

## Two-site modeling of rainfall based on the Neyman-Scott process

A.-C. Favre<sup>1</sup>

Institut National de la Recherche Scientifique, Eau Terre et Environnement, Université du Québec, Ste-Foy, Canada

A. Musy

Hydrology and Water Resources Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland

S. Morgenthaler

Basic Sciences Department, Swiss Federal Institute of Technology, Lausanne, Switzerland

Received 27 March 2002; revised 27 July 2002; accepted 27 July 2002; published 19 December 2002.

[1] On the basis of the Neyman-Scott process a model is proposed to generate rainfall at two sites simultaneously. This model takes into account the correlation, which is produced by generating bivariate correlated variates. The cross correlation between the sites has been computed in order to estimate the parameters. Finally, the model is validated for three pairs of stations situated in the Swiss Plateau region. *INDEX TERMS*: 1821 Hydrology: Floods; 1854 Hydrology: Precipitation (3354); 1869 Hydrology: Stochastic processes; *KEYWORDS*: Neyman-Scott model, rainfall generation, Poisson process, multisite model, bivariate generation

**Citation:** Favre, A.-C., A. Musy, and S. Morgenthaler, Two-site modeling of rainfall based on the Neyman-Scott process, *Water Resour. Res.*, 38(12), 1307, doi:10.1029/2002WR001343, 2002.

### 1. Introduction

[2] The design of hydraulic structures requires the determination of a given return period flood. This estimation should be as accurate as possible in order to optimize the construction costs. The classical approach to determine a value corresponding to a non exceeding probability is based on the frequency analysis. However if the available data length is too short the frequency analysis application is limited by the huge uncertainty obtained in extrapolating to high frequencies. A more recent approach consists in generating a precipitation record of a given length with a stochastic rainfall model. This synthetic rainfall is used as input for a rainfall-runoff model (physically-based or conceptual). A statistical analysis of the output, namely the obtained generated flow record, allows to determine the requested design value with the empirical distribution without using extrapolation [Favre, 1997]. Such a rainfall model is punctual and allows to reproduce the precipitation only at a given location. However the spatial variability of precipitation is significant even for small watersheds.

[3] In order to study rainfall on a whole region, it is necessary to develop models that can be used at multiple sites and may be calibrated from a network of rain gauges. In principle, it is possible to apply a fully continuous spatial-temporal model, such as the one developed by Northrop [1996], and to derive properties of the model at a discrete collection of points in space. In practice, however, this is complicated because such models are difficult to calibrate, especially in countries like Switzerland where one has strong orographic effects on rainfall generation process.

[4] An alternative approach consists in developing models that are discrete in the spatial dimension, and express the between-site interactions in a manner which reflects the underlying spatial-temporal structure of rainfall. Such models are called multisite. Generally the inter-site relationships in these models should depend on the distance between the examined sites and on topography, and should preserve the structure of single-site models for their marginal processes at each site.

[5] Cowpertwait [1994] generalizes the Neyman-Scott model for rainfall at a single site, in which storms arrive in a Poisson process, each storm generating a cluster of rain cells, every cell having a random duration and intensity. The model is generalized by allowing each cell to be one of  $n$  types. The duration of each cell is an exponential random variable that has expectation dependent on the cell type. The distribution of cell intensity is also dependent on the cell type, so that the generalized model provides a correlation between the intensities and durations of the rain cells.

[6] Cowpertwait [1995] presents and fits a relatively simple stochastic spatial-temporal model of rainfall in which the arrival times of rain cells occur in a clustered point processes. In the  $x$ - $y$  plane, rain cells are represented as discs; each disc having a random radius; the locations of the disc centers are given by a two-dimensional Poisson process. Multisite second order properties are derived and used to fit the model to hourly rainfall data taken from six sites in the Thames basin, UK.

[7] Kakou [1997] develops multisite models which are generalizations of those proposed by Cox and Isham [1994]. The basic concept is that there is a master point process of storm or rain cell origins which is decomposed into a number of subprocesses, each one corresponding to a different site.

[8] The main difficulty with the above models lies in their estimation of parameters. The fitting is highly complex, and parameters are supposed to be constant without strong

<sup>1</sup>Formerly at Swiss Federal Institute of Technology, Lausanne, Switzerland.

physical justification. Moreover radar measures are used to calibrate the models. Until now, in Switzerland only three radars are available. These new measurement techniques give more a qualitative insight into precipitation than a quantitative one, because of the orography (hidden precipitation) amongst other reasons. Therefore models which depend on radar-derived parameter estimates cannot be effectively applied in mountainous regions. Hopefully in the future radars will give a more insight into the precipitation process and will change the modeling substantially.

[9] We propose a model for two sites based on the Neyman-Scott process. We make the assumption that there is a master process giving rise to the storm at a given station. The correlation between the rainfall series is obtained by generating bivariate correlated random variates.

[10] Section 2 presents the Neyman-Scott Rectangular Pulses Model for a single site. Section 3 outlines the main statistical tools used to build the rainfall model. Section 4 is devoted to the model development. This model is calibrated and validated in section 5. Section 6 provides the major conclusions and the planned improvements.

