WATER RESOURCES RESEARCH, VOL. 22, NO. 8, PAGES 1316–1322, AUGUST 1986

Compatibility of Continuous Rainfall Occurrence Models With Discrete Rainfall Observations

EFI FOUFOULA-GEORGIOU[1]

Department of Civil Engineering, University of Washington, Seattle

PETER GUTTORP

Department of Statistics, University of Washington, Seattle

Inferences about the underlying rainfall generating mechanism made through a model fitted to the sampled rainfall series at a particular time scale are meaningful only if they are invariant under the time scale of measurement at which the fitting of the model is made. Following the work of I. Rodriguez-Iturbe et al. (1984), this assertion is tested by examining the compatibility of the Neyman-Scott (N-S) model for the continuous underlying rainfall with sampled realizations taken over intervals ranging from one hour to one day. The results suggest that the N-S model is inconsistent at different time scales, and therefore mathematical descriptions of rainfall using this model are inevitably limited to the time scale of measurement at which a description of rainfall is sought. These findings are quite general, in that they apply to any model with rainfall deposited instantaneously at times described by a N-S process, regardless of how the amounts associated with the events are distributed, and of the dependence structure of the amounts.

1. INTRODUCTION

Precipitation is a continuous intermittent process over space and time, which is usually recorded as cumulative amounts over fixed time intervals and at fixed locations. If one assumes the existence of an underlying rainfall generating mechanism evolving continuously in time, a challenging problem is to infer the mathematical structure of this unobserved process from its sampled realizations at discrete time intervals, such as hours or days. One way to approach this problem is to hypothesize a model for the continuous underlying process and estimate its parameters by comparing the theoretical properties of the derived discrete processes at different time scales with the ones estimated from the sampled realizations. If the parameters of the hypothesized model are not the same when estimated from data at different time scales, say, hourly and daily, this model is not consistent at both time scales, and little can be said about the underlying generating mechanism [*Rodriguez-Iturbe et al.*, 1984].

The importance of time scale considerations in modeling rainfall was recently pointed out by *Rodriguez-Iturbe et al.* [1984], who illustrated the dependence of the mathematical representation of rainfall on the time scale of measurement. They examined three models for the unobserved continuous time rainfall intensity process. Two of those were marked Poisson processes with different mechanisms for the process of marks: in one model the marks were rainfall volumes per event (Poisson white noise model), and in the other volumes and durations of events (Poisson rectangular pulses model). The third model was a marked Neyman-Scott (N-S) process where the marks were volumes per event (Neyman-Scott white noise model). By comparing the derived second-order properties of the accumulated rainfall amounts over different time scales with their empirical counterparts, they concluded that among the models studied the Neyman-Scott white noise model provided the best description of rainfall at both the hourly and daily time scales. More recently, *Valdes et al.* [1985] reexamined the time scale sensitivity of the above three rainfall models using temporal rainfall series generated from the space-time rainfall intensity model of *Waymire et al.* [1984]. They arrived at conclusions similar to those of *Rodriguez-Iturbe et al.* [1984] regarding consistency of models with respect to second-order statistics, but demonstrated that none of the models were able to preserve the statistics of extremes.

The present paper has two objectives. The first objective is to examine the time scale consistency of the Neyman-Scott process which has been suggested as a physically realistic model for short time increment rainfall occurrences. Our approach in addressing this objective is somewhat different from that of *Rodriguez-Iturbe et al.* [1984] and *Valdes et al.* [1985] in that it employs properties of the discrete time occurrence series and not of the cumulative rainfall amounts, thus avoiding assumptions regarding the correlation and distributional properties of the instantaneous rainfall amounts and cross-correlation properties between occurrences and amounts. The second objective is to present methods of estimating the continuous underlying N-S model from the discrete rainfall occurrence series and examine the sufficiency of second order statistics of the discrete occurrence or amounts process in identifying the underlying continuous model. The reader unfamiliar with point process terminology and methods will find the review paper of *Waymire and Gupta* [1981], and the paper of *Guttorp* [1986] useful in understanding the mathematical development presented in sections 2 and 3. Alternatively, the reader interested only in the results and their applications may skip directly to section 4.

The paper contains two parts. In the first part (which consists of sections 1 to 4) we study the compatibility of the N-S model for the underlying rainfall generating mechanism. We seek descriptions of the underlying rainfall occurrence process,

[1]Now at Department of Civil Engineering, Iowa State University, Ames.

Paper number 5W4301.
0043-1397/86/005W-4301$05.00

N_t, by means of the discrete time occurrence series $Z_k(\Delta)$, and not the cumulative rainfall amounts series $Y_k(\Delta)$. Thus our approach does not require assumptions about the distributional properties of the instantaneous rainfall amounts and the correlation properties of occurrences and amounts. The discrete time occurrence series $Z_k(\Delta)$ is defined as

$$Z_k(\Delta) = 1(N((k\Delta, (k+1)\Delta]) > 0) \qquad k \geq 0 \tag{1}$$

where $N(A)$ denotes the number of events of N_t falling into the set A, and $1(E)$ is 1 if E occurs and 0 otherwise. The only model hypothesized for N_t is a Neyman-Scott model, since, as is shown analytically in the next section, Poisson-based models are unlikely to be adequate when a significant dependence in the cumulative rainfall amounts at any time scale of interest exists. It will be shown that the parameters of the N-S model estimated from the series $Z_k(\Delta)$ at different time scales Δ ($\frac{1}{24}, \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{2}$, and 1 days) are not consistent, implying that the N-S model does not provide an adequate description for N_t compatible with sampled realizations at different time scales.

