$	•	•		Spatially varying intensity scaling field	(-)
ψ_m			•	Intensity scaling at a specific location	(-)

^aCircles indicate the parameters used by each model. The fields $\rho(\mathbf{x})$ and $\psi(\mathbf{x})$ are functions of location, \mathbf{x} .

each statistic is fitted, in accordance with the uncertainty of each observed statistic or as most appropriate for a particular hydrological application. The scaling term g_s is either one for a probability dry or correlation statistic or the annual mean of the statistic. Analytical expressions for $g(\cdot)$ are available for expected statistics of any accumulation period of the STNSRP process at any location: for the mean, variance, lag-autocovariance and probability of a dry period (PDry) by *Cowpertwait* [1995]; dry-dry and wet-wet transition probabilities by *Cowpertwait* [1994]; and the third order central moment by *Cowpertwait* [1998]. Expressions relating the expected covariance between two locations are also available [*Cowpertwait*, 1995]. For the model to be identifiable Ω must include at least the same number of statistics as there are parameters to be fitted, both first and second order statistics, PDry and cross-correlation, and either the transition probabilities or the autocorrelation.

3.2. Analytical Properties of the NSAR Model

[22] The vector Ψ is used to parameterize the NSAR model's intensity scaling field as for the STNSRP model. Similarly the spatially varying raincell density field is characterized by point values ρ_m at nodes positioned at the locations $\{\mathbf{x}_m\}$ of the M rain gauges. The NSAR model parameterization is then $\{\lambda, \beta, \rho, \gamma, \eta, \xi, \Psi\}$, where $\rho \equiv [\rho_m]$, requiring a total of $(5 + 2M)$ parameters. The raincell density field has a more complex influence on the expected statistics of the model than the intensity scaling field, and so it is necessary to specify the form of the spatial interpolation used prior to fitting the model. Interpolation of nodal values using inverse square distance was chosen as it is a simple scheme which provides a smooth but responsive density field, $\rho(\mathbf{x}|\rho)$, at any location \mathbf{x} . The interpolated field is given by equation (3) with interpolation weights given by equation (4).

$$\rho(\mathbf{x} | \rho) = \sum_{m=1}^M w_m(\mathbf{x}) \rho_m, \quad (3)$$

$$w_m(\mathbf{x}) = \frac{|\mathbf{x} - \mathbf{x}_m|^{-2}}{\sum_{k=1}^M (|\mathbf{x} - \mathbf{x}_k|^{-2})}. \quad (4)$$

[23] Analytical expressions for the expected statistics of the NSAR process sampled at any point are now derived in terms of the expressions available for the STNSRP process. The NSAR process generates the storm origin and the raincell origin, duration, intensity, and radius in the same way as for the STNSRP process. Both the NSAR and the STNSRP models' Poisson raincell generation processes lead to a Poisson random number of raincells that all influence a rain gauge located at \mathbf{x} , $C_{\mathbf{x}}$. However, for the NSAR model $C_{\mathbf{x}}$ is nonhomogeneous in space whereas for the STNSRP model it is homogeneous. The NSAR process sampled at \mathbf{x} is therefore equivalent to an NSRP process at \mathbf{x} with common parameters equal and $\nu = E(C_{\mathbf{x}})$. An STNSRP process sampled at \mathbf{x} , with common parameters equal, is also equivalent provided ρ , ν , and γ are related by equation (1).

[24] The probability of a raincell with center \mathbf{x} and exponentially distributed radius (with parameter γ) influencing a rain gauge at \mathbf{x}_m is the survivor function of the radius random variable at the distance $|\mathbf{x} - \mathbf{x}_m|$, i.e., the probability of the radius being greater than the distance from \mathbf{x} to \mathbf{x}_m . The expected number of raincells reaching \mathbf{x}_m , $E(C_{\mathbf{x}_m})$, in a storm due to raincells with centers following a nonhomogeneous 2-D Poisson process defined by $\rho(\mathbf{x}|\rho)$ may then be evaluated as

$$E(C_{\mathbf{x}_m}) = \iint_{\mathbf{x} \in \mathbb{R}^2} e^{-\gamma|\mathbf{x} - \mathbf{x}_m|} \rho(\mathbf{x} | \rho) d\mathbf{x}. \quad (5)$$

[25] Equations (3), (4), and (5) may be combined and the order of integration and summation swapped to obtain an expression for the NSRP raincell number parameter at rain gauge m equivalent to the NSAR process sampled there,

$$\nu_m(\gamma, \rho) = \sum_{n=1}^M a_{mn}(\gamma) \rho_n, \quad (6)$$

where

$$a_{mn}(\gamma) = \iint_{\mathbf{x} \in \mathbb{R}^2} \frac{e^{-\gamma|\mathbf{x} - \mathbf{x}_m|} |\mathbf{x} - \mathbf{x}_n|^{-2}}{\sum_{k=1}^M (|\mathbf{x} - \mathbf{x}_k|^{-2})} d\mathbf{x} \quad (7)$$

is the expected number of raincells affecting a rain gauge at \mathbf{x}_m due to a unit density at the node located at \mathbf{x}_n .

[26] From these considerations, it follows that at rain gauge m the analytical expression for an expected single-site statistic, $g_{\text{NSAR},m}(\cdot)$, of the NSAR process may be related to the equivalent analytical expressions for the expected STNSRP process, $g_{\text{ST},m}(\cdot)$, and the single-site NSRP process, $g_{\text{SS},m}(\cdot)$, as shown in equation (8),

$$\begin{aligned} g_{\text{NSAR},m}(\lambda, \beta, \rho, \gamma, \eta, \xi, \Psi) &= g_{\text{ST},m}\left(\lambda, \beta, \frac{\nu_m(\gamma, \rho)\gamma^2}{2\pi}, \gamma, \eta, \xi, \Psi\right) \\ &= g_{\text{SS},m}(\lambda, \beta, \nu_m(\gamma, \rho), \eta, \xi, \Psi_m). \end{aligned} \quad (8)$$

[27] The expected cross correlation with distance relation is not a single-site statistic in the interpretation used for equation (8). Instead the cross-correlation between two rain gauges, m_1 and m_2 , a distance d apart may be approximated by two different estimates (using equation (8)) of the cross correlation, at distance d , evaluated at each of the two rain gauges. The arithmetic mean of these estimates provides a good approximation of the NSAR cross-correlation properties,

$$\begin{aligned} \text{corr}_{\text{NSAR},m_1,m_2}(\dots) &\approx \frac{1}{2} \left(\text{corr}_{\text{ST}}\left(d; \lambda, \beta, \frac{\nu_{m_1}(\gamma^2)}{2\pi}, \gamma, \eta, \xi, \Psi\right) \right. \\ &\quad \left. + \text{corr}_{\text{ST}}\left(d; \dots, \frac{\nu_{m_2}(\gamma^2)}{2\pi}, \dots\right) \right), \end{aligned} \quad (9)$$

where $\text{corr}_{\text{ST}}(d; \lambda, \beta, \rho, \gamma, \eta, \xi, \Psi)$ is the analytically expected cross-correlation at distance d for the equivalent STNSRP process at any location. Together the model properties described in equations (8) and (9) form the basis of the fitting algorithm of the NSAR process.

3.3. The NSAR Fitting Algorithm

[28] The algorithm to fit the NSAR model to a set of statistics Ω consists of five steps applied independently for each calendar month. NSAR model parameters are identified using numerical fits of the NSRP and STNSRP processes in accordance with the analytical relationships identified in section 3.2. First the best fit STNSRP process is identified (step 1). Then, allowing only the raincell number parameter and scaling field parameter to vary, the best fit NSRP process is identified for each rain gauge location (steps 2 and 3). Finally, the raincell density field and scaling field parameters are fitted (steps 4 and 5).

[29] In step 1, an STNSRP multisite fit is made to the sample estimates of the rainfall statistics Ω estimated from the observed records apart from PDry statistics which are fitted to the midpoint of the range of values across the rain gauges. This produces the parameter set $\mathbf{P}_1 = \{\lambda, \beta, \rho, \gamma, \eta, \xi, \Psi\}$, which provides a close homogeneous raincell occurrence fit to the observed rainfall statistics. It is assumed that the parameters λ , β , γ , η , and ξ remain close to optimal despite the introduction of spatial variation described in the following steps.

[30] In step 2, the parameter set, $\mathbf{P}_{2,m}$, is calculated for the single-site NSRP process that is equivalent to a sample at \mathbf{x}_m of the STNSRP process parameterized in step 1. The parameter ν replaces both ρ and γ according to equation (1) and the relevant intensity scaling field parameter is selected,

so $\mathbf{P}_{2,m} = \{\lambda, \beta, \nu, \eta, \xi, \Psi_m\}$. This is repeated for each rain gauge, m .

[31] Step 3 concerns the identification of the optimal raincell number parameter, $\nu_{3,m}$, and scaling field parameter, $\Psi_{3,m}$, for each rain gauge m . These are fitted conditional on the other parameters, $\{\lambda, \beta, \eta, \xi\}$, remaining fixed and the requirement that the mean rainfall at rain gauge m remaining unchanged.

[32] Since the mean is typically fitted well by step 1 the latter condition arises from the assumption that the STNSRP intensity scaling field contains a compensation for the imposed stationarity of rainfall occurrence, which may be removed, as raincell occurrence is allowed to vary. In terms of the model's analytical properties the condition may be written as

$$\mu_h^m(\lambda, \beta, \nu_{3,m}, \eta, \xi, \Psi_{3,m}) = \mu_h^m(\lambda, \beta, \nu, \eta, \xi, \Psi_m), \quad (10)$$

where $\mu_h^m(\cdot)$ is the analytical expression for the mean h hour rainfall at rain gauge m as a function of the model's parameters. Using the expression given by Cowpertwait [1995], this may be expanded as

$$\frac{h\gamma\Psi_{3,m}\nu_{3,m}}{\eta\xi} = \frac{h\lambda\Psi_m\nu}{\eta\xi}, \quad (11)$$

so that the condition may be written as

$$\Psi_{3,m}\nu_{3,m} = \Psi_m\nu. \quad (12)$$

[33] The raincell number parameter and the intensity scaling field are therefore fitted using a separate conditional NSRP numerical optimization for each rain gauge m in turn subject to equation (12) and keeping the remaining parameters fixed. Only the statistics from Ω relevant to m are used in this single-site fit, but a higher weight is applied to the dry period statistics to emphasize the precision with which spatial variability of rainfall occurrence is modeled. The analytical expressions used to fit each of the statistics are obtained using equation (8). Cross-correlation statistics relevant to m are also included by considering the equivalent STNSRP process at m (which forms part of equation (9)). The full approximation described by equation (9) is achieved separately at the two rain gauges. The NSRP parameter set for each rain gauge following step 3 is then $\mathbf{P}_{3,m} = \{\lambda, \beta, \nu_{3,m}, \eta, \xi, \Psi_{3,m}\}$.

