A SIMULATION MODEL FOR INTERMITTENT PROCESSES

by

J. Kelman[x]

SYNOPSIS

A general model for synthesizing intermittent data is introduced. The basic assumption is that the intermittent process results from censoring a non-intermittent continuous valued process. Classical techniques for modelling time persistence in this latter process can then be exploited. Also, the latter process admits immediate known extensions to multivariate situations. Examples related to daily streamflow and to movement of single particles in sand channels are presented.

x) Assistant Professor COPPE-UFRS, Rio de Janeiro, Brazil (former Ph.D. student at Colorado State University, Fort Collins, Colorado, USA).

INTRODUCTION

The need for generating hydraulic data in the study of complex water resources problems is recognized by most engineers. Indeed, explicit solutions are rare; frequently, the only way to extract some probabilistic information about the performance of a system is to measure its response to a set of synthetic traces. This is the so called expertimental (or Monte Carlo) method.

This paper presents a technique for generating traces of an intermittent stochastic process. Here, a stochastic process $\{X_t\}$ is defined to be intermittent if, for any t:

i) $P(X_t < \gamma) = 0$

ii) $0 < P(X_t = \gamma) < 1$

iii) $P(X_t = x) = 0$, for $x > \gamma$.

Thus, $\{X_t\}$ is a mixed process in the sense that it has a positive probability of taking on the discrete value γ (called truncation point), and also the continuum of values greater than γ. Many processes studied in water resources fit the above description, as for example the three following cases: streamflow of rivers that occasionally go dry; daily rainfall; and step length of a bedload particle.

The ensuing sections will cover the following topics:

i) the conceptual model which supports the technique of synthetic data generation,
ii) the estimation of the parameters of the model,
iii) some comments on the generation of multivariate samples,
iv) an example related to daily streamflow and another related to single particle movement in alluvial channels.

THE MODEL

It is hypothesized that an intermittent process, say $\{C_t\}$, is the result of a censoring procedure applied on a continuous process, say $\{M_t\}$. In other words $C_t = M_t$ if $M_t > \gamma$, and $C_t = \gamma$ if $M_t \leq \gamma$. No physical justification is intended for $\{M_t\}$. In fact it is an imaginary process which one can hopefully model more easily than the intermittent process.

Therefore, the problem is shifted from advancing a $\{C_t\}$ model to proposing a $\{M_t\}$ model. The latter ought to have a time persistence mechanism built in, as well as the capability of coping with multivariate case. For the sake of simplicity

the normal autoregressive lag-one model is adopted. With this choice, the marginal distribution $\{M_t\}$ is normal, and therefore the values of $\{C_t\}$ which are greater than the truncation point will have a truncated normal distribution. Real data may not be well fitted by the latter distribution; however, greater versatility of the model can be obtained by further hypothesizing that $\{C_t-\gamma\}$ undergoes a power transformation before coming out as a component of the observable $\{X_t\}$ process.

Fig. 1 synthesizes the proposed filtering procedure. It "routes" independent standard normal pulses to end up as an intermittent, and serially dependent, time series.

Figure 1.

where

$$\{\xi_t\} \text{ is i.i.d.} \sim N(o,1)$$

$$\{M_t: m_t = \mu + \rho(m_{t-1} - \mu) + \sigma\ (1 - \rho^2)^{1/2}\ \xi_t\}$$

$$\{C_t: c_t = \gamma I_{(-\infty,\gamma)}(m_t) + m_t I_{(\gamma,\infty)}(m_t)\}$$

$$\{X_t: x_t = (c_t - \gamma)^{1/\alpha} + \gamma\}$$

Filter 1 introduces time persistence, filter 2 censors the data, and filter 3 increases the possibility of obtaining a good fit of the marginal distribution.

$(C_t - \gamma)^{1/\alpha}$ will have a power transformed truncated normal distribution. Its p.d.f. is given by

$$f_X(x) = \frac{\alpha x^{\alpha-1}\ \phi((x^\alpha - \mu)/\sigma)}{\sigma\ \Phi(\mu/\sigma)}\ I_{(0,\infty)}(x) \qquad (2)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the p.d.f. and the c.d.f. of the standard normal. Fig. 2 displays serveral p.d.f. graphs when $\mu = 0$ and $\sigma = 1$.