In the second part of this paper (which consists of sections 5 and 6) we have examined carefully the model and model fitting procedure of *Rodriguez-Iturbe et al.* [1984] and have made some key observations which lead to the conclusion that the N-S white noise model cannot, in general, be consistent at all time scales. In particular, it is shown that the parameter of the N-S model which is related to the dispersion of events in a cluster is uniquely determined by the rate of decay of the autocorrelation function of the cumulative rainfall amounts and therefore cannot be the same at all time scales. Furthermore, it is shown that the parameters of the final N-S model are highly sensitive to the assumption of the distribution of cluster sizes (Poisson or geometric), and therefore if possible, the selection of this distribution should be made on physical grounds and not simple mathematical convenience.

Finally, some suggestions on integration of our approach with the approach of *Rodriguez-Iturbe et al.* [1984] for the estimation of a continuous-time marked Neyman-Scott process which, although not consistent, preserves both the cumulative rainfall amounts and binary rainfall occurrence series at the time scale of interest, are made. This is especially important for the estimation of the N-S model from daily data where the second autocorrelation coefficient r_2 is usually insignificant and the model can be quite sensitive to the assumed value for r_2.

2. Mathematical Structure of the Underlying Rainfall Occurrence Process

To obtain some insight as to what classes of models may be appropriate for the underlying rainfall occurrence process N_t it is important to study the relation of the properties of N_t to the properties of an observed realization resulting from N_t, such as the cumulative amounts Y_k, at different time scales Δ. Below we compute the second order properties of Y_k for a general marked occurrence process N_t.

Let p_N and $p_{NN}(u)$ denote the rate and second-order product moment of a stationary continuous-time point process (N_t, $0 \leq t \leq T$). These properties are defined as

$$p_N = \lim_{dt \to 0} P\{dN_t = 1\}/dt \tag{2}$$

and

$$p_{NN}(u) = \lim_{dt, du \to 0} P\{dN_t = dN_{t+u} = 1\}/dt\, du \tag{3}$$

where $dN_t = N_{t+dt} - N_t$. The parameter $p_{NN}(u)$ measures the probability density of two events u time units apart. Let $U(\ )$ denote the rainfall intensity process. Then, if Y_i stands for the observed rainfall amount in the ith time interval of length Δ, $U(\ )$ is white noise, and X is distributed as $U(\Delta)$,

$$\text{Cov}\,(Y_1, Y_{k+1}) = E\int_0^{\Delta}\int_{k\Delta}^{(k+1)\Delta} U(t)U(s)\,dN_t\,dN_s - E^2(X)p_N{}^2\Delta^2 \tag{4}$$

Conditioning on $N(0, \Delta] = n$, $N((k\Delta, (k+1)\Delta]) = m$, the first term in (4) becomes

$$E(E^2(X)nm) = E^2(X)\int_0^{\Delta}\int_{k\Delta}^{(k+1)\Delta} p_{NN}(t-s)\,dt\,ds$$
$$= E^2(X)\Delta\int_{-\infty}^{\infty} h_\Delta(v - k\Delta)p_{NN}(v)\,dv$$

where

$$h_\Delta(s) = 1 - |s|/\Delta \qquad -\Delta \leq s \leq \Delta$$
$$h_\Delta(s) = 0 \qquad \text{otherwise} \tag{5}$$

[compare *Guttorp and Thompson*, 1984]. Noticing that

$$p_N{}^2\Delta^2 = \Delta\int_{-\infty}^{\infty} h_\Delta(v - k\Delta)p_N{}^2\,dv \tag{6}$$

we find that

$$\text{Cov}\,(Y_1, Y_{k+1}) = E^2(X)\Delta\int_{-\infty}^{\infty} h_\Delta(v - k\Delta)c_{NN}(v)\,dv \tag{7}$$

where $c_{NN}(u) = p_{NN}(u) - p_N{}^2$ is the covariance density of the point process N_t. In the case of a Poisson process, the covariance density c_{NN} is identically zero, whence the same is true regarding the covariance between amounts in different time intervals. Therefore if significant correlation of rainfall amounts at different time scales is found, a model with independent counts, such as the Poisson model, is not expected to give a good description of the continuous process N_t.

A similar computation shows that

$$\text{Var}\,(Y_1) = p_N\Delta\ \text{Var}\,(X) + V_\Delta E^2(X) \tag{8}$$

where $V_\Delta = \text{Var}\,(N_\Delta)$ is the variance-time curve of N_t evaluated at $t = \Delta$ days. Note from (7) and (8) how the covariance and variance of the rainfall amounts depend on the width Δ of the discretization interval.

The conclusion of *Rodriguez-Iturbe et al.* [1984] that Poisson based models did not provide adequate descriptions of N_t for the Denver data can now be reached by alternate computations. Assume, as in their case, that X_i are independently exponentially distributed, so that $E(X^2) = 2E^2(X)$. Solving (8) for V_Δ and substituting data from Table 1 of Rodriguez-Iturbe et al. for hourly ($\Delta = \frac{1}{24}$) and daily ($\Delta = 1$) time scales, we find that

$$V_1 - V_{1/24} = 40.906/E^2(X) - 1.76$$
$$E(X) = 2.125/p_N \tag{9}$$

(It should be emphasized that throughout this paper, Δ is in days, and hence for example, $\Delta = \frac{1}{24}$ corresponds to an hourly sampling interval.) Note that since $V_1 > V_{1/24}$, we must have $E(X) < 4.82$, and $p_N > 0.441$. Assuming a linear variance-time curve (this is typically the case asymptotically; compare *Cox and Isham* [1980, chapter 2]) we find from (9) that the slope of the variance time curve is $9.06p_N{}^2 - 1.76$, which for $p_N = 1.69$, the rate of occurrence of the N-S model found by Rodriguez-Iturbe et al. for Denver data, gives $V_t = 25.05t$. Notice that for a Poisson process $V_t = p_N t$, so in this case we would have $V_t = 1.69t$. Thus the data suggest substantial overdispersion, or clustering, relative to a Poisson model.