[34] In step 4 the vector ρ parameterizing the spatially varying raincell density field, $\rho(\mathbf{x}|\rho)$, is fitted to the raincell number fits, $\{\nu_{3,m}\}$, from the previous step. Simply setting ρ_m at each rain gauge using equation (1) is inappropriate as this would assume a uniform density field. Instead equation (6) describes both the nonuniform density field and the spatial influence of the raincells. Using this the fitted vector ρ_4 may be obtained by fitting the $\{\nu_m(\gamma, \rho)\}$ properties of the NSAR process to the fitted $\{\nu_{3,m}\}$, by minimizing the magnitude of the error term ε_m in the expression

$$\nu_{3,m} = \nu_m(\gamma, \rho) + \varepsilon_m. \quad (13)$$

[35] For practical applications equation (13) may be solved using standard matrix inversion to find an exact

solution. This is summarized in equation (14) in vector notation where $\nu_3 = [\nu_{3,m}]$, the matrix $\mathbf{A} = [a_{mn}]$, and the error $\varepsilon = [\varepsilon_m]$ is zero valued. The parameter set following step four is then $\mathbf{P}_4 = \{\lambda, \beta, \rho_4, \gamma, \eta, \xi, \Psi_3\}$,

$$\rho_4 = \mathbf{A}^{-1} \nu_3. \quad (14)$$

[36] Step 5 concerns the identification of the optimum feasible spatially varying raincell density field and the corresponding intensity scaling field. While step 4 guarantees an optimal fit to the raincell number parameters, fitted in step 3, it does not guarantee a feasible solution, i.e., one that conforms to the standard definition of a Poisson density field. By definition the density field of a Poisson process must be nonnegative, i.e., $\rho(\mathbf{x}|\rho) \geq 0$. For the field description given by equations (3) and (4), this is equivalent to requiring that $\rho_n \geq 0$ for all nodes n . If ρ_4 satisfies this constraint it is confirmed to be the optimal feasible solution, i.e., $\rho_5 = \rho_4$ and $\Psi_5 = \Psi_3$.

[37] For other cases the field of quadratic programming (QP) [e.g., Hillier and Liberman, 1974], an extension of linear programming with the useful property that it is constrained to a nonnegative domain, provides an appropriate algorithm to find an optimum feasible solution to equation (13). QP has previously been used in hydrology to devise correctly constrained parameterizations for unit hydrograph linear models [Natale and Todini, 1976]. Expanding the squared magnitude of the error vector ε^2 using equation (6) and reorganizing terms we can write

$$D_2(\rho) \equiv \frac{1}{2} \left((\nu_3)^2 - \varepsilon^2 \right) = (\mathbf{A}^T \nu_3)^T \rho - \frac{1}{2} \rho^T (\mathbf{A}^T \mathbf{A}) \rho, \quad (15)$$

in which we define the objective function $D_2(\rho)$. This objective function increases with decreasing ε vector magnitude (because ν_3 is constant for this step) and presents step 5 in the form of a standard QP maximization problem. Optimum nonnegative parameters, ρ_5 , are then determined using a standard algorithm. Finally, equation (12) is updated for the current step,

$$\psi_{5,m} = \frac{\Psi_{3,m} \nu_{3,m}}{\nu_{5,m}}, \quad (16)$$

where

$$\nu_5 = \mathbf{A} \rho_5. \quad (17)$$

[38] The fitted NSAR parameter set is then $\mathbf{P}_5 = \{\lambda, \beta, \rho_5, \gamma, \eta, \xi, \Psi_5\}$.

4. Simulating the Raincell Generation Process

4.1. An Efficient Algorithm for the Simulation of STNSRP Raincells

[39] Both the NSAR and the STNSRP processes simulate raincell centers according to a Poisson process on an infinite plane. To simulate such a process an algorithm is therefore required to reduce this infinite process to a finite one, as infinite processes are not computable.

[40] Consider an STNSRP spatially homogeneous raincell generation process over an infinite plane partitioned into inner and outer simulation regions. The inner region is a

finite rectangular area containing all of the calibration rain gauges and all locations at which the simulation process is to be sampled. The outer region is infinite and comprises the rest of the plane.

[41] Here a new finite, exact, and efficient algorithm is used, the derivation of which is provided in Appendix A. An algorithm with similar properties was demonstrated by Leonard *et al.* [2006] for a circular inner region. In contrast a rectangular region is likely to be more computationally efficient than a circular region as a circle containing a randomly positioned set of rain gauges is likely to have greater area (and therefore require greater computational effort) than a similarly defined rectangle. Further, the new derivation is more general and easily adapted to other geometric shapes.

[42] The efficient simulation of the homogeneous spatial raincell process proceeds separately for the inner and outer regions. The simulation of raincell centers in the rectangular inner region is finite and straightforward, and the radii are sampled from an exponential distribution with parameter γ as usual. For the outer region, the total number of relevant raincells is first sampled as a Poisson random variable with mean ρ_y (see equation (A9)). For each such raincell, the distance of the raincell's center from the inner region, x , is sampled from the cumulative distribution function given by equation (A8), its location is fully realized by uniform sampling from all locations at this distance, and its radius is exponentially distributed, with parameter γ , conditional on it being greater than x . Details of the derivation of the algorithm are presented in Appendix A and the algorithm is demonstrated in section 5.

4.2. Nonhomogeneous Spatial Simulation of Raincells

[43] For an NSAR simulation an algorithm is required to sample raincell incidence within a storm according to an infinite nonhomogeneous 2-D Poisson process. Here the algorithm for a homogeneous process described in section 4.1 is extended using an acceptance-rejection algorithm to provide a finite and exact simulation of relevant raincells arising from a nonhomogeneous 2-D Poisson process. For the purposes of this description, we will consider the simulation of relevant raincells (with radius parameter γ) sampled according to a nonhomogeneous and infinite Poisson process with a density field, $\rho(\mathbf{x}|\rho)$, as described in section 3.2.

[44] For each node, m , in turn, the relevant raincells corresponding to a homogeneous occurrence process with density ρ_m are generated for both the inner and outer regions using the algorithm described in section 4.1. Once the location of each raincell is known, \mathbf{x} say, an acceptance-rejection rule is imposed. The raincell is discarded with probability $(1 - w_m(\mathbf{x}))$, where $w_m(\mathbf{x})$ is the interpolation weight as described in equation (4), and which is independent of the model parameters.

[45] The set of raincells generated and accepted for all of the nodes provides an accurate simulation of relevant raincells of the field, $\rho(\mathbf{x}|\rho)$, parameterized by the vector $\rho \equiv [\rho_m]$. This may be shown from first principles as follows. Consider a small area, δA , located at \mathbf{x} . This area will receive an accepted raincell according to node m with probability $w_m(\mathbf{x}) \rho_m \delta A$. The total number of accepted raincells contributed from all nodes will then be $\rho(\mathbf{x}|\rho) \delta A$ (from equation (3)), which is the required nonhomogeneous Poisson density. The

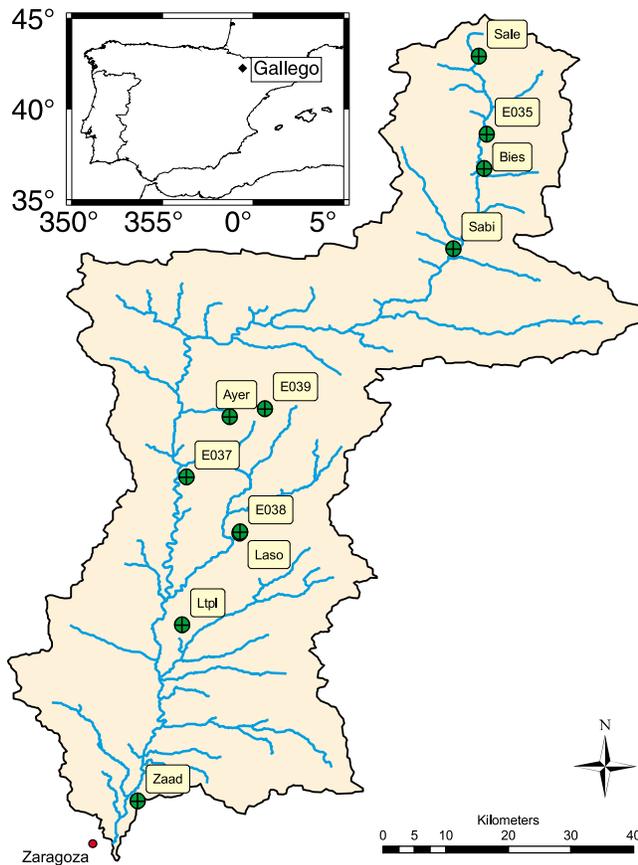


Figure 2. Locations of the seven daily and four 15 min rain gauges in the Gallego catchment above Zaragoza, Spanish Pyrenees. Laso and E038 lie close to each other.

accuracy of the simulation algorithm is demonstrated for the case study application in section 5.

5. Application to the Gallego

5.1. The Gallego Catchment Case Study

[46] The Gallego catchment was selected for the AQUATERRA project [Gerzabek et al., 2007] as a suitable case study for interdisciplinary investigation of the combined

effects of climate change, land use management, river regulation, and pollution. These investigations required a suitable model of the rainfall regime which was found to exhibit extreme orographic effects. Therefore, the catchment provided a challenging case study for the evaluation of the new NSAR methodology. A successful application in such an extreme environment would imply that the model would be suitable for a wide range of European catchments with important, but smaller, magnitude orographic effects.