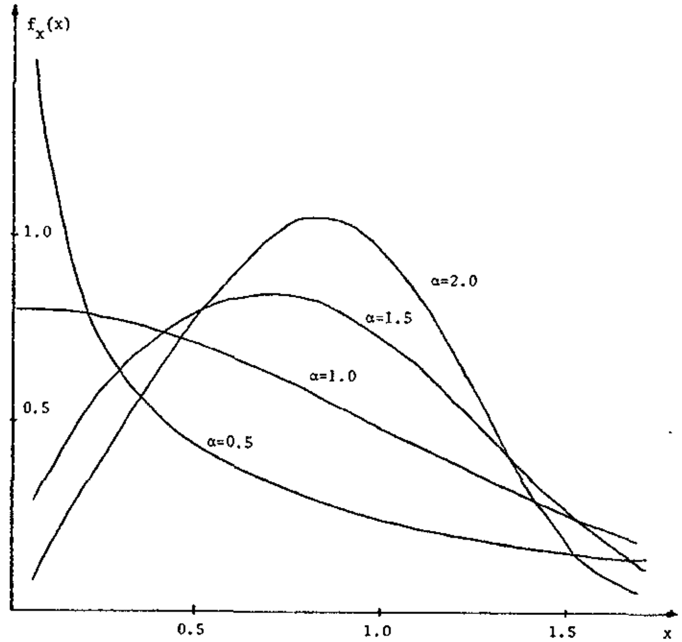

Figure 2. Probability Density Function of the Power Transformed Truncated Normal for $\mu = 0$ and $\sigma = 1$

THE ESTIMATION PROCEDURE

Given a time series $x_1, x_2, x_3, \ldots$ of some intermittent process, a method must be found to estimate the parameters μ, σ, ρ, and α. Among several alternatives, the maximum likelihood estimation procedure is selected because of its large sample properties. These properties allow setting up tests for relevant questions, as, for example, the stationarity and/or serial independence of data. Also, the asymptotic distribution of such estimators are known. These features, however, will not be

dealt with in this paper.

For random samples, and for α fixed equal to 1, estimators for μ and σ are available in the literature, see for example (1), However, when the observations are not independent, as is the case here, the estimation procedure is understandably more cumbersome. The likelihood function, evaluated in a straightforward way, would include a m fold integral of the multivariate normal p.d.f. for each run of m observations equal to γ. As is well known, this integral is not analytically available for $m > 3$. In order to avoid this difficulty, it is assumed that the pairs (X_1, X_2), (X_3, X_4), (X_5, X_6),... are independent, and the corresponding likelihood is to be maximized. This author's experience with generated time series, i.e. in situations when the "population" values are known, supports the accuracy of the results yielded by the above simplifying approximation.

The estimation problem then boils down to evaluating the parameters of a bivariate distribution. Suppose a sample (x_t, x_{t+1}), $t = 1,2,\ldots m$ is available. Without loss of generality, let $\gamma = 0$. Define the three following events:

$$A_1 = \{X_t = 0,\ X_{t+1} = 0\}$$
$$A_{2t} = \{X_t = x_t,\ X_{t+1} = y_t\}$$
$$A_{3t} = \{X_t = z_t,\ X_{t+1} = 0\} \cup \{X_t = 0,\ X_{t+1} = z_t\}$$
$$0 < x_t,\ y_t,\ z_t.$$

Assume that for the sample at hand the events A_1, A_{2t}, A_{3t} occur respectively n_1, n_2, and n_3 times. The likelihood function is then

$$L(\mu,\sigma,\rho,\alpha) = \frac{n!}{n_1!\ n_2!\ n_3!} P(A_1))^{n_1} \prod_{t=1}^{n_2} P(A_{2t}) \prod_{t=1}^{n_3} P(A_{3t}) \qquad (3)$$