On the basis of the above results, we have examined only the Neyman-Scott model as a possible model for N_t. The parameters of this model have been estimated by comparing the theoretical first- and second-order properties of the discretized processes $Z_k(\Delta)$ with those of the observed series, at different time scales Δ. Estimation methods are discussed in the next section.

3. ESTIMATION OF CONTINUOUS-TIME POINT PROCESSES FROM DISCRETE OCCURRENCE SERIES

Let $(N_t, 0 \leq t \leq T)$ be the counting process of onset times of rainfall events (i.e., N_t is the number of rainfall events that started before time t). In order to fit (and test the fit of) models of the point process N_t, one needs exact records of onset times. Usually, such records are unavailable, and onset times must be inferred from observations of the amounts over fixed time intervals, such as 1 hour or 1 day. This has two drawbacks: the arrival onset times are unknown, so some rounding error is inevitable (discretization), and the actual number of events in each time interval is unknown, in that the presence only or absence of rainfall is recorded (clipping). The discretized and clipped time series $Z_k(\Delta)$ was defined in (1) in terms of the continuous point process N_t. In this section, methods of estimation of the parameters of N_t from properties of the $Z_k(\Delta)$ series are presented.

3.1. *Maximum Likelihood Estimators*

Below we develop the necessary theory to derive the likelihood function of the sequence $(Z_k, k = 1, \cdots, n)$ in terms of the zero-probability function

$$\zeta(A) = P(N(A) = 0) \tag{10}$$

of the underlying point process N_t. Here $N(A)$ counts the number of events of N_t that fall in the set A.

Let x and y be the strings of 0's and 1's. We combine strings by concatenation, so e.g., $x1y$ stands for the string that first coincides with x, then has a 1 followed by the string y. Let $x \cdot y$ denote all strings starting with x, then having an arbitrary symbol (i.e., either 0 or 1) and then the string y. The following trivial lemma will allow computation of the probability $P(z)$ of any string z, in terms of (A) evaluated at relatively simple sets A.

Lemma:

$$P(y1x) = P(y \cdot x) - P(y0x)$$

We use the lemma to reduce the given string z to only the symbols 0 and $\cdot$ by removing all 1's starting from the left. An example given in the appendix will make the procedure clear.

This procedure, albeit straightforward, will be fairly tedious for a typical data set. The resulting likelihood function will be a linear combination of the zero-probability function evaluated at unions of disjoint intervals. Expressions of $\zeta(A)$, where A is the union of two such intervals, are given in the work by *Guttorp* [1986], and the methods there are easily (but tediously) extended to more general sets of the required form.

3.2. *Method of Moments*

The statistical properties of $Z_k(\Delta)$ for several models for N_t have been computed by *Guttorp* [1986]. Here, only the properties of interest for the case of a Neyman-Scott process with geometrically distributed cluster sizes are given:

$$m_\Delta = E(Z_k)/\Delta = (1 - B) \tag{11}$$

$$\begin{aligned} c_k &= E(Z_l Z_{l+k})/\Delta^2 = B(D_k - B) \qquad k \geq 1 \\ c_k &= E(Z_l Z_{l+k})/\Delta^2 = B(1 - B) \qquad k = 0 \end{aligned} \tag{12}$$

and therefore

$$r_k = (D_k - B)/(1 - B) \qquad k \geq 1 \tag{13}$$

where

$$B = e^{-\lambda\Delta}\left[\frac{p}{1-(1-p)e^{-\beta\Delta}}\right]^{\lambda/\beta} \tag{14}$$

$$D_k = e^{-\lambda\Delta}\left[\frac{p+(1-p)e^{-\beta(k-1)\Delta}(1-e^{-\beta\Delta})}{1-(1-p)e^{-\beta\Delta}(1-e^{-\beta(k-1)\Delta}(1-e^{-\beta\Delta}))}\right]^{\lambda/\beta} \tag{15}$$

In the above equations, λ is the rate of the primary events (cluster centers); p is the parameter of a geometric distribution of the number of (secondary) events per cluster center; and β is the parameter of the exponential distribution of the time between secondary events and their cluster centers. Since the N-S model has three parameters, denoted by $\boldsymbol{\theta} = (\lambda, p, \beta)$, the expressions for m_Δ, r_1, and r_2 are adequate for the estimation.

Observe from (13)–(15) that the values of r_k are insensitive to the values of the parameter β, in that large changes in β may result in only small changes in r_k. Numerical simulation revealed that the derivatives of r_k with respect to λ are indeed very small (order of 10^{-3}) at the regions of β which seem to be of interest to rainfall modeling and therefore care must be exercised in solving these equations. Also notice that for appropriate combinations of parameter values, D_2 may take on values very close (or equal, in some cases) to D_1, thus making the system of equations underdetermined. In those cases, several N-S models give rise to discrete realizations $Z_k(\Delta)$ with the same first three moments. In view of the above, these equations were tested through simulation before being used for the rainfall data analysis. For example, given the insensitivity of the system to large changes in β, it is important to test the variability in β induced by the sampling variability in r_1 and r_2.