[47] The Gallego catchment (4009 km²) lies to the south of the Pyrenees (Figure 2) and is a subcatchment of the Ebro River (85,000 km²). The Ebro is a typical Mediterranean River characterized by low summer flows and higher winter flows and has a semiarid climate with a mean annual precipitation of ~350 mm. Rainfall in the Gallego is particularly influenced by orography: the daily occurrence more than doubles and the monthly amounts increase between 2 and 7 times across the catchment.

[48] Seven daily and four 15 min time series were selected after error checking and considering location and record length (Table 2). Their distribution throughout the catchment is shown in Figure 2. The seven daily rain gauges were partitioned into calibration, {Ayer, Ltpl, Sabi, Sale, Zaad}, and validation, {Bies, Laso}, groups for model evaluation.

[49] The 6 years of available 15 min rainfall observations were insufficient to establish reliable sample estimates of rainfall statistics or extremes at each rain gauge. Instead it proved possible to establish catchment-wide nonlinear downscaling regression relationships between hourly and daily statistics estimated for each complete month of records for all six of the four 15 min records together. The hourly variance may be estimated as $VarH = (0.016797) VarD^{0.87418}$ and the proportion of dry hours (with less than 0.1 mm accumulation), PDH, may be estimated from equation (18). These relationships have R^2 values, of 88% and 82%, respectively, where $VarD$ is daily variance and $PDD_{1.0}$ is the proportion of dry days with less than 1.0 mm of rainfall.

$$\ln\left(\frac{PDH}{1 - PDH}\right) = (1.4310) + (0.98659) \ln\left(\frac{PDD_{1.0}}{1 - PDD_{1.0}}\right). \quad (18)$$

[50] The observed rainfall properties were then summarized as a set of daily and hourly rainfall statistics. For each of the seven daily rain gauges, the daily mean, variance,

Table 2. Properties of the (a) Seven Daily and (b) Four 15 min Rain Gauge Datasets Used in the Gallego Catchment Case Study^a

Identifier	Name	Easting (m)	Northing (m)	Altitude (m)	First Year	Length (years)	Annual Rainfall (mm)
(a) Daily Data							
Calibration							
Ayer	Ayerbe	690,578	4,682,490	582	1945	48	651
Ltpl	El Temple	686,242	4,649,530	335	1943	50	392
Sabi	Sabiñanigo	716,824	4,710,850	790	1941	52	807
Sale	Sallent de Gallego la Sarra	718,562	4,741,140	1460	1960	33	1490
Zaad	Zaragoza Aula Dei	682,093	4,621,640	225	1950	42	367
Validation							
Bies	Biescas “E.I.A.”	720,026	4,723,670	855	1941	39	855
Laso	La Sotonera “Embalse”	692,654	4,664,422	413	1941	62	471
(b) 15 min Data							
E035		720,084	4,729,052	1085	2000	6	
E037		685,698	4,672,628	423	2000	6	
E038		692,715	4,664,605	415	2000	6	
E039		694,826	4,684,074	653	2000	6	

^aThe calibration and validation groupings of the daily rain gauges are indicated. Easting and northing are given in UTM coordinates.

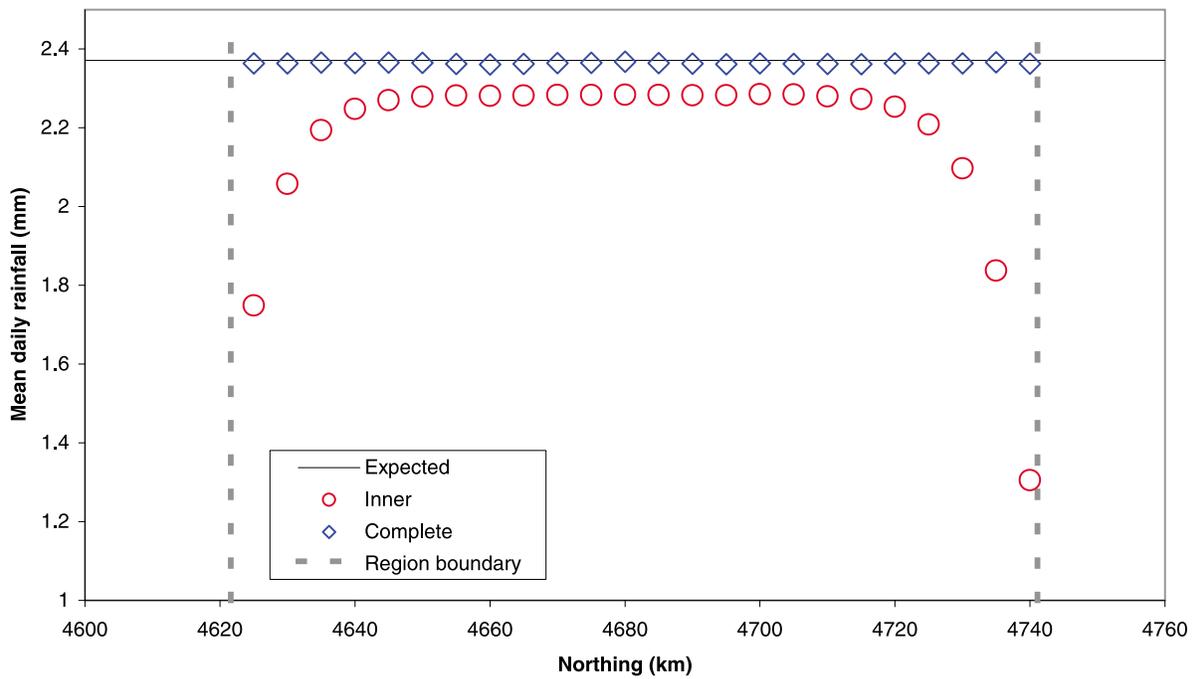


Figure 3. Comparison of the mean daily rainfall amounts sampled from two spatial simulation algorithms along a north-south transect through the center of the inner region with the value expected for the STNSRP process (Expected). For the “Inner” simulation raincell centers were generated with a uniform rate but only within the inner region. The “Complete” simulation additionally generated relevant raincells in the outer region using the efficient algorithm. The inner region boundaries are also indicated (Region boundary).

proportion of dry days (less than 0.2mm, PDD), lag-1 autocorrelation (AC), skewness coefficient and spatial cross correlation between the rain gauges were estimated. Hourly rainfall properties, $VarH$ and PDH , were also estimated for each daily rain gauge by applying the downscaling regression relationships established for the catchment. These statistics were then partitioned into calibration and validation datasets according to the originating daily rain gauge. The 10 intergroup daily cross-correlation statistics were assigned to the validation dataset. The calibration and validation datasets then comprised 45 and 25 unique statistics, respectively, for each calendar month and so provide a good basis for the calibration, validation, and evaluation of the stochastic rainfall simulation models.

5.2. Test of the Efficient STNSRP Simulator

[51] First, the accuracy of the efficient STNSRP raincell simulation scheme described in section 4.1 is demonstrated. A rectangular inner region was prepared for the Gallego rain gauges. A test parameter set was chosen: each parameter with a constant value for all calendar months; using a mean raincell radius of 5 km (i.e., $\gamma = 0.2 \text{ km}^{-1}$); with all ψ_m parameters equal; and with typical values for the other parameters. The STNSRP process was then simulated twice for 1000 years with these parameters: (1) simulating raincells only within the rectangular inner region (Inner) and (2) additionally simulating relevant raincells occurring in the outer region using the efficient simulation procedure (Complete). The two simulations were sampled at 24 locations along a north-south transect through the mid-point of the inner simulation region. Figure 3 compares the mean

rainfall amounts simulated at each location with the values expected for the STNSRP process. The missing contributions from the outer region clearly lead to a low bias in the simulated rainfall amounts for the Inner simulation with the bias increasing as the region’s boundary is approached. Even in the center of the inner region the proximity of the region’s eastern and western boundaries has an influence. The Complete simulation is shown to be an excellent match to the expected properties of the STNSRP process; this provides a practical demonstration of the efficient algorithm and illustrates the need for an efficient means to sample the process in the outer simulation region.

5.3. The STNSRP Model Calibration

[52] To provide a comparison with the NSAR model and to illustrate the limitations of the STNSRP process for catchments exhibiting widely varying rainfall occurrence, a fit was made to the observed properties of the five calibration rain gauges (see Table 2). The observed calibration statistics and the STNSRP model fit to these are shown in Figure 4. For validation purposes, the intensity scaling field values at the two validation rain gauges were estimated by averaging the values at their two nearest rain gauges to provide a simple interpolation. A 1000 year simulation was then generated and sampled at the daily rain gauges locations. Analysis of the simulated time series sampled at the calibration rain gauge locations (not shown) confirmed that the model’s behavior was an excellent match to the fits shown in Figure 4.

[53] From Figure 4 the observed mean, variance and proportion dry statistics are seen to vary considerably across