For $(U,V)' \sim N\left[\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \begin{pmatrix}\sigma^2 & \sigma^2\rho\\ \sigma^2\rho & \sigma^2\end{pmatrix}\right]$,

$$P(A_1) = P(U < a,\ V < a) = \int_{-\infty}^{0}\int_{-\infty}^{0} f_{U,V}\ (u,v)\ du\ dv, \qquad (4)$$

where

$$f_{U,V}(u,v) = \frac{1}{2\pi\sigma^2(1-\rho^2)^{1/2}} \exp\left\{\frac{-1}{2(1-\rho^2)} Q\right\} \text{ and}$$

$$Q = \left(\frac{u-\mu}{\sigma}\right)^2 - 2\rho\left(\frac{u-\mu}{\sigma}\right)\left(\frac{v-\mu}{\sigma}\right) + \left(\frac{v-\mu}{\sigma}\right)^2$$

Similarly,

$$P(A_{2t}) = f_{U,V}(x_t^\alpha, y_t^\alpha)\, du\, dv = J\left[\frac{u,v}{x,y}\right] f_{U,V}(x_t^\alpha, y_t^\alpha)\, dx\, dy$$

where the Jacobian

$$J\left[\frac{u,v}{x,y}\right] = \begin{vmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{vmatrix} = \begin{vmatrix} \alpha x_t^{\alpha-1} & 0 \\ 0 & \alpha y_t^{\alpha-1} \end{vmatrix} = \alpha^2 (x_t y_t)^{\alpha-1}$$

Therefore,

$$P(A_{2t}) = \alpha^2 (x_t y_t)^{\alpha-1} f_{U,V}(x_t^\alpha, y_t^\alpha)\, dx\, dy \qquad (5)$$

Finally,

$$P(A_{3t}) = f_U(z_t^\alpha)\, du \int_{-\infty}^{0} f_{V/U}\,(v/u = z_t^\alpha)\, dv =$$

$$= f_u(z_t^\alpha)\, du \int_{-\infty}^{0} \frac{1}{\sigma(1-\rho^2)^{1/2}}\, \phi\left[\frac{v - \rho\, z_t^\alpha - u(1-\rho)}{\sigma\,(1-\rho^2)^{1/2}}\right] dv$$

$$P(A_{3t}) = \frac{\alpha\, z_t^{\alpha-1}}{\sigma}\, \phi\left[\frac{z_t^\alpha - \mu}{\sigma}\right] \Phi\left[\frac{-\rho\, z_t^\alpha - \mu(1-\rho)}{\sigma\,(1-\rho^2)^{1/2}}\right] dz \qquad (6)$$

The random variables $\frac{U-\mu}{\sigma}$ and $\frac{V-\mu}{\sigma}$ might be expressed as

$$U = \rho^{1/2} W_1 + (1-\rho)^{1/2}\, W_2$$

$$V = \rho^{1/2} W_1 + (1-\rho)^{1/2}\, W_3 \qquad (7)$$

where W_1, W_2, W_3 are independent standard normal. From Eqs. (4) and (7)

$$P(A_1) = \int_{-\infty}^{\infty} \phi(t) \left[\Phi\left[\frac{-\mu - \rho^{1/2}\sigma t}{\sigma(1-\rho)^{1/2}}\right]\right]^2 dt \qquad (8)$$

From Eqs. 3, 5, 6, and 8, after dropping the subscripts:

$$\log L = LL\ (\mu,\sigma,\rho,\alpha) = C + n_1 \log \int_{-\infty}^{\infty} \phi(t) \left[\Phi\left[\frac{-\mu - \rho^{1/2}\sigma t}{\sigma\,(1-\rho)^2}\right]\right]^2 dt$$

$$+ (2n_2 + n_3) \log \frac{\alpha}{\sigma} - n_2\left[\frac{\log(1-\rho^2)}{2} + \frac{\mu^2}{(1+\rho)\sigma^2}\right]$$

$$+ \sum^{n_2} \left[(\alpha-1)\log(xy) + \frac{2\mu(1-\rho)(x^\alpha + y^\alpha) - (x^{2\alpha} + y^{2\alpha}) + 2\rho(xy)^\alpha}{2(1-\rho^2)\sigma^2}\right]$$

$$+ \sum^{n_3} \left[(\alpha-1)\log z + \log \Phi\left[\frac{-z^\alpha - \mu(1-\rho)}{\sigma\,(1-\rho^2)^{1/2}}\right] + \log \phi\left[\frac{z^\alpha - \mu}{\sigma}\right]\right] \qquad (9)$$

where C is a constant.