Five hundred continuous-time occurrences were generated from a Neyman-Scott model with known parameters $\boldsymbol{\theta}$ and were subsequently discretized at different time scales corresponding to 1, 2, 4, 6, 12, and 24 hours. The length of the sequences corresponds to twenty years of hourly data or a maximum of 3000 events, whichever is reached first. Using the discrete-time series, the parameters of the Neyman-Scott model were estimated. Table 1*a* shows the simulation results for $\boldsymbol{\theta} = (0.10, 0.05, 5.00)$, Table 1*b* for $\boldsymbol{\theta} = (0.10, 0.40, 1.00)$, and Table 1*c* for $\boldsymbol{\theta} = (0.30, 0.15, 17.00)$. The parameter vectors $\boldsymbol{\theta}$ were chosen so as to cover ranges of parameter values thought of being representative of rainfall data at different time scales (see next section).

From Tables 1*a*–1*c* it is observed that in general, the underlying N-S process is well identified and estimated from the observed discrete realizations. The failures (no solution of the equations) are many times caused by small negative empirical covariances c_1 or c_2, while only positive theoretical covariances of $Z_k(\Delta)$ are permissible when N_t is a N-S process (see

TABLE 1*a*. Mean Value and Standard Deviation of the Parameters (λ, p, β) of the Neyman-Scott Model Fitted to Discretized Data at Time Scales $\Delta = \frac{1}{24}, \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{2}$, and 1 Days

Δ	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$	1 ($f = 72$)
			$\hat{\lambda}$			
Mean	0.105	0.103	0.102	0.102	0.102	0.101
Standard	0.029	0.012	0.009	0.009	0.009	0.009
			$\hat{p}$			
Mean	0.050	0.051	0.051	0.052	0.052	0.080
Standard	0.013	0.007	0.009	0.011	0.017	0.056
			$\hat{\beta}$			
Mean	5.144	5.081	5.027	5.027	5.083	4.601
Standard	1.671	0.701	0.478	0.479	0.690	1.824

Results are based on 500 replicates; f is the number of failures. Population parameters are $\lambda = 0.10$, $p = 0.05$, and $\beta = 5.00$.

equation (12) in which D_1, $D_2 > B$, and $B > 0$ for all $\boldsymbol{\theta}$). Depending on the values of the parameters, problems of nonuniqueness and nonidentifiability are experienced at different regions of discretization time interval. All these problems are well explained by examining the Jacobian of the system and the theoretical values of m, c_1, and c_2 at the values of the population parameters.

4. Analysis of Rainfall Occurrence Data

The Neyman-Scott model was fitted to rainfall occurrence data from Denver, Colorado (1952–1972) and Seattle-Tacoma (Sea-Tac) International Airport, Washington (1965–1982). The discretized occurrence series at time scales Δ equal to $\frac{1}{24}, \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{2}$, and 1 days were created from hourly data, and the method of moments fitting procedure described in section 3 was used for the estimation. Each month was assumed a homogeneous period and was fitted separately. In order to avoid the abrupt transition from the one month to the next (end effects), the series were adjusted so that each month starts at the beginning of the first rainy period completely included in the month and ends at the end of the last dry period completely included in the month. Table 2 shows the parameters of the N-S model fitted to data from Denver, and Table 3 to data from the Sea-Tac Airport. It can be seen that the estimated parameters differ drastically at each time scale and they exhibit a very smooth trend, unlikely to have resulted from numerical problems.

Therefore we conclude that there is not a unique N-S model which, when discretized at time scales Δ, results in series $Z_k(\Delta)$ with the first three moments equal to the moments of the rainfall occurrence series at the appropriate time scales. Introduction of higher-order moments in the estimation has not been examined herein. However, it has been verified that autocorrelation coefficients at higher lags are adequately preserved.

TABLE 1*b*. Same as Table 1*a*, But With Population Parameters $\lambda = 0.10$, $p = 0.40$, and $\beta = 1.00$

Δ	$\frac{1}{24}$ ($f = 101$)	$\frac{1}{12}$ ($f = 69$)	$\frac{1}{6}$ ($f = 31$)	$\frac{1}{4}$ ($f = 4$)	$\frac{1}{2}$	1
			$\hat{\lambda}$			
Mean	0.205	0.147	0.095	0.095	0.100	0.100
Standard	0.035	0.040	0.034	0.024	0.007	0.004
			$\hat{p}$			
Mean	0.780	0.573	0.371	0.377	0.397	0.398
Standard	0.097	0.140	0.130	0.093	0.025	0.020
			$\hat{\beta}$			
Mean	13.420	3.526	1.150	1.006	1.009	1.018
Standard	9.582	2.441	0.726	0.441	0.168	0.094

TABLE 1*c*. Same as Table 1*a*, But With Population Parameters $\lambda = 0.30$, $p = 0.15$, and $\beta = 17.00$

Δ	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$ ($f = 106$)	$\frac{1}{2}$ ($f = 268$)	1 ($f = 275$)
			$\hat{\lambda}$			
Mean	0.303	0.302	0.302	0.300	0.296	0.291
Standard	0.027	0.017	0.015	0.015	0.016	0.021
			$\hat{p}$			
Mean	0.149	0.150	0.148	0.164	0.458	0.718
Standard	0.017	0.016	0.030	0.066	0.097	0.105
			$\hat{\beta}$			
Mean	17.087	17.091	17.349	16.994	6.419	2.038
Standard	2.088	1.306	2.407	5.503	2.373	1.202

5. Comments on the Neyman-Scott White Noise Model

The Neyman-Scott white noise model of interest consists of a Neyman-Scott model (with Poisson distributed cluster sizes) for the occurrence of the instantaneous bursts and an exponential distribution for the rainfall magnitudes associated with each burst. Here the necessary properties of the accumulated rainfall amounts $Y(\Delta)$, where Δ is the time scale of measurement in days, are given. For more information on this model, see the paper of *Rodriguez-Iturbe et al.* [1984].