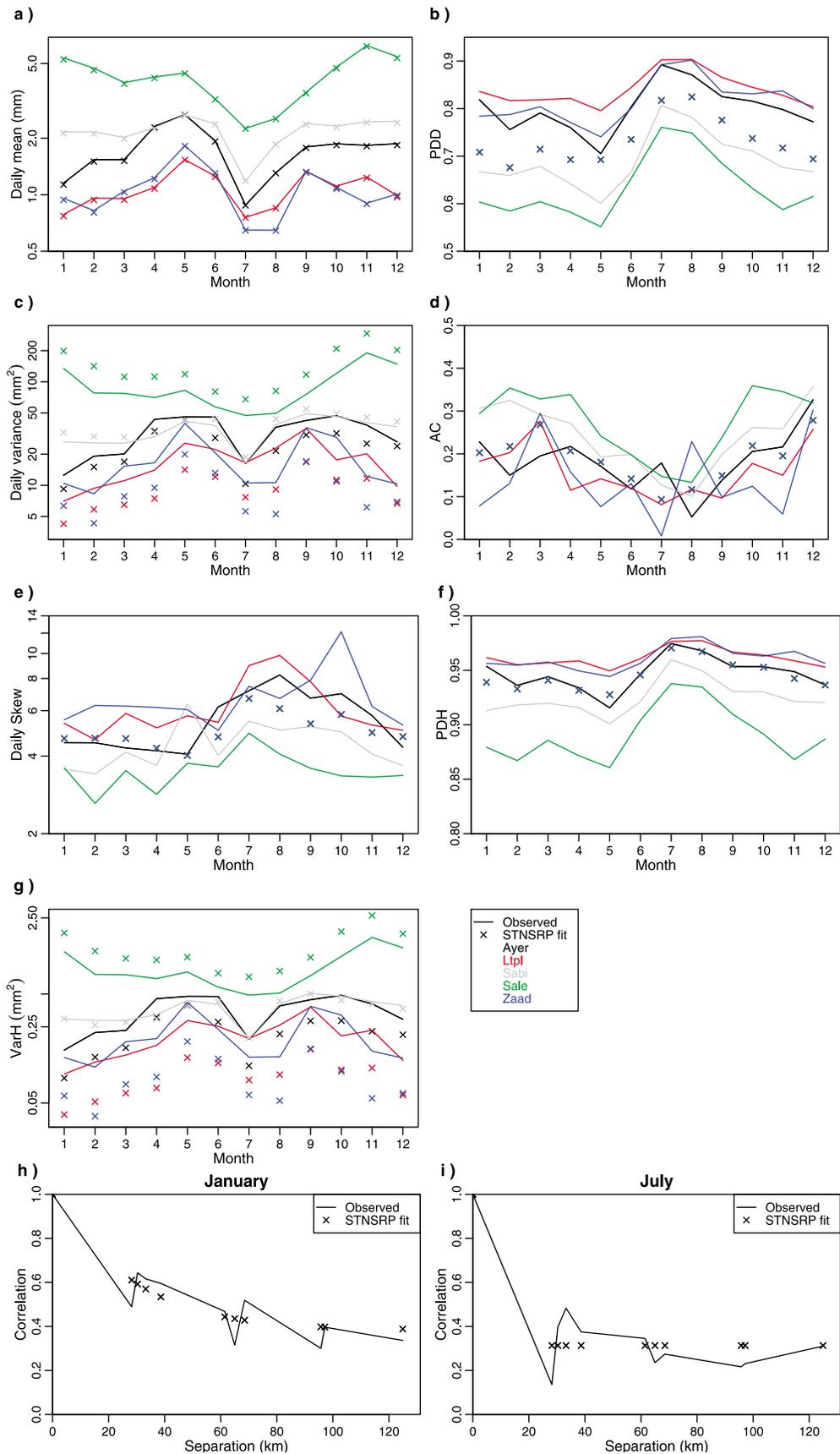


Figure 4

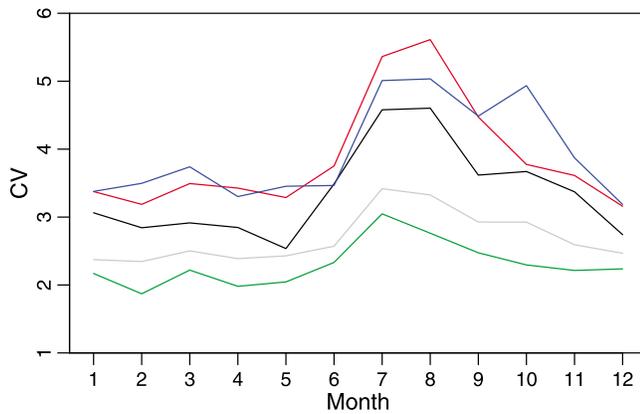


Figure 5. The annual cycles of the coefficient of variation (CV) for each calibration rain gauge (colors as for Figure 4).

the catchment. The STNSRP process fits the mean rainfall well as it is able to represent the nonhomogeneous amounts process. The two spatial cross-correlation plots show how rainfall is less correlated in summer for nearby rain gauges than in winter, possibly due to an increase in convective compared to frontal activity. This pattern is well fitted by the model. However, the fitted PDD, PDH, lag-1 autocorrelation and skewness coefficient statistics are spatially uniform and so each month's fit for these statistics is the same for all rain gauges.

[54] Of particular note for hydrological applications is the observed daily rainfall occurrence, as indicated by $(1 - \text{PDD})$. At the wetter rain gauges this is at least double that at the drier rain gauges. Therefore the STNSRP model conceptualization of homogeneous rainfall occurrence in space is clearly not appropriate for this catchment. Further, the use of the intensity scaling field in the STNSRP model implicitly assumes that dimensional statistics vary in proportion to an appropriate power of the mean (e.g., that the daily coefficient of variation [CV] is spatially uniform). However, as Figure 5 shows, the observed data exhibits a seasonally variable spatially nonhomogeneous CV, the driest rain gauges and summer months having the highest CVs. Amounts scaling therefore results in fitted daily variance statistics with excessive spatial variability as shown in Figures 4c and 4g for *VarD* and *VarH*.

5.4. The NSAR Model Calibration

[55] The parameters of the NSAR process were fitted to the observed calibration dataset according to the procedure described in section 3.3. The fitted raincell density field for December is shown in Figure 6. This illustrates how equations (3) and (4) interpolate the fitted nodal parameters, the differing influences of the calibration and validation rain gauges on the field and the increase of the raincell density field toward the North of the catchment. Raincells, with a

mean radius of 12.3 km ($\gamma = 0.0813 \text{ km}^{-1}$), were generated for 250 simulated storms using the nonhomogeneous algorithm described in section 4.2. These raincell centers are illustrated as points in Figure 6. Within the inner region, which contains all seven daily rain gauges, the density was found to be consistent with the contours. As required for the outer region, the density can be seen to increase with the fitted field, decrease with distance from the inner region (as fewer relevant raincells are simulated) and match the density of the inner region at the boundary.

[56] For validation purposes, intensity scaling field values were estimated at the two validation rain gauges by simply averaging the values at the two neighboring rain gauges. A 1000 year spatial-temporal NSAR simulation was then generated and sampled at the daily rain gauge locations. The sample statistics of the time series sampled at the five calibration rain gauges are shown in Figures 7 and 8. The simulated mean and proportion dry statistics (Figures 7a, 7b, and 7f) are excellent matches to the observed. These demonstrate successful modeling of the observed spatial heterogeneity of both rainfall occurrence and amounts at daily and hourly scales using a spatially and temporally continuous process, despite the extreme orographic effects present in the Gallego catchment. Comparison with Figures 4a, 4b, and 4f clearly shows the improvement of the NSAR model over the STNSRP model. It should also be noted that the simulated results for the mean (Figure 7a) closely match the observed (shown) and the fitted NSAR model properties (not shown) and so demonstrate that the simulation process is correctly implemented.

[57] The simulation accuracy of the daily variance statistics is also excellent (Figure 7c) and the systematic biases noted for the STNSRP model (Figure 4c) have been eliminated. The NSAR model is therefore able to correctly simulate the observed spatially nonhomogeneous CV (Figure 5), overcoming the homogeneous CV limitation of the STNSRP model. The simulated hourly variance also seems improved (Figures 7g and 4g). The skewness coefficient typically has a large sample variability and is biased by the length of observations (short records generally have lower skewness coefficients than longer records) [Wallis *et al.*, 1974]. The NSAR simulated skewness coefficient follows the spatial pattern of the observations but appears to be over simulated for the lower altitude rain gauges. In contrast the STNSRP fit simply matched the spatially averaged skewness (Figure 4e). The lag-1 autocorrelation statistic improves upon the uniform value fitted for the STNSRP model. The cross correlations for pairs of calibration rain gauges (Figure 8) are seen to match observations well and to exhibit similar variability to the observed data whereas a smooth curve is typically produced by the STNSRP process (e.g., Figure 4h and Burton *et al.* [2008]). This application demonstrates 12 independent monthly calibrations following a seasonal cycle and a corresponding simulation of the NSAR methodology and so provides confidence that the fitting methodology is not data dependent.

Figure 4. Observed and STNSRP model fits to (a–e, h, i) daily and (f, g) hourly calibration rainfall statistics for the Gallego. The colors in plots a–g indicate statistics relevant to individual rain gauges as shown in the key. Note that the y axis is logarithmic for mean, variance, and skewness coefficient statistics. Daily cross correlation between pairs of rain gauges is shown as a function of separation for (h) January and (i) July.