## 2. Neyman-Scott Rectangular Pulses Model

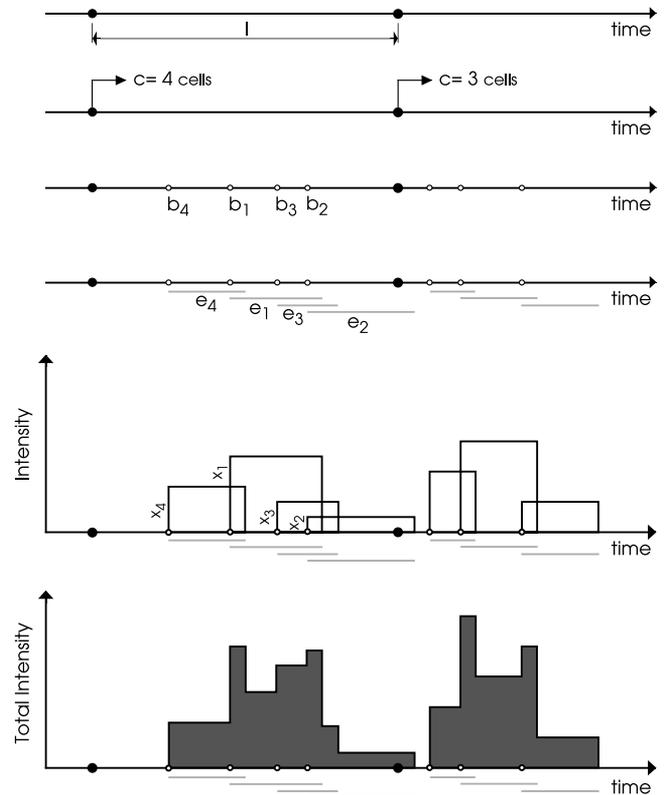
[11] The Neyman-Scott is a clustered point process model in time. This process is used to reproduce the rainfall in the following way: each rainfall event is originated by a triggering mechanism, the origin of the event (representing, for example, the arrival of a mid-latitude atmospheric front or more generally, a storm), that primes several elementary precipitation cells. The Neyman-Scott model is described by three independent elementary stochastic processes: (1) a process that defines the storms origin, (2) a process that determines the number of rain cells generated by each storm, and (3) a process that defines the cells origin. Storm origins are governed by a Poisson process with parameter  $\lambda$  (Neyman-Scott is a cluster Poisson process). At a point on the ground the storm is conceptualized as a random number  $C$  of rain cells. The distribution for  $C$  could be either geometric or Poisson. The cell origins are independently separated from the storm origin by distances which are exponentially distributed with parameter  $\beta$ , no cell origins being located at the storm origin. A rectangular pulse is associated independently with each cell origin, its duration and depth being independent. They are assumed to be exponentially distributed with parameter  $\eta$  and  $1/\mu_X$ , respectively. Figure 1 shows a schematic depiction of this process.

[12] The precipitation intensity at time  $t$ ,  $Y(t)$ , is given by the sum of the intensities of the individual cells active at time  $t$

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u) dN(t-u), \quad (1)$$

where  $X_u(\kappa)$  is the random depth of the pulse originating at time  $u$ , measured a time  $\kappa$  later and where  $\{N(t)\}$  counts occurrences in the Poisson process of pulse origins. An estimation method of the five parameters based on the method of moments has been proposed by Favre [1998] and produced unbiased estimators.

[13] Finally the Neyman-Scott model is generally used to generate series of rainfall at hourly time steps. It has been used rarely for below 1 hour temporal resolution [e.g., Calenda and Napolitano, 1999]. Practical applications of the model demonstrate that NSRPM is able to reproduce the



**Figure 1.** Construction of the Neyman-Scott rectangular pulses model. The procedure first generates the origin of the storms ( $l_i$ ), then the characteristics related to the rain cells, namely, their number ( $c_i$ ), position ( $b_i$ ), duration ( $e_i$ ), and depth ( $x_i$ ).

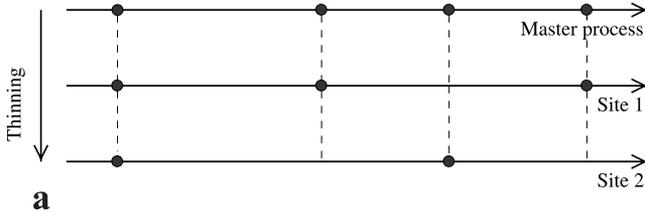
extremes of precipitation [Entekhabi et al., 1989, Figure 2; Velghe et al., 1994].

## 3. Statistical Tools

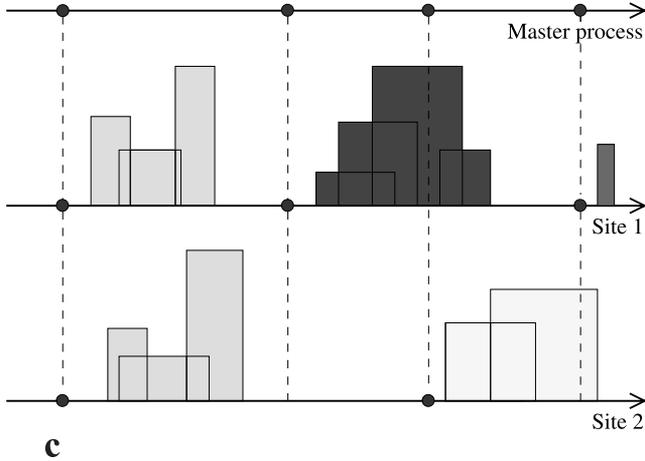
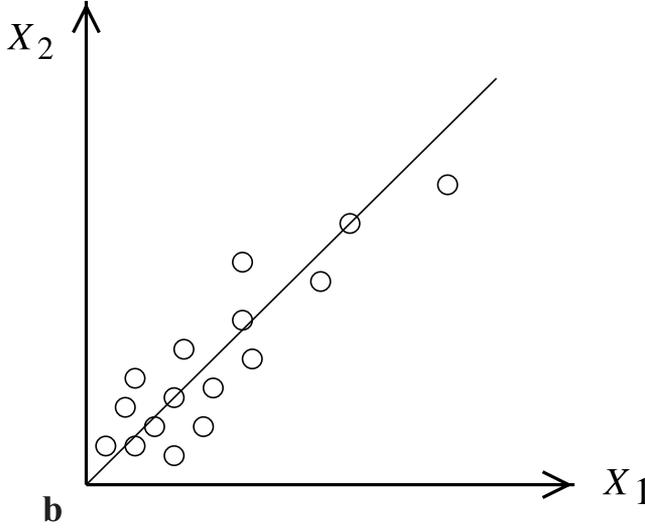
[14] This section presents bivariate random vectors generators used for the multisite model building, namely for the exponential and geometric distributions.