Let $E(Y)$, Var (Y), and c_k denote the mean, variance, and lag-k autocovariance of the cumulative rainfall amounts Y. Then, *Rodriguez-Iturbe et al.* [1984] give

$$E(Y) = \Delta\lambda E(X)E(\nu) \tag{16}$$

$$\mathrm{Var}\,(Y) = \theta_1\Delta + \frac{2\theta_2}{\beta^2}(\beta\Delta - 1 + e^{-\beta\Delta}) \tag{17}$$

$$c_k = \frac{\theta_2}{\beta^2}(1 - e^{-\beta\Delta})^2 e^{-\beta(k-1)\Delta} \qquad k \geq 1 \tag{18}$$

where

$$\theta_1 = 2\lambda E^2(X)E(\nu) \tag{19}$$

$$\theta_2 = \lambda\beta E^2(X)E(\nu(\nu - 1))/2 \tag{20}$$

In the above equations the convention is made, as before, that λ and β will always be given in days^{-1} and Δ in days; ν denotes the number of events in a cluster. (A typographical error in equation (42) of *Rodriguez-Iturbe et al.* [1984], which has θ_2 not divided by 2, has been corrected here.) From the above equations, several observations are made as follows.

5.1. *Observation 1*

The autocorrelation function of the cumulative rainfall amounts $Y(\Delta)$ resulting from the integration of a Neyman-Scott white noise model has a Markovian dependence structure, which depends only on the cluster size and dispersion of events in a cluster.

Specifically, it can be shown that

$$r_k = r_1 e^{-\beta(k-1)\Delta} \tag{21}$$

where

$$r_1 = \frac{(1 - e^{-\beta\Delta})^2}{2\left[\dfrac{2\beta E(\nu)}{E(\nu(\nu - 1))}\Delta + (\beta\Delta - 1 + e^{-\beta\Delta})\right]} \tag{22}$$

TABLE 2. Parameters (λ, p, β) of the Neyman-Scott Model Fitted to Rainfall Occurrences at Denver, Colorado (Period 1952–1972)

Δ	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$	1
			January			
$\hat{\lambda}$	0.206	0.174	0.168	0.170	0.184	···
$\hat{p}$	0.049	0.077	0.131	0.189	0.249	···
$\hat{\beta}$	11.022	6.441	4.240	3.695	3.330	···
			February			
$\hat{\lambda}$	0.244	0.221	0.200	0.206	0.226	···
$\hat{p}$	0.033	0.039	0.097	0.136	0.070	···
$\hat{\beta}$	10.978	8.977	6.339	5.211	9.894	···
			March			
$\hat{\lambda}$	0.307	0.272	0.219	0.205	0.217	···
$\hat{p}$	0.026	0.032	0.075	0.087	0.141	···
$\hat{\beta}$	12.326	9.378	4.768	4.030	3.395	···
			April			
$\hat{\lambda}$	0.308	0.286	0.257	0.264	0.260	0.235
$\hat{p}$	0.026	0.049	0.100	0.133	0.272	0.372
$\hat{\beta}$	12.097	9.361	5.488	5.340	3.406	1.758
			May			
$\hat{\lambda}$	0.385	0.351	0.316	0.324	0.216	0.310
$\hat{p}$	0.040	0.073	0.126	0.182	0.312	0.350
$\hat{\beta}$	12.762	8.334	4.542	3.672	0.600	1.957
			June			
$\hat{\lambda}$	0.371	0.372	0.331	0.336	0.219	0.311
$\hat{p}$	0.148	0.205	0.350	0.456	0.450	0.573
$\hat{\beta}$	16.650	11.815	4.373	3.360	0.409	1.532
			July			
$\hat{\lambda}$	0.364	0.356	0.377	···	···	0.282
$\hat{p}$	0.198	0.292	0.424	···	···	0.586
$\hat{\beta}$	21.261	14.465	12.358	···	···	0.849
			August			
$\hat{\lambda}$	···	0.321	0.330	0.341	···	0.286
$\hat{p}$	···	0.325	0.447	0.278	···	0.482
$\hat{\beta}$	···	14.530	11.322	21.496	···	3.411
			September			
$\hat{\lambda}$	0.280	0.271	0.256	0.251	0.236	0.273
$\hat{p}$	0.048	0.090	0.206	0.249	0.402	0.353
$\hat{\beta}$	13.861	9.875	4.896	3.730	1.711	2.509
			October			
$\hat{\lambda}$	0.207	0.166	0.149	0.152	0.152	0.183
$\hat{p}$	0.027	0.041	0.081	0.151	0.192	0.389
$\hat{\beta}$	11.527	7.304	4.480	3.502	3.085	2.835
			November			
$\hat{\lambda}$	0.251	0.225	0.177	0.192	0.178	···
$\hat{p}$	0.029	0.060	0.125	0.134	0.111	···
$\hat{\beta}$	13.513	9.564	4.301	5.243	4.812	···
			December			
$\hat{\lambda}$	0.237	0.217	0.180	0.182	0.178	0.169
$\hat{p}$	0.060	0.094	0.153	0.150	0.345	0.381
$\hat{\beta}$	12.463	8.543	4.238	4.737	2.189	1.631

Here, Δ is the discretization interval in days.

is a function of only β and the first two moments of v, and not of the other two parameters of the model λ, and $E(X)$. Note that this result is contrary to the implication of *Rodriguez-Iturbe et al.* [1984] that the autocorrelation function depends on all the parameters of the model.