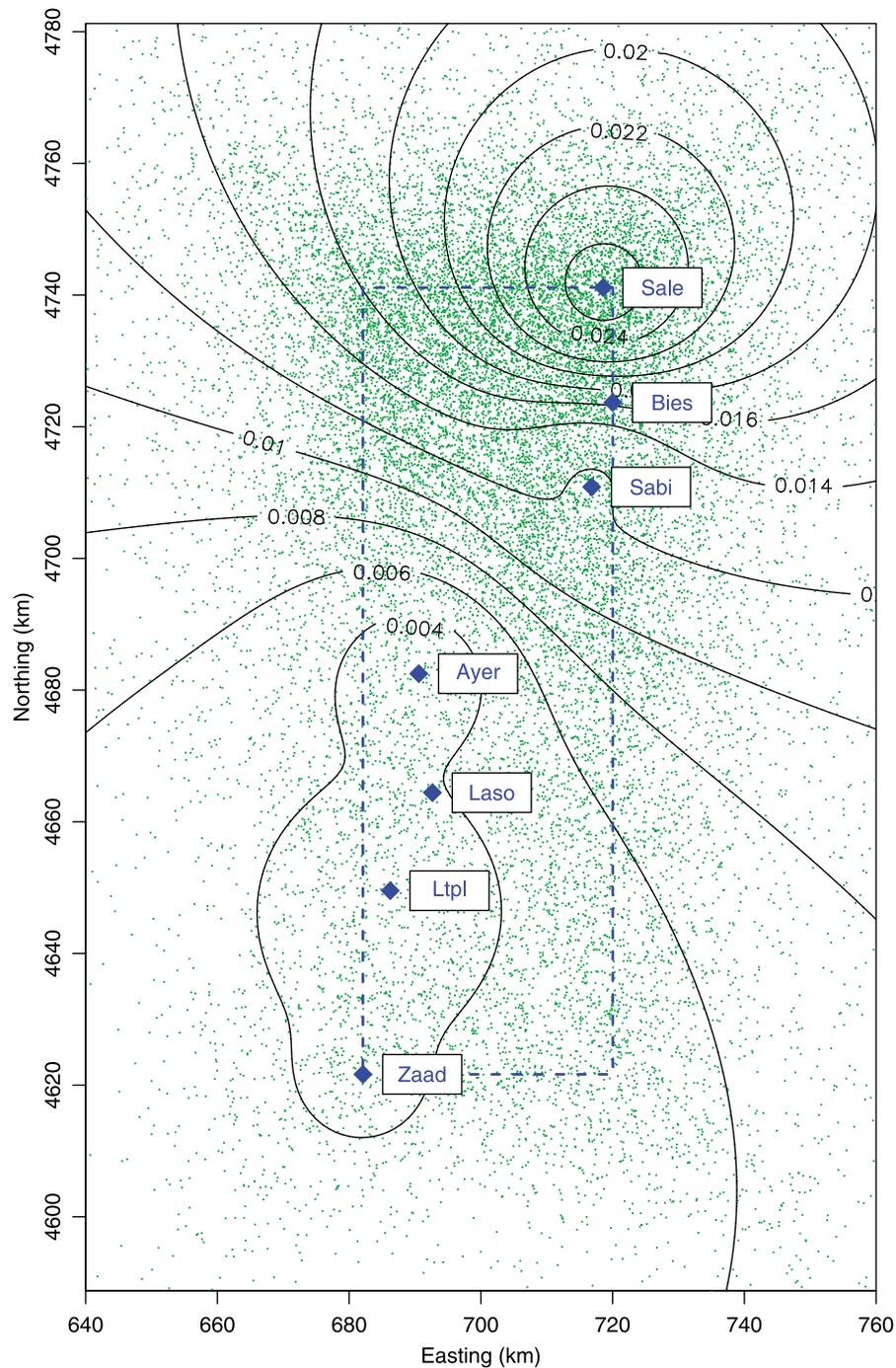


Figure 6. The fitted raincell density field for December shown as contours with an interval of 0.002 km^{-2} . The points indicate raincell centers generated for 250 storms simulated using the nonhomogeneous algorithm described in section 4.2. Daily rain gauges (diamonds) and the inner region boundaries (dashed line) are also shown.

5.5. Validation Results

[58] A validation of the stochastic rainfall models is now presented in which properties of the simulated rainfall fields are compared with observed rainfall properties that were not included in the model calibration. Previously, *Cowpertwait* [2006] demonstrated the sampling of an STNSRP process at locations which were not used in the calibration of the model. Here we use this property as the basis of a validation

of the both the NSAR and the STNSRP models by testing such samples against historical records that were not included in model calibration. The observed and simulated rainfall statistics are shown in Figure 9 for time series sampled at the validation rain gauge locations. Despite the rain gauge location being the only property used to obtain the time series, excellent matches are obtained for the mean, PDD, *VarD*, and PDH, providing a successful validation of

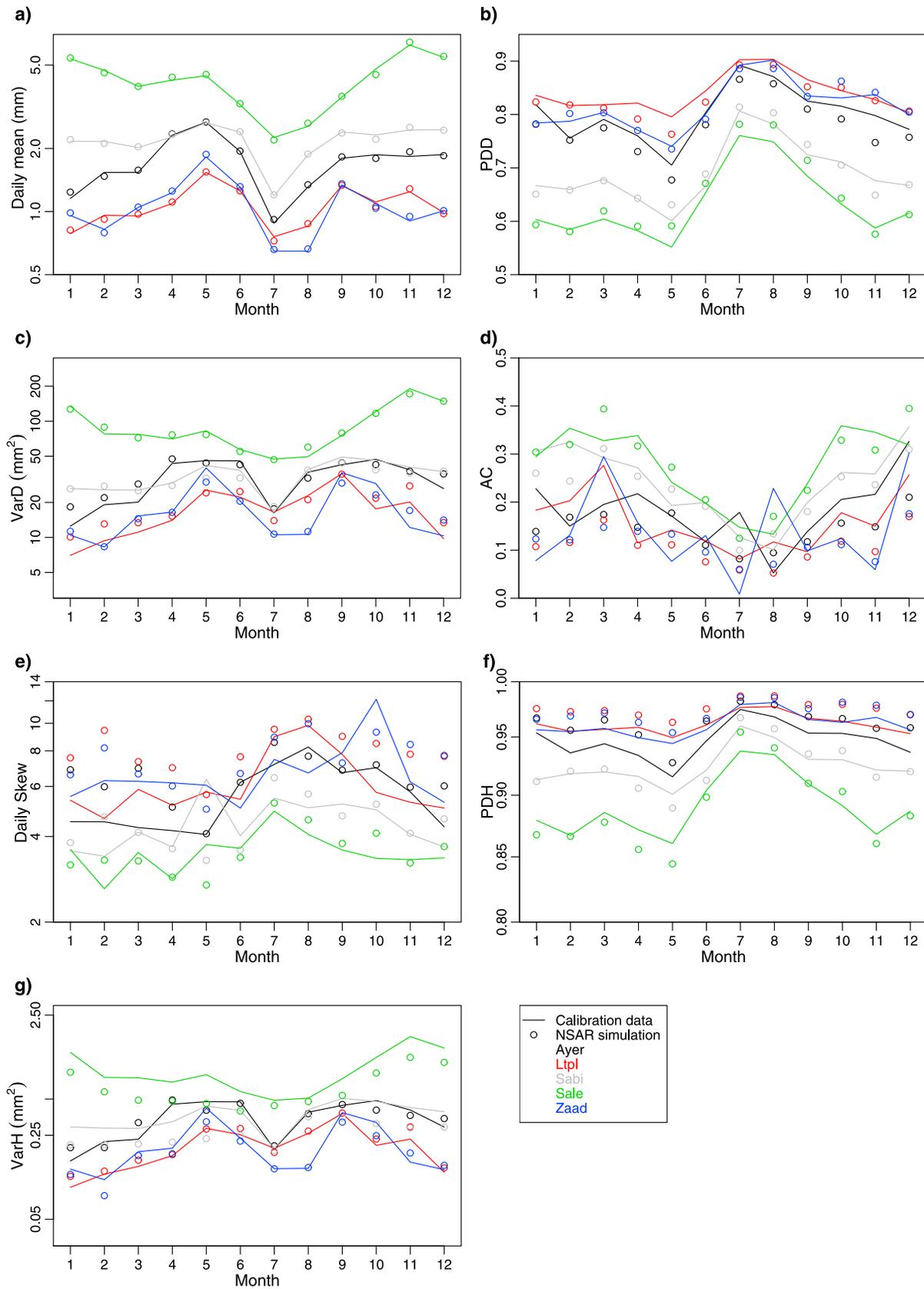


Figure 7. Comparison of the calibration dataset and the NSAR simulated rainfall statistics. The calibration dataset comprises (a–e) observed daily statistics and (f, g) estimated hourly statistics. Colors correspond to rain gauges.

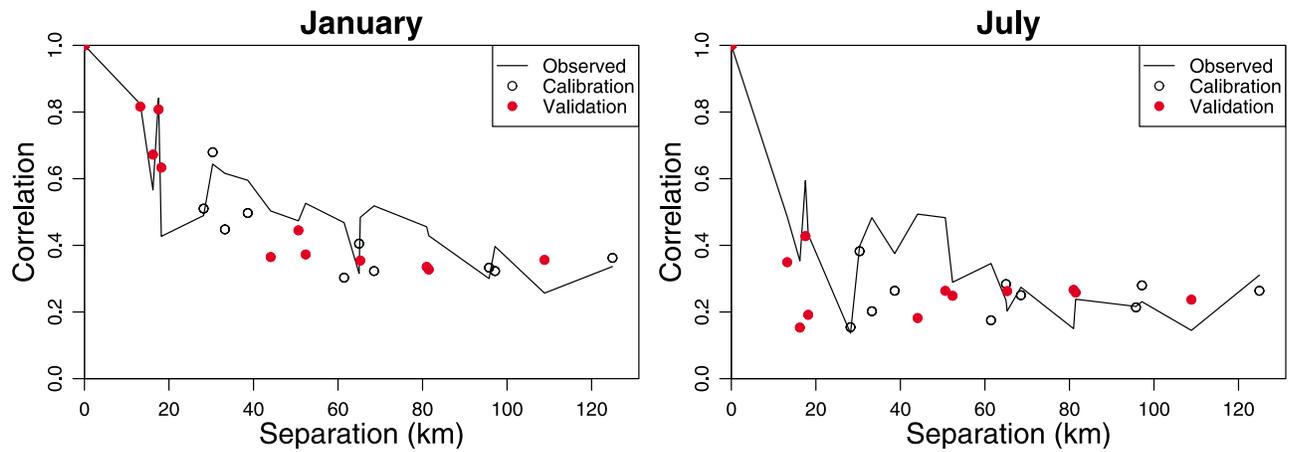


Figure 8. The daily cross-correlation statistics estimated from the NSAR simulation for the calibration and validation rain gauge pairings. These are compared with the observed dataset and plotted against rain gauge separation for January and July.

the model's heterogeneous rainfall occurrence properties. However, the model's representation of $VarH$, AC , and skew appears slightly worse than for the calibration results. Finally, Figure 8 shows the simulated cross correlation of pairs of rain gauges where the pairing was not used in the calibration: the simulated values are an excellent match to the observed datasets.

[59] Table 3 provides a summary analysis of the accuracy with which the observed statistics are matched by time series sampled at the locations of the calibration and validation rain gauges for both the STNSRP and the NSAR models. Since the validation dataset has less variance than the calibration data for most statistics, the calibration and validation scores cannot be directly compared in this table. However, a reference value is provided, "Reference", in which observations are similarly compared to the mean of the observed *calibration* statistics. The Reference, NSAR, and STNSRP scores may be compared for each group of sites. Table 3 shows quantitatively that the NSAR model significantly improves the representation of the mean, PDD, $VarD$, AC , PDH, and $VarH$ for the calibration dataset compared with the STNSRP model. Of these statistics the validation results confirm an improvement in the mean, PDD, $VarD$, and PDH statistics. Of the remaining validation statistics AC , $VarH$, and cross correlation appear to be modeled with a similar quality whilst skew may not be modeled as well as for the STNSRP model. The STNSRP model's validation results are shown to improve on the Reference value for the mean, $VarD$, and $VarH$ results. Additionally, the NSAR model improves on the PDD and PDH results. While improved calibration results alone could simply be a consequence of an increase in the number of model parameters, the validation results demonstrate that the NSAR model has a significantly improved representation of the rainfall field compared with the STNSRP model as the new model demonstrates an improved predictive ability.