The estimates $\hat{\mu}$, $\hat{\sigma}$, $\hat{\rho}$, and $\hat{\alpha}$ ought to be found in such a way that the likelihood function, or its logarithm, is maximum at this particular point. A numerical procedure, the Newton-Raphson algorithm, will be adopted:

$$H\Delta = D \qquad (10)$$

where

H is the Hessian matrix of the LL function, namely

$$H = \begin{bmatrix} \partial LL/\partial\mu^2 & \partial^2 LL/\partial\mu\partial\sigma & \partial^2 LL/\partial\mu\partial\rho & \partial^2 LL/\partial\mu\partial\alpha \\ & \partial^2 LL\partial\sigma^2 & \partial^2 LL/\partial\sigma\partial\rho & {}^2 LL/\partial\sigma\partial\alpha \\ & & \partial^2 LL/\partial\rho^2 & \partial^2 LL/\partial\rho\partial\alpha \\ & & & \partial^2 LL/\partial\alpha^2 \end{bmatrix}$$

and

$$\Delta' = \left[\mu_{old} - \mu_{new},\ \sigma_{old} - \sigma_{new},\ \rho_{old} - \rho_{mew},\ \alpha_{old} - \alpha_{new}\right]$$

and

$$D' = \left[\partial LL/\partial\mu,\ \partial LL/\partial\sigma,\ \partial LL/\partial\rho,\ \partial LL/\partial\alpha\right]$$

The first and second derivatives of LL, needed to evaluate Eq. 10, are furnished in the appendix.

When the analyst wants to produce traces of several dependent time series (multivariate case), the cross-correlations ought to be taken into consideration. For this situation the following simplified estimation procedure is proposed:

a) For each "station" (time series) find the marginal parameters μ_j, σ_j, ρ_j, α_j according to algorithm (10). If ℓ is the number of stations, $j = 1, 2, \ldots, \ell$

b) Find each lag-zero cross-correlation coefficient ρ_{jk}, $1 \le j < k \le \ell$, using only the data from station j and k.

Suppose that for each pair of stations a sample (x_{tj}, x_{tk}), $t = 1, 2, \ldots, m$ is available. "t" is the time index and "j", "k" are the station indexes.

Assume the sample space is divided into the four events

$$A_1 = \{X_{tj} = 0,\ X_{tk} = 0\}$$

$$A_{2t} = \{X_{tj} = x_{tj},\ X_{tk} = y_{tk}\}$$

$$A_{3t} = \{X_{tj} = x_{tj},\ X_{tk} = 0\}$$

$$A_{4t} = \{X_{tj} = 0,\ X_{tk} = y_{tk}\}$$

$$0 < x_{tj},\ y_{tk}$$

If the above events occurred respectively n_1, n_2, n_3, and n_4 times, the likelihood function is

$$L(\rho_{jk}) = \frac{n!}{n_1!\, n_2!\, n_3!\, n_4!} P(A_1)^{n_1} \prod_{t=1}^{n_2} P(A_{2t}) \prod_{t=1}^{n_3} P(A_{3t}) \prod_{t=1}^{n_4} P(A_{4t}) \tag{11}$$