5.2. *Observation 2*

In view of the above observation, a simple estimation procedure of the N-S white noise model with Poisson distributed cluster sizes is proposed.

1. Fit an exponential function to the autocorrelation function (ACF) r_k, $k \geq 1$ of the amounts, or at least to the part of the ACF desired to be preserved. The parameter of the exponential function gives the value of the parameter β of the N-S white noise model.

2. Since r_1 depends only on β and $E(v)$, substitute in (22) the value of β and solve for $E(v)$.

3. Then, from $E(Y)$ and Var (Y), solve for the other two parameters to get

$$E(X) = \text{Var}\,(Y)/[2 + E(v)(\beta\Delta - 1 + e^{-\beta\Delta})/\beta] \tag{23}$$

$$\lambda = E(Y)/(E(X)E(v)) \tag{24}$$

Other one parameter distributions for v will yield slightly different algebraic expressions (see section 6).

5.3. *Observation 3*

The above estimation procedure will give a fit which preserves the mean, variance, and r_1 of the cumulative rainfall amounts exactly, and the ACF r_k, $k \geq 2$ to the desired degree of accuracy. Since, in general, it is not true that the rate of decay of the ACF of the cumulative rainfall amounts is the same for all discretization time intervals (e.g., 1 hour, 2 hours, $\cdots$, 24 hours), it is expected that the fitted N-S models at

TABLE 3. Parameters (λ, p, β) of the Neyman-Scott Model Fitted to Rainfall Occurrences at Seattle-Tacoma Airport (Period 1965–1982)

Δ	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$	1
			January			
$\hat{\lambda}$	1.204	0.953	0.811	0.724	0.582	0.443
$\hat{p}$	0.028	0.058	0.077	0.075	0.152	0.184
$\hat{\beta}$	18.008	9.041	5.796	4.754	2.124	1.014
			February			
$\hat{\lambda}$	0.917	0.688	0.538	0.528	0.469	0.441
$\hat{p}$	0.032	0.053	0.076	0.104	0.166	0.254
$\hat{\beta}$	15.814	8.237	5.238	3.506	2.042	1.132
			March			
$\hat{\lambda}$	0.876	0.678	0.520	0.463	0.405	0.403
$\hat{p}$	0.053	0.082	0.079	0.096	0.156	0.215
$\hat{\beta}$	16.058	8.155	5.066	3.596	2.051	1.772
			April			
$\hat{\lambda}$	0.813	0.671	0.534	0.471	0.469	···
$\hat{p}$	0.083	0.131	0.147	0.161	0.178	···
$\hat{\beta}$	18.274	8.975	5.088	3.756	3.860	···
			May			
$\hat{\lambda}$	0.471	0.419	0.372	0.340	0.281	0.295
$\hat{p}$	0.086	0.134	0.207	0.237	0.359	0.446
$\hat{\beta}$	19.395	12.000	6.706	4.433	1.645	1.676
			June			
$\hat{\lambda}$	0.513	0.433	0.388	0.350	0.265	0.249
$\hat{p}$	0.069	0.130	0.181	0.240	0.340	0.454
$\hat{\beta}$	20.056	9.875	6.596	4.121	1.087	0.832
			July			
$\hat{\lambda}$	0.288	0.249	0.247	0.231	0.224	···
$\hat{p}$	0.055	0.117	0.194	0.216	0.475	···
$\hat{\beta}$	15.540	8.641	6.916	5.815	3.691	···
			August			
$\hat{\lambda}$	0.564	0.410	0.308	0.305	0.242	0.289
$\hat{p}$	0.065	0.092	0.132	0.155	0.241	0.236
$\hat{\beta}$	17.797	8.212	3.537	3.630	1.564	3.908
			September			
$\hat{\lambda}$	0.664	0.524	0.394	0.357	0.319	0.280
$\hat{p}$	0.056	0.098	0.124	0.132	0.252	0.307
$\hat{\beta}$	19.268	9.398	4.696	3.828	1.791	1.018
			October			
$\hat{\lambda}$	0.622	0.561	0.443	0.392	0.374	0.360
$\hat{p}$	0.046	0.071	0.113	0.138	0.248	0.639
$\hat{\beta}$	14.908	10.326	4.741	3.580	2.539	0.409
			November			
$\hat{\lambda}$	1.017	0.805	0.645	0.607	0.573	0.142
$\hat{p}$	0.037	0.082	0.112	0.123	0.251	0.143
$\hat{\beta}$	19.329	9.029	5.090	4.496	2.445	0.103
			December			
$\hat{\lambda}$	1.245	0.948	0.813	0.749	0.630	0.681
$\hat{p}$	0.025	0.055	0.083	0.098	0.202	0.316
$\hat{\beta}$	18.790	8.452	5.374	4.224	1.934	1.772

Here, Δ is the discretization interval in days.