### 3.1. Bivariate Exponential Variates

[15] The bivariate distribution used in our approach should be mathematically tractable, easy to generate and be able to reproduce all kinds of positive dependence. The bivariate distributions of Lawrence and Lewis [1983] are particularly flexible in that sense. In their article, they introduce some models for constructing pairs of dependent exponential random variables as random-coefficient linear functions of pairs of independent exponential random variables. The Lawrence and Lewis models offer the following advantages: (1) their construction is very simple; (2) they require only two independent exponential random variables and one or two binary random variables for their construction; (3) there is a broad range of the attainable dependency, as measured linearly by the product-moment correlation, or monotonically by the Spearman correlation; (4) the models are analytically tractable so that, in some cases, closed form joint probability density functions and regressions can be obtained. For our purposes it is desirable to model two aspects of the joint distribution: one depend-



### Bivariate exponential family



**Figure 2.** Schematic depiction of the two-site Neyman-Scott rectangular pulses model: Master process and bivariate correlated generation.

ency measure, such as the product-moment correlation, and a probability statement, such as  $Pr\{W_1 > W_2\}$ . For these reasons the following two parameters model called (EP3) has been chosen:

$$(EP3) \begin{cases} W_1 = \tau_1 Z_1 + I_1 Z_2 & 0 \leq \tau_1 \leq 1 \\ W_2 = \tau_2 Z_2 + I_2 Z_1 & 0 \leq \tau_2 \leq 1, \end{cases} \quad (2)$$

where  $Z_1, Z_2$  are independent, identically and exponentially distributed random variables of unit means and  $I_1,$

$I_2 \sim \mathcal{B}(1, 1 - \tau_i)$ ,  $i = 1, 2$  so that  $Pr\{I_i = 0\} = 1 - Pr\{I_i = 1\} = \tau_i$ ,  $i = 1, 2$ . We use the following notation in this document

$$W \sim \mathfrak{E}(\tau_1, \tau_2) = \begin{pmatrix} \tau_1 & I_1 \\ I_2 & \tau_2 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}.$$

[16] The joint distribution of Lawrence and Lewis variates has been computed by Favre [2001].

### 3.2. Bivariate Geometric Variates

[17] The theory leading to the construction of the used bivariate geometric variates has been developed by Arnold [1967] using the additivity property of a distributions family.

[18] This property of closeness, under addition of independent random variables, has led to inversion-free methods for generating bivariate geometric random vectors, known as trivariate reduction methods [Cherian, 1941; David and Fix, 1961; Mardia, 1970; Schmeiser and Lal, 1982].

[19] Suppose we have  $G_1 \sim \mathcal{G}(\theta_1), G_2 \sim \mathcal{G}(\theta_2), G_3 \sim \mathcal{G}(\theta_3)$  three independent geometric random variables with parameters  $\theta_i$ ,  $i = 1, 2, 3$ . The bivariate geometric variates are then obtained taking the minimum in the following way

$$\begin{aligned} \tilde{G}_1 &= \min(G_1, G_2), \\ \tilde{G}_2 &= \min(G_2, G_3). \end{aligned} \quad (3)$$

We use the notation  $(\tilde{G}_1, \tilde{G}_2) \sim \mathfrak{G}(\theta_1, \theta_2, \theta_3)$ . The distribution of  $\tilde{G}_1$  and  $\tilde{G}_2$  can be directly obtained using the properties of the minimum function leading to  $\tilde{G}_1 \sim \mathcal{G}(\theta_1 + \theta_2 - \theta_1\theta_2)$ ,  $\tilde{G}_2 \sim \mathcal{G}(\theta_2 + \theta_3 - \theta_2\theta_3)$ . As  $\tilde{G}_1$  and  $\tilde{G}_2$  are both functions of  $G_2$ , dependence between these two variables is introduced. The measure of this dependence has been quantified through the correlation by Favre [2001].

$$\begin{aligned} \rho[\tilde{G}_1, \tilde{G}_2] &= \frac{\theta_2(1 - \theta_1 - \theta_3 + \theta_1\theta_3)}{(\theta_1 + \theta_2 + \theta_3 - \theta_1\theta_2 - \theta_1\theta_3 - \theta_2\theta_3 + \theta_1\theta_2\theta_3)} \\ &\quad \times \frac{1}{\sqrt{(\theta_1 - 1)(\theta_3 - 1)}}. \end{aligned} \quad (4)$$