[60] To test the ability of the NSAR model to simulate realistic rainfall extremes, the time series of annual maximum daily rainfall was extracted from the observed record for three rain gauges (with high, medium and low annual rainfall), discarding years containing missing data. The maxima were ranked and are shown on a Gumbel proba-

bility plot in Figure 10. The simulated 1000 year time series corresponding to each rain gauge location was then partitioned into subseries with equal length to the observed series. For each subseries the annual maxima were extracted and ranked. The 10th and 90th percentiles of the distribution of each ranking was then evaluated across all subseries and plotted in Figure 10 for each rain gauge, providing a measure of the variability of the simulated extreme value curve. These plots show an excellent agreement of the observed and simulated daily extreme value properties of the rainfall field, which were not used in model calibration. This strongly suggests that the model provides an excellent representation of the rainfall regime of the catchment.

6. Discussion and Conclusions

[61] The nonhomogeneous spatial activation of raincells (NSAR) model is the first spatially and temporally continuous stochastic rainfall model demonstrated to successfully simulate long hourly time series exhibiting both spatially nonhomogeneous occurrence as well as nonhomogeneous amounts. Previous spatial-temporal models, such as the STNSRP model, have successfully modeled only the nonhomogeneous amounts process. However, the NSAR model implements a modified STNSRP process in which raincell incidence follows a spatially nonhomogeneous Poisson process, under the assumption that rainfall occurrence at all locations is triggered by a single initiating event. The new model is able to simulate rainfall regimes in which rainfall amounts, rainfall occurrence and storm duration properties are spatially nonhomogeneous.

[62] The new NSAR model was evaluated against the STNSRP model, using the Gallego catchment in the Spanish Pyrenees which has a rainfall regime exhibiting extreme orographic influences. The models were calibrated to a set of observed daily statistics and to hourly statistics, estimated using regionalized relationships, for five rain gauges. Validation data was not used in model calibration and comprised extreme value statistics and daily and hourly statistics for two additional rain gauges. This provides the first validation of an STNSRP or NSAR process at a location not used in model fitting.

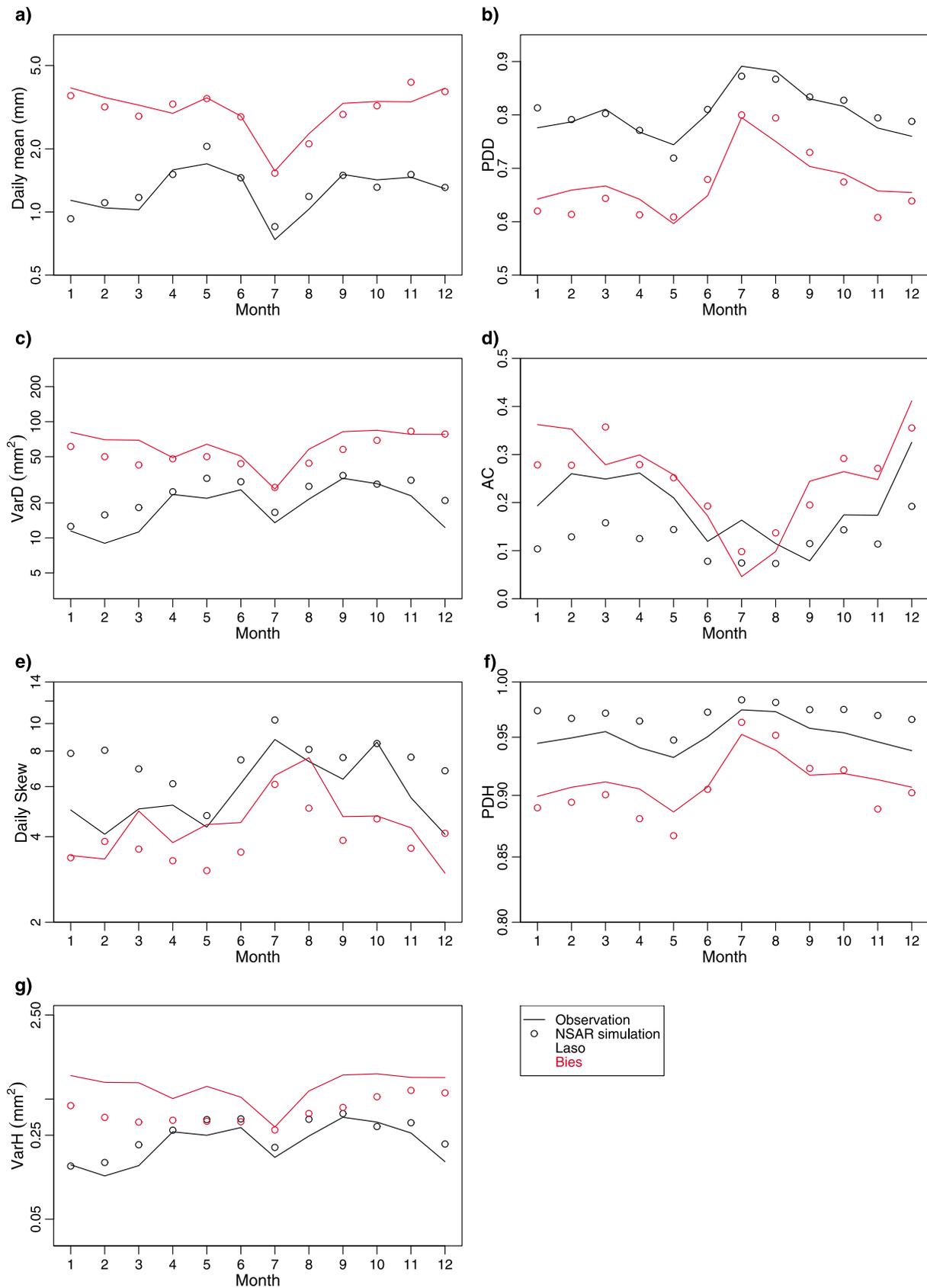


Figure 9. Comparison of observations and the corresponding samples from the NSAR simulation for two validation rain gauges, which were not used in the model calibration. Colors correspond to rain gauges.

Table 3. A Comparison of the Statistics of the Time Series Simulated by the STNSRP and the NSAR Models for the Calibration and Validation Groups of Rain Gauges^a

Model	Group	Daily						Hourly	
		Mean (mm)	PDD (-)	Variance (mm ²)	AC (-)	Skew (-)	XCorr (-)	PDH (-)	Variance (mm ²)
Reference	Calibration	1.248	0.081	33.6	0.067	1.49	-	0.027	0.29
STNSRP	Calibration	0.101	0.078	25.8	0.069	1.61	0.09	0.027	0.25
NSAR	Calibration	0.065	0.019	5.2	0.058	1.53	0.11	0.012	0.14
Reference	Validation	0.998	0.067	25.7	0.063	1.00	-	0.020	0.24
STNSRP	Validation	0.285	0.062	17.7	0.069	1.06	0.10	0.022	0.15
NSAR	Validation	0.259	0.025	11.6	0.071	1.60	0.10	0.017	0.19

^aEach value is the root mean square difference between the observed and simulated statistics of the relevant rain gauges for all months. "Reference" is a similar evaluation of the monthly mean of the observed calibration statistics.

[63] The STNSRP model, which has a homogeneous rainfall occurrence process, was shown to be too inflexible to model the catchment's nonhomogeneous rainfall regime. Spatial heterogeneities in the mean, daily, and hourly proportion dry, daily variance, and daily coefficient of variation were shown to be better simulated by the NSAR model compared with the STNSRP model for both calibration and validation datasets. Other statistics were found to be of a similar quality for both models. Extreme value statistics were shown to be well simulated by the NSAR model providing a further validation. The new model is therefore demonstrated to significantly improve on the spatial modeling of orographically influenced rainfall compared with the STNSRP model. Sampling the process at subdaily time steps and for locations not included in model calibration also show that the new methodology goes considerably beyond daily multisite approaches.

[64] The increased complexity of the NSAR model requires more parameters (a total of 15 for the Gallego application) than the STNSRP model (11) both of which were fitted to 45 calibration statistics. This number of parameters is comparable with the closest models demonstrated in the literature by Favre et al. [2002] and Wheeler et al. [2000]. The validation results suggest that the NSAR model is not overparameterized, overfitted or poorly identified as such models have limited ability to interpolate or extrapolate. Therefore, the additional model complexity is shown to improve the representation of the rainfall regime.

[65] The spatial interpolation property of the model, demonstrated by the validation results, suggests that the number of model parameters indicated by the linear expression in section 3.2 is likely to be an overestimate as the number of calibration rain gauges increases. Further, the potential to relate spatial parameters to catchment topographic properties may also lead to a substantial reduction in the number of model parameters.

[66] An algorithm was presented which improves the efficiency and accuracy of the simulation of circular raincells arising according to a homogeneous Poisson process on an infinite plane, as required for the STNSRP model. Raincells are conditionally simulated at any distance outside an inner rectangular region, containing all locations at which the process may be sampled. The condition being that the raincells are large enough to affect the inner region. The algorithm was demonstrated to provide an accurate STNSRP process for a rectangular inner region and may be slightly more efficient than an alternative algorithm appropriate for a circular inner region [Leonard et al., 2006]. The derivation of the properties of the conditional process may be easily adapted to other geometric shapes and the algorithm may be relevant to other models using spatial or volumetric Poisson processes.