Following a procedure analogous to the univariate case, the problem boils down to

$$\max_{\rho_{jk}} \Bigg\{ n_1 \log \int_{-\infty}^{-\hat\mu_j/\hat\sigma_j} \int_{-\infty}^{-\hat\mu_k/\hat\sigma_k} \frac{1 \exp}{2\pi\,(1-\rho_{jk}^2)^{1/2}} (x_j^2 - 2\rho_{jk}x_jx_k + x_k^2)\, dx_j dx_k - \frac{n_2 \log(1-\rho_{jk}^2)}{2} - \frac{1}{s(1-\rho_{jk}^2)} \sum^{n_2} \Bigg[\left(\frac{x_j - \hat\mu_j}{\hat\sigma_j}\right)^2 - 2\rho_{jk}\left(\frac{x^{\alpha}_j - \hat\mu_j}{\hat\sigma_j}\right)\left(\frac{y^{\alpha}_k - \hat\mu_k}{\hat\sigma_k}\right) + \left(\frac{y^{\alpha}_k - \hat\mu_k}{\hat\sigma_k}\right)^2 \Bigg] + \sum^{n_3} \Bigg[\log \Phi\left(\frac{-\hat\mu_k - \rho_{jk}\hat\sigma_k/\hat\sigma_j (x^{\alpha}_j - \hat\mu_j)}{\hat\sigma_k \;\; 1-\rho_{jk}^2} \right) \Bigg] + \sum^{n_4} \Bigg[\log \Phi\left(\frac{-\hat\mu_j - \rho_{jk}\hat\sigma_j/\hat\sigma_k (y^{\alpha}_k - \hat\mu_k)}{\hat\sigma_j\,(1-\rho_{jk}^2)^{1/2}} \right) \Bigg] \Bigg\} \tag{12}$$

DATA GENERATION

Once the parameters are estimated, the generation of synthetic traces is accomplished simply by following the stepwise procedure illustrated in Fig. 1.

In the multivariate case the deviate $\xi_{t,j}$; $j = 1, 2, \ \ ,\ell$ are not independent. One way of generating $\xi_{t,j}$ is by the use of:

$$\underline{\xi}_t = A\underline{\eta}_t \tag{13}$$

in which A is a $\ell x \ell$ matrix and $\underline{\eta}_t$ is a $\ell x1$ vector of independent standard normal deviated. Therefore, the covariance matrix associated with $\underline{\eta}_t$ is $I\ell$, the $\ell x \ell$ unit matrix. Consequently, the covariance matrix associated with $\underline{\xi}_t$ is AA'. On the other hand linear auto-regressive equations for stations "j" and "k" are:

$$M_{t,j} = \mu_j + \rho_j(M_{t-1,j} - \mu_j) + \sigma_j\,(1-\rho_j^2)^{1/2}\xi_{t,j} \qquad (14)$$

$$M_{t,k} = \mu_k + \rho_k(M_{t-1,k} - \mu_k) + \sigma_k\,(1-\rho_k^2)^{1/2}\xi_{t,k} \qquad (15)$$

Multiplying Eqs. 14 and 15 and finding the expected values, one gets:

$$\mathrm{corr}(\xi_{t,j}, \xi_{t,k}) = \frac{\rho_{j,k}(1-\rho_j\rho_k)}{\left\{(1-\rho_j^2)(1-\rho_k^2)\right\}^{1/2}} \qquad (16)$$

Hence the (j,k) element of matrix AA', $j \neq k$, is given by Eq. 16. The diagonal elements, of course, are unity. Several methods are available for finding a matrix A when AA' is given; (2) contains a straightforward one.

It whould be pointed out that with this approach the higher-than-lag-zero cross-correlations of the $\{M_t\}$ historical process will not be reproduced by the synthetic traces.

EXAMPLES

The validity of the model can not be established in general terms. *In fact no claim is made about its universality. It* is merely suggested that some intermittent time series may well be simulated through the use of the proposed scheme. This can be best realized with the help of a couple of examples, which follow.

The first example is related to simulation of daily streamflow. Many attempts have been made to develop models that would enable the production of synthetic traces of this stochastic process. The motivation for these efforts is found on the need of better designs, as well as operation policies, for water resources systems, particularly with respect to the features related to flood control. The weird characteristics of the process, however, has imposed severe limitations on the success of these attempts. A review of material related to this subject might be found in (3).