[20] Suppose we want to generate bivariate geometric random variables  $\tilde{G}_1 = \mathcal{G}(\tilde{\theta}_1)$  and  $\tilde{G}_2 = \mathcal{G}(\tilde{\theta}_2)$  with a given correlation  $\rho_0 = \rho[\tilde{G}_1, \tilde{G}_2]$ , using the trivariate reduction method defined in the above section. The parameters  $\theta_1, \theta_2, \theta_3$  of the basic geometric variables used for generation can be obtained with equation (4) by the resolution of the following system of three equations in three unknown variables

$$\begin{cases} \tilde{\theta}_1 = \theta_1 + \theta_2 - \theta_1\theta_2 \\ \tilde{\theta}_2 = \theta_2 + \theta_3 - \theta_2\theta_3 \\ \rho_0 = \rho(\theta_1, \theta_2, \theta_3) \end{cases} \Rightarrow \begin{cases} \theta_3 = \frac{\tilde{\theta}_2(C\tilde{\theta}_1^2 + \tilde{\theta}_1 - 1)}{(\tilde{\theta}_1 - 1)(C\tilde{\theta}_1\tilde{\theta}_2 + 1)} \\ \theta_2 = \frac{\tilde{\theta}_2 - \theta_3}{1 - \theta_3} \\ \theta_1 = \frac{\theta_3(1 - \tilde{\theta}_1) + \tilde{\theta}_1 - \tilde{\theta}_2}{1 - \tilde{\theta}_2}, \end{cases} \quad (5)$$

where  $C = \rho_0 \frac{\sqrt{(\tilde{\theta}_1 - 1)(\tilde{\theta}_2 - 1)}}{\tilde{\theta}_1\tilde{\theta}_2}$ .

### 4. Model Development

[21] An extensive analysis of the rainfall series available on the Swiss Plateau exhibits the superposition of two processes [Favre and Overney, 1999]: the first one is a precipitation occurrence process active at all stations, while the second one is a rain cell generating process active at single stations with a stronger random component. Based on

**Table 1.** Statistical Structure of TS-NSRPM

Random Variable	Notation	Distribution	Parameters
Storms origin	$(L_1, L_2)$	Poisson	$\lambda_1, \lambda_2, \Lambda$
Number of cells	$(C_1, C_2)$	geometric	$\mu_C^{(1)}, \mu_C^{(2)}, \rho_C$
Cells position	$(B_1, B_2)$	exponential	$\beta_1, \beta_2$
Cells duration	$(E_1, E_2)$	exponential	$\eta_1, \eta_2, \nu_1, \nu_2$
Cells intensity	$(X_1, X_2)$	exponential	$\mu_X^{(1)}, \mu_X^{(2)}, \phi_1, \phi_2$

this finding, and following *Cox and Isham* [1994], we assume that storms arrive in the studied region in the form of a temporal Poisson process with rate  $\Lambda$ : this is called the master process. The evolution of rainfall at any given site is thought to be caused by a thinning of the master process, that is, a random selection with probability  $p$  of the storms that form the master process [Rényi, 1959]. A basic property of the Poisson process [see, e.g., *Resnick*, 1992] ensures that the storm origins at station  $i$  follow a Poisson distribution of rate  $\lambda_i = \Lambda p_i$ .

[22] At each individual site  $i$  the model is required to have the Neyman-Scott structure with five parameters  $\Theta_i = (\lambda_i, \eta_i, \beta_i, \mu_C^{(i)}, \mu_X^{(i)})$ , where  $\lambda_i = \Lambda p_i$ .

[23] It is quite obvious that the idea of a master process and thinning is not capable to take account of the whole complex structure of correlation between stations. To go further we must build a mathematically tractable model based on multivariate statistics in which the correlation between stations is handled through the generation of correlated random variables. We thus look for bivariate distributions with a given marginal and a given dependence structure. The previous section presented methods appropriate for this purpose.

[24] Our data analysis showed that the parameter  $\beta$ , which governs the position of cells relative to the storm origin is the most random component both from the temporal and spatial points of view (throughout the year, for different calibration periods and for one station to another). Thus the

estimation of this parameter is highly uncertain. Based on this remark, the basic version of the model assumes that  $B_1$  and  $B_2$ , the positions of the cells relative to the storm center at site 1 and 2, are independent random variates.

[25] The duration  $E$  and the intensity  $X$  of the cells both follow an exponential distribution. To generate the correlation between durations and intensities at two stations, the bivariate exponential variates (EP3 in equation (2)) have been chosen. Thus  $(X_1, X_2) \sim \mathcal{E}(\phi_1, \phi_2)$  and  $(E_1, E_2) \sim \mathcal{E}(\nu_1, \nu_2)$ . To model the correlation between the number of cells the trivariate reduction method, as defined in equation (3), has been used so that  $(C_1, C_2) \sim \mathcal{G}(\theta_1, \theta_2, \theta_3)$ . Note that according to equation (5), as soon as  $\rho_0$  is fixed, the parameters  $\theta_1, \theta_2, \theta_3$  are determined as functions of  $\mu_C^{(1)}$  and  $\mu_C^{(2)}$ . Figure 2 shows a simplified representation of the two-site model.

[26] The major theoretical advantage of the model is that the basic property of the thinning of Poisson processes mentioned above and the marginal distributional properties lead directly to the marginal distribution being a Neyman-Scott process at both sites.

[27] The proposed model is called TS-NSRPM (Two-Site Neyman-Scott Rectangular Pulses Model). The Table 1 summarizes the different parameters of the TS-NSRPM with bold lettering for those that are added in the passage from NSRPM to TS-NSRPM.