[67] The new NSAR methodology can provide simulated heterogeneous rainfall fields which represent unobserved variability at locations between the rain gauges used in the calibration. Therefore, process sampling may vary according

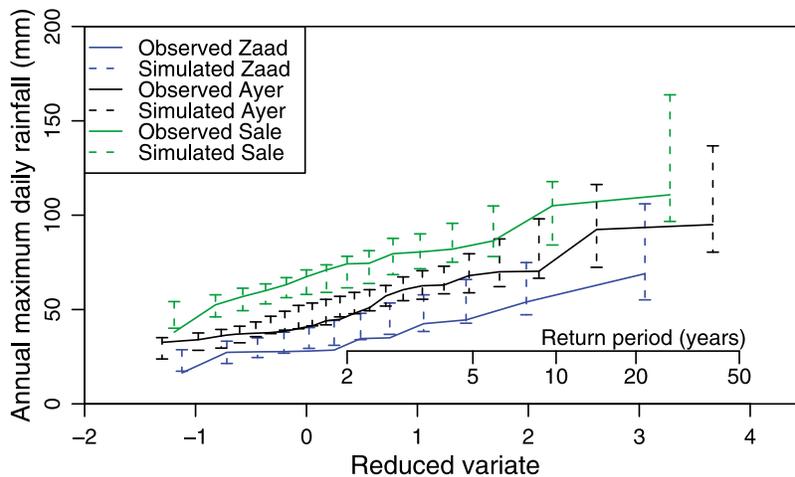


Figure 10. Gumbel plots comparing observed and simulated extreme daily rainfall for Zaad, Ayer, and Sale. The error bars show the 10th and 90th percentiles of the distribution simulated for each statistic.

to the application, e.g., as multisite time series or on a grid for distributed hydrological modeling. In contrast, existing stochastic rainfall models found to successfully represent both nonhomogeneous occurrence and amounts are limited to multisite and/or use discrete daily time steps. The new methodology is an extension of the NSRP model, a well-tested rainfall simulator that has been used in a wide variety of climatic zones, application types and for downscaling climate change scenarios [e.g., see *Burton et al.*, 2008; *Kilsby et al.*, 2007]. Therefore, it provides an important extension to the tools available for distributed hydrological modeling in large river catchments or where orographic effects are important.

Appendix A: Efficient and Accurate Simulation of Raincell Centers for a Rectangular Inner Simulation Region

[68] This appendix provides the derivation of the new finite algorithm for the simulation of circular raincells occurring with a homogeneous Poisson process over an infinite plane as part of the STNSRP simulation algorithm. A summary of the algorithm is provided in section 4.1.

[69] Consider raincells occurring across an infinite plane following a homogeneous Poisson process with density ρ . A finite inner rectangular simulation region R is chosen so as to contain all of the calibration rain gauges and all locations at which the simulation process is to be sampled (Figure A1). Thus, R has a defined area A_R and sides of length w and z . The outer simulation region is the rest of the infinite plane. Simulation of raincells with centers within R is straightforward: the number of centers is sampled as a discrete Poisson random variable with mean ρA_R ; the location of each center is sampled uniformly over the region; their radii are independent and exponentially distributed with parameter γ .

[70] Simply ignoring raincell occurrence in the outer region may considerably underestimate the true STNSRP process at locations where it is sampled, as raincells have spatial extent. Typically a finite approximate algorithm has therefore been used [see *Leonard et al.*, 2006] whereby the inner region is expanded to include a buffer zone of all locations where a generated raincell is likely to affect the original inner region. This approach suffers from two key drawbacks: (1) exponentially generated raincell radii can exceed any finite distance and (2) the increased simulation area significantly increases the computational effort required for the simulation [see *Leonard et al.*, 2006]. The buffer zone approach therefore under-simulates the STNSRP raincell occurrence process and is computationally inefficient, as it generates many irrelevant raincells.

[71] Here an alternative algorithm is derived and demonstrated whereby raincells are only simulated in the outer region if they are relevant to the inner region. This reduces the sampling of the infinite spatial raincell occurrence process to a finite, exact and efficient algorithm.

[72] Consider a thin rectangular “annulus” a distance x from R with thickness δx , within the outer region (Figure A1). The probability of a raincell center occurring in the annulus can be approximated as

$$P(\text{raincell centre in annulus}) \approx 2(w + z + 4x)\rho\delta x \quad (\text{A1})$$

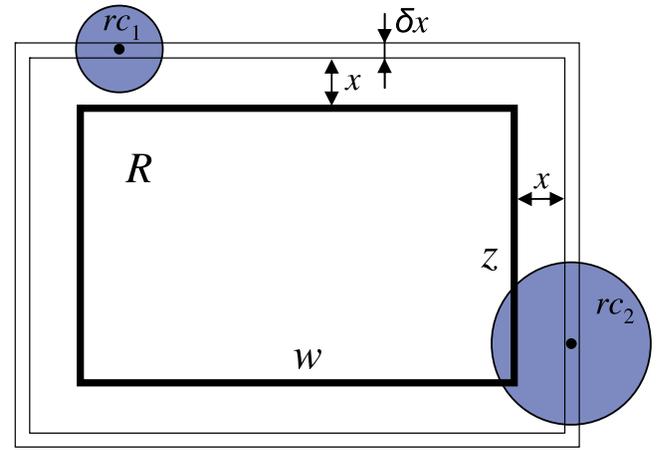


Figure A1. Illustration of a rectangular inner region R with area A_R over which homogeneous generation of circular raincells is required. A rectangular “annulus” lies at a distance x from R with a small width δx . Two circular raincells are illustrated: one, rc_1 , of no consequence to R and a second rc_2 , which affects R .

for terms of up to first order in δx . Such a raincell is considered relevant to R if its radius, an exponential random variable with parameter γ , is greater than x . The probability of such a raincell being relevant is therefore

$$\int_x^\infty \gamma e^{-\gamma x} dx = e^{-\gamma x}. \quad (\text{A2})$$

[73] The probability that a raincell will occur in the annulus and be relevant, p_r , is then the product of equations (A1) and (A2). An unknown monotonically increasing function of x is then considered, $y(x)$. Substituting

$$\delta x \approx \frac{dx}{dy} \delta y \quad (\text{A3})$$

into p_r we obtain

$$p_r \approx 2(w + z + 4x)\rho e^{-\gamma x} \frac{dx}{dy} \delta y. \quad (\text{A4})$$

[74] Should $y(x)$ be chosen to make p_r into a constant multiple of δy then equation (A4) simply describes a homogeneous Poisson process over y . This is satisfied if

$$(w + z + 4x)e^{-\gamma x} \frac{dx}{dy} = c_1 \quad (\text{A5})$$

for some constant c_1 . Separating variables and integrating provides the relationship that

$$c_2 - \frac{(w + z + 4x)e^{-\gamma x}}{\gamma} - \frac{4e^{-\gamma x}}{\gamma^2} = c_1 y \quad (\text{A6})$$

for another constant c_2 . Choosing $y(0) = 0$ and $y(\infty) = 1$ as boundary conditions on y implies that

$$c_1 = c_2 = \frac{\gamma(w + z) + 4}{\gamma^2} \quad (\text{A7})$$

and so $y(x)$ may be expressed as

$$y(x) = 1 - \left[1 + \frac{4x\gamma}{\gamma(w+z)+4} \right] e^{-\gamma x}. \quad (\text{A8})$$

[75] The chosen definition and boundary conditions of the function $y(x)$ allow it to be interpreted as the cumulative distribution function (CDF) of the distance of relevant raincells occurring in the outer region. Since a suitable expression has been found for $y(x)$, all relevant raincells in the outer region may be sampled as a homogeneous Poisson process on the interval $y \in [0,1]$ with a density, ρ_y , given by equation (A9), which derives from equations (A4), (A5), and (A7),

$$\rho_y = 2\rho c_1 = \frac{2\rho}{\gamma^2} (\gamma(w+z)+4). \quad (\text{A9})$$

[76] Thus, for the outer region, the total number of relevant raincells is a Poisson random variable with mean ρ_y . For each such raincell, the distance x of its center from the inner region may be sampled from the CDF given by equation (A8), its location is sampled uniformly from all points at this distance, and finally its radius is exponentially distributed, with parameter γ , conditional on it being greater than x .

[77] The flexibility of this derivation with respect to other geometric shapes may be demonstrated by comparison with the derivation of *Leonard et al.* [2006] who used a circular inner region. Consider a circular inner region with radius r and a thin annulus with radius $r+x$ with thickness δx . The equivalent expression, to equation (A4), for the probability, p_r , of a raincell occurring in the annulus and reaching the inner region is given by equation (A10). The substitution of $y(x)$ and the constant, c_3 , is also shown,

$$p_r \approx 2\pi\rho(r+x)e^{-\gamma x}\delta x \approx 2\pi\rho \left[(r+x)e^{-\gamma x} \frac{dx}{dy} \right] \delta y = 2\pi\rho c_3 \delta y. \quad (\text{A10})$$

[78] Solving for $y(x)$ as for the rectangular region gives

$$y(x) = 1 - \left[1 + \frac{x\gamma}{r\gamma+1} \right] e^{-\gamma x} \quad (\text{A11})$$

and

$$c_3 = \frac{1+r\gamma}{\gamma^2}. \quad (\text{A12})$$

[79] The number of relevant raincells in the outer region will then be a Poisson random variable with mean $\rho_y = 2\pi\rho c_3$. Each raincell's location may be sampled uniformly on the circumference of a circle with radius $(r+x)$ where the CDF of x is given by equation (A11). These two properties confirm the results of *Leonard et al.* [2006] (the parameter ν in paragraph 18 and equation (12) respectively).

Notation

A the $M \times M$ matrix $[a_{mn}]$, -.
 A_R the area of an inner region, km^2 .

δA a small area, km^2 .
 $a_{mn}(\gamma)$ the expected number of raincells affecting rain gauge m due to a unit density at a node n , -.
 c_i constants used for derivation, $i \in \{1,2,3\}$, -.
corr() the expected cross-correlation between two locations at a given distance apart: **corr_{ST}()** for the STNSRP process; **corr_{NSAR, m_1,m_2} ()** for the NSAR process; -.
 C_x the number of raincells occurring over a location \mathbf{x} , -.
 d a distance between two locations, km.
 $D(\cdot)$, $D_2(\cdot)$ objective functions used in model fitting, -.
 g an arbitrary rainfall statistic, varies.
 \hat{g} an observed sample estimate of g , varies.
 $g(\cdot)$ analytical estimate of the expected mean value of g for simulated rainfall in terms of model parameters, varies.
 $g_{\Lambda,m}(\cdot)$ specific form of $g(\cdot)$ for a specific rainfall model, $\Lambda \in \{\text{NSAR (NSAR)}; \text{STNSRP (ST)}; \text{NSRP (SS)}\}$, varies.
 g_s scaling term used to standardize each statistic g for model fitting, varies.
 h aggregation period used in the evaluation of a statistic, h.
 m an index describing a rain gauge or a node, -.
 m_1, m_2 rain gauges used to illustrate the model structure, -.
 n an index describing a node, -.
 M the number of nodes or equally, the number of calibration rain gauges, -.
PDH proportion of dry hours with less than 0.1mm accumulation, -.
PDD $_{\tau}$ proportion of dry days with less than τ mm accumulation, if τ is omitted then it is 0.2mm, -.
 \mathbf{P}_i parameter set generated by steps $i \in \{1,4,5\}$ of the NSAR fitting procedure.
 $\mathbf{P}_{i,m}$ parameter set generated by step $i \in \{2,3\}$ of the NSAR fitting procedure for rain gauge/node m .
 p_r probability that a raincell will occur in an annulus about the inner region and be relevant to the region, -.
 R finite inner rectangular region.
 r radius of a finite inner circular region, km.
VarD variance of the daily rainfall accumulation, mm^2 .
VarH variance of the hourly rainfall accumulation, mm^2 .
 w length of a side of the rectangular inner region, R, km.
 w_g weight applied to statistic g for model fitting, -.
 $w_n(\mathbf{x})$ the inverse square distance weighting of node n at location \mathbf{x} , -.
 x distance to an annulus around an inner region, km.
 δx thickness of a thin annulus around an inner region, km.
 \mathbf{x} an arbitrary location on the infinite simulation surface, km.
 \mathbf{x}_m the location of rain gauge m , km.
 $y(x)$ a monotonically increasing function of x , -.
 δy the change in y corresponding to the change δx , -.
 z length of a side of the rectangular inner region, R, km.
 β raincell waiting time parameter, h^{-1} .