Hydrologists tend to agree that success on stochastically modeling daily runoff can only be achieved through the embodiment of some knowledge about the physical processes that cause runoff. To start with, it should be recognized that the hydrograph rising limb is due mostly to factors external to the

watershed. In other words, it is due to the sources that feed the watershed. On the other hand, the lowering limb, for a given recession period, is governed mostly by the emptying of the watershed. It is, therefore, easier to deterministically explain the latter process than the former one. In a forthcoming paper a model will be presented that assumes the runoff increments (flow on day t subtracted from flow on day t+1) belong to either of two mutually exclusive intermittent processes. For the time being only the positive increments will be studied.

If $\{D_t\}$ represents the daily streamflow process, the limited objective here is, then, to model the process $\{X_t\}$ defined as:

$$X_t = \begin{cases} D_t - D_{t-1}, & \text{for } (D_t - D_{t-1}) > 0 \\ 0 & \text{, otherwise} \end{cases} \qquad (17)$$

$\{X_t\}$ is clearly an intermittent process and hopefully it can be modeled according to the approach introduced in the previous sections. This hypothesis will be tested for the Delaware River at Valley Falls, Kansas, USA. Quimpo (4) reported that the data for the period 1923-1960 is of good quality and free of man-made influences. The drainage area is 922 square miles. The study concentrated on June and July because this is the rainy season, and therefore it is expected to be the season of the year when the positive increments play a relevant role on the general picture of the process.

A *simple evaluation criteria of the performance* of the system was selected: a) goodness of fit of the marginal distribution, b) how the theoretical distribution of run lengths resemble the historical ones. The latter will give an indication of how well the model duplicates the time persistence of data.

The results are summarized in Tables 1 and 2. It is apparent that the marginal fit for June was very good. In fact $P(X_{11}^2 > 12.23) \simeq 0.36$. For July the result was not so encouraging: $P(X_7^2 > 14.84) \simeq 0.04$. However, it seems fair to say that even for July the marginal fit will not exclude the use of the model.

A *approximation* to the theoretical distribution of run lengths could be obtained through the use of an expression due to Saldarriaga et al (5). However, for this particular need it was simpler to use the experimental method: a synthetic trace of the $\{M_t\}$ process was generated and the corresponding sample frequencies were considered approximations of the theoretical ones. The trace was long enough to accomodate 10000 runs. A measurement of the goodness of fit for the distribution of run

Table 1

Results for the Delaware River

JUNE

There are 823 zero values in 1140 observations

Parameters: $\mu = -9.685$ $\quad \sigma = 16.360$
$\rho = 0.258$ $\quad \alpha = 0.364$

Marginal fit of $\{X_t\}$ ⟶ $\chi^2_{11} = 12.23$

Class (c.f.s.)	Frequency Observed	Frequency Theoretical	Class (c.f.s.)	Frequency Observed	Frequency Theoretical
0	.723	.722	5000 - 5500	.002	.003
0 - 500	.162	.158	5500 - 6000	.001	.002
500 - 1000	.026	.030	6000 - 6500	.000	.002
1000 - 1500	.014	.018	6500 - 7000	.001	.002
1500 - 2000	.011	.012	7000 - 7500	.001	.002
2000 - 2500	.011	.009	7500 - 8000	.003	.001
2500 - 3000	.008	.007	8000 - 8500	.003	.001
3000 - 3500	.004	.006	8500 - 9000	.002	.001
3500 - 4000	.007	.005	9000 - 9500	.001	.001
4000 - 4500	.006	.004	9500 -10000	.001	.001
4500 - 5000	.004	.003	> 10000	.010	.009

Distribution of Run Lengths

Run Lengths (r)	P(R = r) Observed	P(R = r) Theoretical	No. of observed Runs
1	0.588	0.604	114
2	0.253	0.236	49
3	0.119	0.094	23
4	0.026	0.041	5
5	0.010	0.013	2
6	0.005	0.007	1

Table 2

Results for the Delaware River

JULY

There are 907 zero values in 1140 observations

$\frac{907}{1140} = 0.796$

Parameters: $\mu = -9.261$ $\quad \sigma = 11.220$
$\rho = 0.367$ $\quad \alpha = 0.350$

Marginal fit of X_t ⟶ $\chi^2_7 = 14.84$

Class (c.f.s.)	Frequency Observed	Frequency Theoretical	Class (c.f.s.)	Observed	Theoretical
0	.796	.795	> 5000	.007	.004
0 - 500	.158	.151		.	.
500 - 1000	.020	.020		.	.
1000 - 1500	.007	.010		.	.
1500 - 2000	.003	.006		.	.
2000 - 2500	.003	.004		.	.
2500 - 3000	.001	.003		.	.
3000 - 3500	.003	.002		.	.
3500 - 4000	.002	.002		.	.
4000 - 4500	.000	.002		.	.
4500 - 5000	.000	.001		.	.