[28] In order to estimate the parameters of the TS-NSRPM, the cross covariance has been derived based on the theory of point processes, and is given in the following proposition: The cross covariance for two aggregated sub-processes is

$$C_{12}^{(h)}(\xi) = A_1 A_2 \left( K + \mathbf{1} \left\{ \frac{\eta_1}{\eta_2 \nu_1} > 1 \right\} K_1 + \mathbf{1} \left\{ \frac{\eta_1 \nu_2}{\eta_2} > 1 \right\} K_2 + \mathbf{1} \left\{ \frac{\eta_1}{\eta_2 \nu_1} > 1 \right\} \mathbf{1} \left\{ \frac{\eta_1 \nu_2}{\eta_2} > 1 \right\} K_3 \right). \quad (6)$$

**Figure 3.** Map of the ANETZ rainfall stations located in the Swiss Plateau.

**Table 2.** Estimated Moments (Hourly Mean, Hourly and Daily Variances, Hourly and Daily Lag 1 Autocovariances) Based on Rainfall Data at Each Station for January 1978–1999

	Payerne	Neuch.	SMA	Kloten	Tänikon	Gütt.
$\mu^{(1)}$	0.079	0.101	0.092	0.085	0.112	0.077
$s^{(1)}$	0.125	0.145	0.130	0.126	0.202	0.115
$s^{(24)}$	19.898	25.088	19.320	19.425	31.870	15.248
$s_{i,i+1}^{(1)}$	0.084	0.095	0.083	0.078	0.132	0.066
$s_{i,i+1}^{(24)}$	6.465	6.177	6.192	5.342	8.720	3.781

where

$$K = \frac{\nu_2^2(M_1 + M_2)}{\eta_2^2} \left( -2 \exp\left(\frac{-kh\eta_2}{\nu_2}\right) + \exp\left(\frac{-h(k-1)\eta_2}{\nu_2}\right) \right. \\ \left. + \exp\left(\frac{-h(1+k)\eta_2}{\nu_2}\right) \right) \\ + \frac{(N_1 + N_2 + N_3 + N_4)}{\beta_2^2} (-2 \exp(-\beta_2 kh) \\ + \exp(-\beta_2 h(k-1)) + \exp(-\beta_2 h(1+k))) \\ + \frac{O_1}{\eta_2^2} (-2 \exp(-\eta_2 kh) + \exp(-\eta_2 h(k-1)) \\ + \exp(-\eta_2 h(1+k))).$$

$$K_1 = \frac{P_1 + P_2}{\eta_1^2 \eta_2^2} \left( -2 \exp\left(\frac{-\eta_1 \eta_2 kh}{\eta_1 - \eta_2 \nu_1}\right) + \exp\left(\frac{-\eta_1 \eta_2 h(k-1)}{\eta_1 - \eta_2 \nu_1}\right) \right. \\ \left. + \exp\left(\frac{-\eta_1 \eta_2 h(1+k)}{\eta_1 - \eta_2 \nu_1}\right) (-\eta_1 + \eta_2 \nu_1)^2 \right),$$

$$K_2 = \frac{Q_1}{\eta_1^2 \eta_2^2} \left( -2 \exp\left(\frac{-\eta_1 \eta_2 kh}{\eta_1 \nu_2 - \eta_1}\right) + \exp\left(\frac{-\eta_1 \eta_2 h(k-1)}{\eta_1 \nu_2 - \eta_2}\right) \right. \\ \left. + \exp\left(\frac{-\eta_1 \eta_2 h(1+k)}{\eta_2 \nu_2 - \eta_2}\right) (\eta_1 \nu_2 - \eta_2)^2 \right),$$

$$K_3 = \frac{Q_2}{\eta_1^2 \eta_2^2} \left( -2 \exp\left(\frac{-\eta_1 \eta_2 kh}{\eta_1 \nu_2 - \eta_2}\right) + \exp\left(\frac{-\eta_1 \eta_2 h(k-1)}{\eta_1 \nu_2 - \eta_2}\right) \right. \\ \left. + \exp\left(\frac{-\eta_1 \eta_2 h(1+k)}{\eta_1 \nu_2 - \eta_2}\right) (\eta_1 \nu_2 - \eta_2)^2 \right).$$

[29] The constants building the above formula are given in appendix A. The proof of this proposition is given by Favre [2001].

## 5. Parameter Estimation and Validation

[30] Three groups of two stations located on the Swiss Plateau have been used to estimate the parameters and

**Table 3.** Estimated Moments (Hourly Mean, Hourly and Daily Variances, Hourly and Daily lag 1 Autocovariances) Based on Rainfall Data at Each Station for June 1978–1999

	Payerne	Neuch.	SMA	Kloten	Tänikon	Gütt.
$\mu^{(1)}$	0.136	0.130	0.193	0.155	0.194	0.153
$s^{(1)}$	0.609	0.471	0.748	0.600	0.838	0.593
$s^{(24)}$	43.985	39.157	60.206	46.781	64.171	46.114
$s_{i,i+1}^{(1)}$	0.203	0.203	0.293	0.208	0.304	0.215
$s_{i,i+1}^{(24)}$	4.271	4.967	7.050	5.478	9.108	4.464

**Table 4.** NSRPM Parameters at Each Station for January 1978–1999

	$\lambda$	$\eta$	$\beta$	$\mu_C$	$\mu_X$
Payerne	$8.74 \times 10^{-3}$	0.982	$5.77 \times 10^{-2}$	11.52	0.769
Neuch.	$1.32 \times 10^{-2}$	1.62	$9.76 \times 10^{-2}$	19.82	0.624
SMA	$1.23 \times 10^{-2}$	1.09	$5.76 \times 10^{-2}$	11.14	0.732
Kloten	$1.16 \times 10^{-2}$	1.40	$7.90 \times 10^{-2}$	13.73	0.748
Tänikon	$1.25 \times 10^{-2}$	1.18	$7.77 \times 10^{-2}$	12.44	0.855
Gütt.	$1.31 \times 10^{-2}$	1.46	$8.27 \times 10^{-2}$	10.42	0.829

validate the two-site Neyman-Scott rectangular pulses model: Payerne and Neuchâtel, Zürich SMA and Zürich-Kloten, Tänikon and Güttingen. Figure 3 shows a map with the location of the stations. The data used are from 1978 to 1999. First the single site parameters have been estimated for January and June, representative winter and summer months, with the procedure based on the method of moments. Tables 2 and 3 show the historical properties of these stations used for the estimation for January and June respectively while Tables 4 and 5 shows the corresponding parameters for the two months.