- γ raincell radius parameter, km^{-1} .
- ε_m an error term used in the derivation of the fitting scheme for the raincell density field, $-$.
- ε the vector $[\varepsilon_m]$, $-$.
- η raincell duration parameter, h^{-1} .
- λ storm origin arrival rate parameter, h^{-1} .
- $\mu_h^m()$ expected h hour rainfall at rain gauge m as a function of model parameters, mm .
- ν expected number of raincells affecting a rain gauge parameter, $-$.
- $\nu_{i,m}$ expected number of raincells per storm at rain gauge m fitted by step $i \in \{3,5\}$ of the NSAR fitting procedure, $-$.
- \mathbf{v}_i the vector $[\nu_{i,m}]$ for $i \in \{3,5\}$, $-$.
- $\nu_m(\gamma, \rho)$ expected number of raincells per storm overlapping rain gauge m , $-$.
- ξ raincell intensity parameter, h/mm .
- ρ uniform spatial density of raincells parameter, km^{-2} .
- $\rho(\mathbf{x})$ spatially varying raincell density, km^{-2} .
- ρ_n raincell density fitted at the node n , km^{-2} .
- $\boldsymbol{\rho}$ the vector $[\rho_n]$, km^{-2} .
- ρ_i the best choice of ρ obtained in step $i \in \{4,5\}$, km^{-2} .
- ρ_y the expected number of relevant raincells occurring in an outer region, $-$.
- $\psi(\mathbf{x})$ intensity scaling field for spatially varying rainfall amounts, $-$.
- ψ_m intensity scaling field at rain gauge m , $-$.
- Ψ a vector of intensity scaling field values for a set of rain gauges, $-$.
- Ψ_i the value of Ψ fitted by step $i \in \{3,5\}$ of the fitting procedure, $-$.
- $\psi_{i,m}$ the intensity scaling field at rain gauge m fitted by step $i \in \{3,5\}$ of the NSAR fitting procedure, $-$.
- Ω a set of rainfall statistics used in fitting a rainfall model.

[80] **Acknowledgments.** This research was supported by the European Union FP6 Integrated Project AquaTerra (project 505428) under the thematic priority sustainable development, global change, and ecosystems and Hayley Fowler was supported by a NERC Postdoctoral Fellowship award (2006–2009) NE/D009588/1. The Gallego rain gauge records were supplied by a website supported by the Ministerio de Medio Ambiente, Madrid and the Confederación Hidrográfica del Ebro (<http://oph.chebro.es/>). The authors appreciate the assistance of a number of colleagues: Bruno Majone, University of Trento, who collected and organized the rainfall records; Ali Ford, Newcastle University, who converted the coordinate systems of the rain gauges' locations; and David Kuntz, University of Tübingen, who helped with the preparation of the map of the Gallego catchment. Finally the contributions of four anonymous reviewers and the editors are gratefully appreciated as this has helped to significantly improve this paper.

References

- Bárdossy, A., and E. J. Plate (1992), Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, *28*(5), 1247–1259, doi:10.1029/91WR02589.
- Barros, A., and D. P. Lettenmaier (1993), Dynamic modeling of the spatial distribution of precipitation in remote mountainous areas, *Mon. Weather Rev.*, *121*(4), 1195–1214.
- Barros, A., and D. P. Lettenmaier (1994), Dynamic modeling of orographically induced precipitation, *Rev. Geophys.*, *32*(3), 265–284, doi:10.1029/94RG00625.
- Buishand, T. A., and T. Brandsma (2001), Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling, *Water Resour. Res.*, *37*(11), 2761–2776, doi:10.1029/2001WR000291.
- 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. Softw.*, *23*, 1356–1369, doi:10.1016/j.envsoft.2008.04.003.
- Chandler, R. E., and H. S. Wheeler (2002), Analysis of rainfall variability using generalised linear models: A case study from the west of Ireland, *Water Resour. Res.*, *38*(10), 1192, doi:10.1029/2001WR000906.
- Cowpertwait, P. S. P. (1991), Further developments of the Neyman-Scott clustered point process for modeling rainfall, *Water Resour. Res.*, *27*(7), 1431–1438, doi:10.1029/91WR00479.
- Cowpertwait, P. S. P. (1994), A generalized point process model for rainfall, *Proc. R. Soc. London, Ser. A*, *447*, 23–37.
- 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 of rainfall for the Thames catchment, UK, *J. Hydrol.*, *330*, 586–595, doi:10.1016/j.jhydrol.2006.04.043.
- Cowpertwait, P. S. P., C. G. Kilsby, and P. E. O'Connell (2002), A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes, *Water Resour. Res.*, *38*(8), 1131, doi:10.1029/2001WR000709.
- Favre, A. C., A. Musy, and S. Morgenthaler (2002), Two-site modeling of rainfall based on the Neyman-Scott process, *Water Resour. Res.*, *38*(12), 1307, doi:10.1029/2002WR001343.
- Fowler, H. J., C. G. Kilsby, P. E. O'Connell, and A. Burton (2005), A weather-type conditioned multisite stochastic rainfall model for the generation of scenarios of climatic variability and change, *J. Hydrol.*, *308*, 50–66.
- Gerzabek, M. H., et al. (2007), The integrated project AquaTerra of the EU sixth framework lays foundations for better understanding of river-sediment-soil-groundwater systems, *J. Environ. Manage.*, *84*, 237–243, doi:10.1016/j.jenvman.2006.10.014.
- Hill, F. F., K. A. Browning, and M. J. Bader (1981), Radar and rain gauge observations of orographic rain over south Wales, *Q. J. R. Meteorol. Soc.*, *107*, 643–670.
- Hillier, F. S., and G. J. Liberman (1974), *Operations research*, 2nd ed., 800 pp., ISBN 0-8162-3856-1, Holden-Day, San Francisco.
- Kilsby, C. G., P. D. Jones, A. Burton, A. C. Ford, H. J. Fowler, C. Harpham, P. James, A. Smith, and R. L. Wilby (2007), A daily weather generator for use in climate change studies, *Environ. Modell. Softw.*, *22*, 1705–1719.
- Leonard, M., A. V. Metcalfe, and M. F. Lambert (2006), Efficient simulation of a space-time Neyman-Scott rainfall model, *Water Resour. Res.*, *42*, W11503, doi:10.1029/2006WR004986.
- Mehrotra, R., and A. Sharma (2007), A semi-parametric model for stochastic generation of multisite daily rainfall exhibiting low-frequency variability, *J. Hydrol.*, *335*, 180–193, doi:10.1016/j.jhydrol.2006.11.011.
- Mezghani, A., and B. Hingray (2009), A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin, *J. Hydrol.*, *377*, 245–260, doi:10.1016/j.jhydrol.2009.08.033.
- Natale, L., and E. Todini (1976), A stable estimator for linear-models. 1. Theoretical development and Monte-Carlo experiments, *Water Resour. Res.*, *12*(4), 667–671, doi:10.1029/WR012i004p00667.
- Onof, C., R. E. Chandler, A. Kakou, P. S. P. Northrop, H. S. Wheeler, and V. Isham (2000), Rainfall modeling using Poisson-cluster processes: A review of developments, *Stoch. Env. Res. Risk A.*, *14*(6), 384–411.
- Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London, Ser. A*, *410*, 269–288.
- Roe, G. H. (2005), Orographic precipitation, *Annu. Rev. Earth Planet. Sci.*, *33*, 645–671.
- Segond, M. L., C. Onof, and H. S. Wheeler (2006), Spatial-temporal disaggregation of daily rainfall from a generalized linear model, *J. Hydrol.*, *331*, 674–689.
- Srikanthan, R. (Sri), and G. G. S. Pegram (2009), A nested multisite daily rainfall stochastic generation model, *J. Hydrol.*, *371*, 142–153, doi:10.1016/j.jhydrol.2009.03.025.
- Stehlik, J., and A. Bárdossy (2002), Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, *J. Hydrol.*, *256*(1–2), 120–141.

- Wallis, J. R., N. C. Matalas, and J. R. Slack (1974), Just a moment!, *Water Resour. Res.*, *10*(2), 211–219, doi:10.1029/WR010i002p00211.
- Wheater, H. S., V. S. Isham, D. R. Cox, R. E. Chandler, A. Kakou, P. J. Northrop, L. Oh, C. Onof, and I. Rodriguez-Iturbe (2000), Spatial-temporal rainfall fields: Modeling and statistical aspects, *Hydrol. Earth Syst. Sci.*, *4*(4), 581–601.
- Wilby, R. L., O. J. Tomlinson, and C. W. Dawson (2003), Multisite simulation of precipitation by conditional resampling, *Climate Res.*, *23*(3), 183–194.
- Wilks, D. S. (1998), Multisite generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, *210*, 178–191.
- Wilks, D. S., and R. L. Wilby (1999), The weather generation game: A review of stochastic weather models, *Prog. Phys. Geog.*, *23*(3), 329–357.
-
- A. Burton, H. J. Fowler, C. G. Kilsby, and P. E. O’Connell, Water Resource Systems Research Laboratory, School of Civil Engineering and Geosciences, Cassie Building, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK. (aidan.burton@ncl.ac.uk)