these time scales will not have the same parameter β. In fact, at the daily time scale, where for practical purposes $r_2 \simeq 0$, the system of equations is underdetermined, and an infinite number of N-S models can be found which preserve $E(Y)$, Var (Y), and r_1 exactly and have $r_2 \simeq 0$. For example, the models M_1 and M_2, where

$$M_1: \begin{bmatrix} \lambda = 0.090 \text{ days}^{-1} \\ E(X) = 1.186 \text{ mm} \\ E(v) = 19.91 \\ \beta = 3.5 \text{ days}^{-1} \end{bmatrix} \quad M_2: \begin{bmatrix} \lambda = 0.090 \text{ days}^{-1} \\ E(X) = 2.821 \text{ mm} \\ E(v) = 8.34 \\ \beta = 3.0 \text{ days}^{-1} \end{bmatrix}$$

have the same $E(Y)$, Var (Y), and r_1, and practically zero r_2 ($r_2 = 0.004$ for M_1, and 0.008 for M_2). The first model was fitted by *Rodriguez-Iturbe et al.* [1984] to the May 15–June 15 daily rainfall data from Denver, and the second was fitted by us by simply fixing the value of β to 3 days^{-1}. It is thus obvious that more information is needed to be able to discriminate one model from the other. A possible way to proceed would be through an integrated estimation approach in which the properties of both the Y_k and Z_k series are used. Although we have not pursued this idea in the present paper, simple calculations show that models M_1 and M_2 give series Z_k with properties close to each other, but not quite the same. For example, $m_\Delta(M_1) = 0.847$ days^{-1} and $m_\Delta(M_2) = 0.859$ days^{-1}, $c_0(M_1) = 0.130$, and $c_0(M_2) = 0.121$, which implies that at large time scales, such as days, difficulties may be encountered in model selection even with this integrated approach. For small time scales there is no problem, and the N-S parameters can be uniquely determined by the proposed estimation procedure.

6. Sensitivity of the N-S Model on the Assumed Distribution for the Cluster Size

Two versions of the Neyman-Scott model have been used for modeling rainfall. The one assumes a geometric distribution for the number v of events in a cluster [e.g., *Kavvas and Delleur*, 1981; *Ramirez and Bras*, 1985], and the other assumes a Poisson distribution [e.g., *Smith and Karr*, 1985; *Rodriguez-Iturbe et al.*, 1984]. In most of the times, the selection of the one distribution versus the other has been decided upon mathematical convenience. For example, *Smith and Karr* [1985] could only derive maximum likelihood estimators for the case of a Poisson distribution, while the authors of the present paper can only obtain closed form expressions for the properties of the $Z_k(\Delta)$ series for the case of a geometric distribution. The question we pose is to what extent the assumed distribution for v affects the parameters of the final N-S model. This is especially important in the cases where these parameters have been given physical interpretations by many authors.

The properties of the cumulative rainfall amounts $Y(\Delta)$ which have been computed by *Rodriguez-Iturbe et al.* [1984] for the case of a Poisson distribution for v can be easily computed for the case of a geometric distribution. For the approach to these computations, the reader is referred to the paper of Rodriguez-Iturbe et al., while here only the final expressions are given. $E(Y)$, Var (Y), and c_k are again given by (16)–(18), where θ_1 and θ_2 are now given as

$$\theta_1 = 2\lambda E^2(X)/p \tag{25}$$

$$\theta_2 = \lambda\beta E^2(X)(1-p)/p^2 \tag{26}$$

(instead of equations (19) and (20)), and where p is the parameter of the geometric distribution for v. It can be easily shown that the so derived amounts have again a Markovian dependence (equation (21)), where now r_1 is given as

$$r_1 = \frac{(1-e^{-\beta\Delta})^2}{2\left[\dfrac{p}{1-p}\beta\Delta + (\beta\Delta - 1 + e^{-\beta\Delta})\right]} \tag{27}$$

In the sequel we will use the notation NSG and NSP for the N-S model with geometric and Poisson distribution for v, respectively. Using the fitting procedure proposed in the previous section (observation 2), the NSG model is fitted to the hourly rainfall data of the period May 15–June 15 of the Denver station. This model is then compared to the NSP model fitted to the same data by Rodriguez-Iturbe et al.

As was discussed earlier, the parameter β of the N-S model is uniquely determined by the desired approximation of the ACF of the series Y. Keeping therefore the same (indeed very satisfactory) approximation as that of Rodriguez-Iturbe et al. (see Figure 7 in their paper), we have determined the value of $\beta = 4.8$ days^{-1}. Preservation of r_1 in (27) computes $p = 0.1042$. Then, solving (16) and (17) (where θ_1 and θ_2 are given in (25) and (26)), we find $E(X) = 0.416$ mm, and $\lambda = 0.53$ days^{-1}. Compare now the models

$$\text{NSP}: \begin{bmatrix} \lambda = 0.098 \text{ days}^{-1} \\ E(X) = 1.264 \text{ mm} \\ E(v) = 17.20 \\ \beta = 4.8 \text{ days}^{-1} \end{bmatrix} \quad \text{NSG}: \begin{bmatrix} \lambda = 0.53 \text{ days}^{-1} \\ E(X) = 0.416 \text{ mm} \\ E(v) = 9.597 \\ \beta = 4.8 \text{ days}^{-1} \end{bmatrix}$$

where NSP is the one given by Rodriquez-Iturbe et al. (Table 1 in their paper). Obviously, the assumed form of the distribution of the cluster size drastically affects the parameters of the fitted N-S model.