[31] For the two-site model, six additional parameters remain to be estimated, namely  $\Lambda$ ,  $\rho_C$ ,  $\nu_1$ ,  $\nu_2$ ,  $\phi_1$ ,  $\phi_2$ . The only realistic way to estimate the parameters is to use a pseudogeneralized method of moments. Alternative estimation procedure, such as maximum likelihood estimates, besides involving heavy mathematical complexity are not available and are not computable because the distribution function of the rainfall intensity for a given scale of aggregation is not known. A crucial factor in the parameters estimation procedure is the choice of the model properties to be included in the objective function. A selection of some joint properties of the two stations is required for a satisfactory representation of the spatial structure of the rainfall process. Also, properties at various levels of aggregations should be used in order to capture the temporal behavior of the process. Kakou [1997] initially used two lag 0 cross correlations, one at an aggregation level of 1 hour and one of higher level, for instance 6, 12, or 24 hours, but the obtained fits were rather poor. Subsequently, the latter statistic was replaced by the lag 1 cross correlation of hourly data, which provides additional information about the motion of rainfall event, and significantly improves the properties of the fitted processes. Thus, throughout the analyses presented here, we used the lag 0 and lag 1 cross correlation at hourly time step. Another important property to be included is the probability that two sites are simultaneously dry during a time interval of given length. We use this characteristic at hourly and daily levels in our procedure. As each couple consists of stations that are close together, it is assumed that  $\Lambda = \max(\lambda_1, \lambda_2)$ , and thus that

**Table 5.** NSRPM Parameters at Each Station for June 1978–1999

	$\lambda$	$\eta$	$\beta$	$\mu_C$	$\mu_X$
Payerne	$1.99 \times 10^{-2}$	3.52	$1.78 \times 10^{-1}$	5.58	4.31
Neuch.	$1.95 \times 10^{-2}$	2.10	$1.35 \times 10^{-1}$	5.27	2.66
SMA	$2.80 \times 10^{-2}$	2.62	$1.52 \times 10^{-1}$	5.78	3.12
Kloten	$2.29 \times 10^{-2}$	3.40	$1.56 \times 10^{-1}$	6.33	3.63
Tänikon	$2.56 \times 10^{-2}$	2.63	$1.20 \times 10^{-1}$	5.41	3.69
Gütt.	$2.35 \times 10^{-2}$	3.44	$1.87 \times 10^{-1}$	6.36	3.52

**Table 6.** TS-NSRPM Parameters of Three Couples of Stations for January 1978–1999

	Payerne Neuch.	Kloten SMA	Tänikon Gütt.
$\lambda$	$1.32 \times 10^{-2}$	$1.23 \times 10^{-2}$	$1.31 \times 10^{-2}$
$\rho_C$	0.90	0.90	0.90
$\phi_1$	0.11	0.98	0.31
$\phi_2$	0.95	0.04	0.45
$\nu_1$	0.22	0.17	0.04
$\nu_2$	0.84	0.92	0.86

the station with the biggest number of storms builds the master process. With this assumption we have that  $p_1 = \frac{\lambda_1}{\max(\lambda_1, \lambda_2)}$  and  $p_2 = \frac{\lambda_2}{\max(\lambda_1, \lambda_2)}$ . Computations of the cross correlation function (see equation (6)) show that in order to obtain sufficiently high values of correlation, it is necessary that  $\rho_C$  be large. Thus we set  $\rho_C = 0.9$  in the winter period, respectively  $\rho_C = 0.7$  for the summer period. The remaining estimation parameters procedure is divided into two steps: (1) Computation of equation (6) with  $h = 1$ ,  $k = 0$  and  $k = 24$  in order to fit the observed correlation values. Since two equations are used to estimate 4 parameters, several sets of parameters are able to reproduce the desired cross correlations. (2) Simulation with these parameters sets of precipitation series and computation of the probability that two sites are simultaneously dry at hourly and daily levels. The combination of these two steps leads to the choice of the best parameters set regarding the dryness probability. Note that in the second step, only the parameters  $\nu_1$  and  $\nu_2$  are in fact fitted as the probability of two sites being dry does not depend on  $\phi = (\phi_1, \phi_2)$ . Tables 6 and 7 show the obtained parameters for January and June respectively.

[32] A successful multisite model should be capable of describing the main between-site dependencies of the rainfall process. The adequacy of the fit is assessed via comparison between observed and estimated properties, both marginal and joint, that have not been used in the parameter estimation procedure. Tables 8 and 9 show the observed and estimated properties for Payerne and Neuchâtel.

[33] With reference to Tables 8 and 9, we notice that at all levels of aggregation, the probability of two sites dry and the cross correlation are estimated accurately. The same observations can be done for the two other pairs of stations. The cross correlations are slightly underestimated which is not surprising as we assumed that the position of the cells from the storm origin are independent. The probability that both sites are dry is slenderly overestimated. The results in summer are slightly worse than in winter: this can be due to the assumption that  $\Lambda = \max(\lambda_1, \lambda_2)$ , which is more constraining in summer. Moreover the convective precipitation in summer are more difficult to

**Table 7.** TS-NSRPM Parameters of Three Couples of Stations for June 1978–1999

	Payerne Neuch.	Kloten SMA	Tänikon Gütt.
$\lambda$	$1.99 \times 10^{-2}$	$2.80 \times 10^{-2}$	$2.56 \times 10^{-2}$
$\rho_C$	0.70	0.70	0.70
$\phi_1$	0.32	0.40	0.26
$\phi_2$	0.58	0.38	0.52
$\nu_1$	0.16	0.45	0.16
$\nu_2$	0.55	0.28	0.72

**Table 8.** Cross Properties (Probability of Both Sites Dry and Cross Correlation for Four Different Lags) Between Payerne and Neuchâtel for January<sup>a</sup>

Level of Aggregation		Probability of Both Sites Dry	Cross Correlation			
			Lag 0	Lag 1	Lag 2	Lag 3
1 h	$\mathcal{H}$	<b>0.814</b>	<b>0.652</b>	<b>0.580</b>	0.454	0.379
	$\mathcal{M}$	0.774	0.681	0.594	0.424	0.357
6 h	$\mathcal{H}$	0.684	0.786	0.427	0.223	0.160
	$\mathcal{M}$	0.725	0.710	0.372	0.206	0.128
12 h	$\mathcal{H}$	0.600	0.832	0.330	0.163	0.094
	$\mathcal{M}$	0.652	0.785	0.312	0.147	0.068
24 h	$\mathcal{H}$	<b>0.475</b>	0.851	0.290	0.095	0.114
	$\mathcal{M}$	0.514	0.813	0.243	0.041	0.109

<sup>a</sup> $\mathcal{H}$  are the observed values. Characteristics in **bold** have been used in the model fitting.

model due to their unexpected characteristics. In fact thunderstorms can happen very locally and it may rain at a station while not at the neighboring station even if they are close together.

## 6. Conclusion

[34] The spatial component of precipitation has been taken into account in the development of a two-site Neyman-Scott Rectangular Pulses Model. The idea of the proposed approach is to take into account the correlation between stations through the generation of bivariate correlated random variables. An application of the model for three couples of sites showed that it is able to reproduce both cross correlation at several levels of aggregation and lag, and probability of precipitation occurrence at several levels of aggregation. Two advantages of the proposed model are its simplicity and flexibility.

[35] There are several possibilities for future work. Considering that the cells position are correlated would further increase the flexibility of the model. More generally, a challenging approach would be to generalize the model for more than two sites. Coupling it with a rainfall-runoff model in hydrological applications, would lead to the quantification of the differences as compared to the inde-

**Table 9.** Cross Properties (Probability of Both Sites Dry and Cross Correlation for Four Different Lags) Between Payerne and Neuchâtel for June<sup>a</sup>

Level of Aggregation		Probability of Both Sites Dry	Cross Correlation			
			Lag 0	Lag 1	Lag 2	Lag 3
1 h	$\mathcal{H}$	<b>0.834</b>	<b>0.425</b>	<b>0.327</b>	0.228	0.172
	$\mathcal{M}$	0.812	0.438	0.335	0.214	0.118
6 h	$\mathcal{H}$	0.672	0.715	0.293	0.117	0.017
	$\mathcal{M}$	0.785	0.640	0.272	0.098	0.019
12 h	$\mathcal{H}$	0.569	0.784	0.193	0.005	0.019
	$\mathcal{M}$	0.640	0.669	0.140	0.012	0.000
24 h	$\mathcal{H}$	<b>0.417</b>	0.808	0.120	0.065	0.035
	$\mathcal{M}$	0.428	0.750	0.094	0.000	0.000

<sup>a</sup> $\mathcal{H}$  are the observed values. Characteristics in **bold** have been used in the model fitting.

pendent modelling of flows. Intuitively the extreme flows should be by the way better reproduced.

### Appendix A: Cross-Covariance Function

[36] The constants appearing in the cross-covariance function are the following:

$$A_1 = \lambda_1 p_2 E[C_1 C_2] \frac{\beta_1 \beta_2}{\beta_1 + \beta_2},$$

$$A_2 = E[X_1 X_2],$$

$$A_3 = \nu_1 \nu_2,$$

$$M_1 = -\frac{A_3 \nu_1 \nu_2^2}{(\eta_1 \nu_2 + \eta_2 \nu_1)(\eta_2 - \beta_2 \nu_2)},$$

$$M_2 = \frac{A_3 \nu_1 \nu_2^2}{(-\eta_1 + \eta_1 \nu_2 + \eta_2 \nu_1)(\eta_2 - \beta_2 \nu_2)},$$

$$N_1 = \frac{A_3 \nu_1 \nu_2}{(\eta_1 + \beta_2 \nu_1)(\eta_2 - \beta_2 \nu_2)},$$

$$N_2 = -\frac{\eta_2 \nu_1 (-\eta_2 + \beta_2 \nu_2 + A_3 \nu_2 \eta_2 - A_3 \nu_2 \beta_2)}{(\eta_2 - \beta_2)(\beta_2 \eta_2 \nu_1 + \eta_1 \eta_2 - \eta_1 \beta_2)(\eta_2 - \beta_2 \nu_2)},$$

$$N_3 = \frac{(-1 + A_3)(-1 + \nu_1 \nu_2) \eta_1 \eta_2}{(\beta_2 \eta_2 \nu_1 + \eta_1 \eta_2 - \eta_1 \beta_2)(-\beta_2 \eta_1 \nu_2 + \eta_2 \beta_2 + \eta_1 \eta_2)},$$