7. Conclusions

In this paper the issue of compatibility of the continuous-time Neyman-Scott (N-S) occurrence process N_t, with sampled discrete occurrence series $Z_k(\Delta)$ at time scales ranging from 1 hour to 1 day, has been studied. Our results agree with those of *Valdes et al.* [1985] and show that the instantaneous N-S model is not a consistant model for rainfall series recorded over different sampling intervals. These findings are quite general in that they apply to any model with rainfall deposited instantaneously at times described by a Neyman-Scott process, regardless of how the amounts associated with the events are distributed, and of the dependence structure of the amounts. The implications of this time scale inconsistency are twofold. First, a N-S model fitted to a particular time scale should not be used for extrapolations at other time scales, since, for example, the model fitted to hourly rainfall amounts does not preserve the daily amounts. Second, the N-S model does not provide an adequate description of the underlying rainfall generating mechanism, and thus no physical meaning should be attached to its parameters.

Several observations regarding the N-S white noise model have been made. First, the autocorrelation function of the rainfall amounts series resulting from integration of the N-S white noise model has a Markovian dependence structure depending only on the parameters of the clusters, and not the parameter of the instantaneous rainfall amounts. This observation leads to a simple fitting procedure. Second, the assumed distribution (Poisson or geometric) for the cluster size plays an important role in the final N-S model. Therefore since selection of the one distribution versus the other does not seem possible based solely on physical considerations, it does not seem possible to attach a physical meaning to the model parameters, and the N-S process must be interpreted as a descriptive and time scale–dependent model, rather than a prescriptive one.

APPENDIX: EXAMPLE DEMONSTRATING THE LIKELIHOOD OF Z_k IN TERMS OF THE ZERO PROBABILITY FUNCTION OF N_t

Let the series Z_k consist of the string $z = 01001$. Then,

$$P(z) = P(01001) = P(0 \cdot 001) - P(00001)$$

$$P(0 \cdot 001) = P(0 \cdot 00 \cdot) - P(00000)$$

$$P(00001) = P(0000 \cdot) - P(00000)$$

so

$$P(z) = P(0 \cdot 00 \cdot) - P(0 \cdot 000) - P(0000 \cdot) + P(00000)$$

In terms of $\zeta(A)$ we get (compare the computations in the work by *Guttorp* [1986]

$$P(z) = \zeta((0, \Delta]U(2\Delta, 4\Delta]) - \zeta((0, \Delta]U(2\Delta, 5\Delta]) - \zeta((0, 4\Delta]) + \zeta((0, 5\Delta])$$

Expressions for $\zeta(A)$ where A is the union of two such intervals are given in the work by *Guttorp* [1986], and the methods there are easily (but tediously) extended to more general sets of the required form.

The resulting likelihood function is a function of the parameters of the continuous-time model. The method of maximum likelihood calls for maximizing this function. In general, this optimization problem may be quite involved. The Poisson process case, however, is quite simple. Since

$$\zeta\left(\bigcup_1^m (k_i\Delta, (k_i + l_i)\Delta)\right) = \exp\left(-\rho\sum_1^m l_i\right)$$

provided that $k_i + l_i \leq k_{i+1}$, we get

$$l(\rho) = \log P(z; \rho) = N_n \log(1 - e^{-\rho\Delta}) - (n - N_n)\rho\Delta \qquad \text{(A1)}$$

so the maximum likelihood estimate $\hat{\rho}$ is the solution to $l(\rho) = 0$, or

$$\hat{\rho} = \frac{1}{\Delta} \log(1 - \bar{N}) \qquad \text{(A2)}$$

where $\bar{N} = N_n/n$. This is the same estimator as that obtained by the method of moments, equating the observed moment $\bar{N}$ to its expected value $\Delta m_P(\Delta)$.

Acknowledgments. This work had partial support from SIAM Institute for Mathematics and Society (SIMS) and from NSF through grant MCS-8302573. This research was performed while the second author was visiting the Department of Statistics, University of British Columbia. Computer resources were provided by the College of Engineering, University of Washington.

REFERENCES

Cox, D. R., and V. Isham, *Point Processes*, Chapman and Hall, London, 1980.

Guttorp, P., On binary time series obtained from continuous time point processes describing rainfall, *Water Resour. Res.*, *22*(6), 897–904, 1986.

Kavvas, M. L., and J. W. Delleur, A stochastic cluster model for daily rainfall sequences, *Water Resour. Res.*, *17*(4), 1151–1160, 1981.

Ramirez, J., and R. L. Bras, Conditional distributions of Neyman-Scott models for storm arrivals and their use in irrigation scheduling, *Water Resour. Res.*, *21*(3), 317–330, 1985.

Rodriguez-Iturbe, I., V. K. Gupta, and E. Waymire, Scale considerations in the modeling of temporal rainfall, *Water Resour. Res.*, *20*(11), 1611–1619, 1984.

Smith, J. A., and A. F. Karr, Statistical inference for point process models of rainfall, *Water Resour. Res.*, *21*(1), 73–79, 1985.

Valdes, J. B., I. Rodriguez-Iturbe, and V. K. Gupta, Approximations of temporal rainfall from a multidimensional model, *Water Resour. Res.*, *21*(8), 1259–1270, 1985.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 2, A review of the theory of point processes, *Water Resour. Res.*, *17*(5), 1273–1285, 1981.

Waymire, E., V. K. Gupta, and I. Rodriguez-Iturbe, A spectral theory of rainfall intensity at the meso-β scale, *Water Resour. Res.*, *20*(10), 1453–1465, 1984.

E. Foufoula-Georgiou, Department of Civil Engineering, Iowa State University, Ames, IA 50011.

P. Guttorp, Department of Statistics, University of Washington, Seattle, WA 98195.

(Received September 16, 1985;
revised May 6, 1986;
accepted May 7, 1986.)