$$N_4 = -\frac{\eta_1 \nu_2 (-\eta_1 - \beta_2 \nu_1 + A_3 \nu_1 \eta_1 + A_3 \nu_1 \beta_2)}{(\eta_1 + \beta_2)(\eta_2 \eta_1 + \eta_2 \beta_2 - \beta_2 \eta_1 \nu_2)(\eta_1 + \beta_2 \nu_1)},$$

$$O_1 = -\frac{1}{\eta_2(\eta_2 - \beta_2)},$$

$$P_1 = \frac{(\eta_2 \nu_1 - \eta_1)(\eta_1 - \eta_1 \nu_2 - \eta_2 \nu_1 + A_3 \nu_1 \nu_2 \eta_2)}{(\beta_2 \eta_2 \nu_1 + \eta_1 \eta_2 - \eta_1 \beta_2) \eta_2 (-\eta_1 + \eta_1 \nu_2 + \eta_2 \nu_1)},$$

$$P_2 = -\frac{(-1 + A_3)(\eta_2 \nu_1 - \eta_1)(-1 + \nu_1 \nu_2)}{(\beta_2 \eta_2 \nu_1 + \eta_1 \eta_2 - \eta_1 \beta_2)(-\eta_1 + \eta_1 \nu_2 - \eta_2 + \eta_2 \nu_1)},$$

$$Q_1 = \frac{(\eta_2 - \eta_1 \nu_2)(\eta_1 \nu_2 - \eta_2 + \eta_2 \nu_1 - A_3 \nu_1 \nu_2 \eta_1)}{\eta_1(\eta_2 \eta_1 + \eta_2 \beta_2 - \beta_2 \eta_1 \nu_2)(\eta_1 \nu_2 - \eta_2 + \eta_2 \nu_1)},$$

$$Q_2 = \frac{(-1 + A_3)(\eta_2 - \eta_1 \nu_2)(-1 + \nu_1 \nu_2)}{(-\eta_1 + \eta_1 \nu_2 - \eta_2 + \eta_2 \nu_1)(-\beta_2 \eta_1 \nu_2 + \eta_2 \beta_2 + \eta_1 \eta_2)}.$$

[37] **Acknowledgments.** The financial support provided by the Swiss National Science Foundation (SNF) is gratefully acknowledged. The authors wish also to thank the anonymous reviewers of the paper. This research has been carried out at the Swiss Federal Institute of Technology while the main author was a Ph.D. student.

### References

- Arnold, B. C., A note on multivariate distributions with specified marginals, *J. Amer. Statist. Assoc.*, 62, 1460–1461, 1967.
- Calenda, G., and F. Napolitano, Parameter estimation of Neyman-Scott process for temporal point rainfall simulation, *J. Hydrol.*, 225, 45–66, 1999.
- Cherian, K. C., A bi-variate correlated gamma-type distribution function, *J. Indian Math. Soc. N. S.*, 5, 133–144, 1941.
- 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.
- Cox, D. R., and V. Isham, *Statistics For the Environment*, 2, *Water Related Issues*, vol. 2, pp. 3–18. John Wiley, New York, 1994.
- David, F. N., and E. Fix, Rank correlation and regression in a nonnormal surface, in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 177–197, Univ. Calif. Press, Berkeley, Calif., 1961.
- Entekhabi, D., I. Rodriguez-Iturbe, and P. S. Eagleson, Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation, *Water Resour. Res.*, 25, 295–302, 1989.
- Favre, A.-C., Modélisation stochastique de séries de précipitations, Application à la détermination des crues de fréquence rare, *Communication*, 5, 21–65, 1997.
- Favre, A.-C., The Neyman-Scott rectangular pulses model for precipitation: Parameter estimation and confidence interval determination, paper presented at International Meeting on Statistical Climatology, Am. Meteorol. Soc., Whistler, Canada, 1998.
- Favre, A.-C., Single and multi-site modelling of rainfall based on the Neyman-Scott process, Ph.D. thesis, Swiss Fed. Inst. of Technol., Lausanne, 2001.
- Favre, A.-C., and O. Overney, Investigations of the temporal and spatial properties of rainfall series: Some insights into rainfall modelling, in *Envirowater*, vol. II, pp. 20–37, IATE/HYDRAM, PPUR, Lausanne, Switzerland, 1999.
- Kakou, A., Point process based models for rainfall, Ph.D. thesis, Dep. of Stat. Sci., Univ. Coll. London, London, 1997.
- Lawrence, A. J., and P. Lewis, Simple dependent pairs of exponential and uniform random variables, *Oper. Res.*, 31(6), 1179–1197, 1983.
- Mardia, K. V., *Families of Bivariate Distributions*, Hafner, Darien, Conn., 1970.
- Northrop, P., Modelling and statistical analysis of spatial-temporal rainfall fields, Ph.D. thesis, Univ. Coll. London, London, 1996.
- Rényi, A., On the dimension and entropy of probability distributions, *Acta Math. Acad. Sci. Hungary*, 10, 193–215, 1959.
- Resnick, S., *Adventures in Stochastic Processes*, Birkhäuser Boston, Cambridge, Mass., 1992.
- Schmeiser, B. W., and R. Lal, Bivariate gamma random vectors, *Oper. Res.*, 30(2), 355–374, 1982.
- 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.

A.-C. Favre, INRS-ETE, Université du Québec, CP 7500, Ste-Foy, Québec (QC), G1V 4C7, Canada. (anne-catherine\_favre@inrs-ete.quebec.ca)

S. Morgenthaler, Basic Sciences Department, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland. (stephan.morgenthaler@epfl.ch)

A. Musy, Hydrology and Water Resources Laboratory, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland. (andre.musy@epfl.ch)