DISTRIBUTION OF RUN LENGTHS

Run Lengths (r)	P(R = r) Observed	P(R = r) Theoretical	No. of observed Runs
1	0.658	0.633	104
2	0.241	0.222	38
3	0.089	0.087	14
4	0.006	0.035	1
5	0.006	0.015	1
6	0.000	0.005	0

lengths to be accomplished due to the scarce number of cells. However, a mere inspection seems to indicate that the model does reproduce the time persistence of the data. This assertion is still more acceptable when one realizes that the proportionally large differences between the observed and theoretical frequencies occur for long runs, where there is only a handfull of observations, and of course where the reliability of the observed frequencies is the lowest.

The second example is related to the simulation of single particles movement in sand channels. Hung (1975) conducted extensive experimental as well as theoretical work on this topic: he labeled a sand particle with Cesium 137, which has a half-life of about 30 years, and studied its displacement in a flume 200 ft. long, 8 ft. wide and 4 ft. deep. Hung's first experimental run, which data will be used hereafter, had the following hydraulic characteristics: discharge 12.39 c.f.s.; depth 1 ft.; flume slope 1.667×10^{-3}; Froude numer 0.273. The bed material was coarse sand, with $d_{50} = 1.12$ mm and gradation = 1.51.

From the several variables studied by Hung only the longitudinal displacement will be of interest here. Let the step length X_t be the length of the jump that the particle underwent "instantaneously" on time t. The position of the particle as a function of time is illustrated by Fig. 3. The

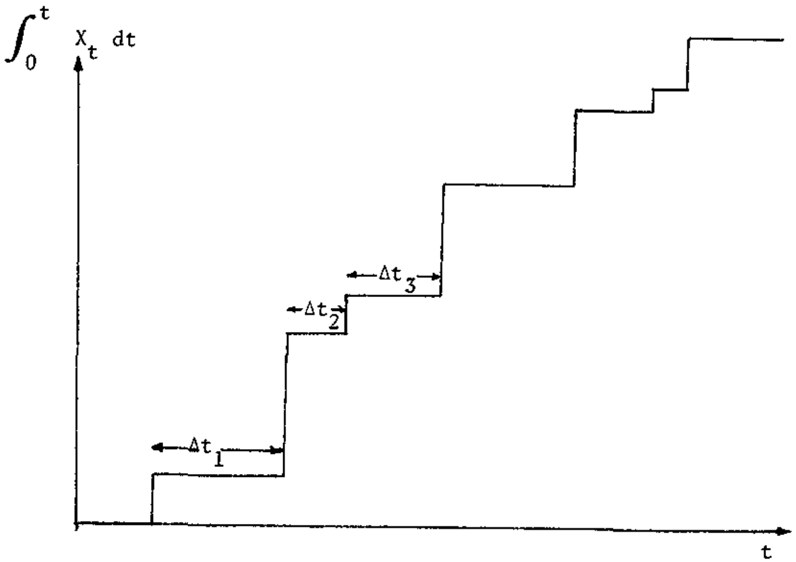

Figure 3.

$\Delta t_1, \Delta t_2, \ldots$, are called resting periods.

For the sake of this example the time axis was divided into two minute cells and if two or more jumps had occurred in the same cell they would be added up yielding a single value. Therefore, the modified time series $\{X'_t\}$ is a succession of positive and zero observations, fitting well, then, the framework described in the previous sections.

The results obtained for the Hung's run 1-2 data are summarized in Table 3. The marginal fit was very good. In fact $P(\chi_9^2 > 7.216) \simeq 0.615$. An interesting fact was the low value estimated for the serial correlation of the hypothetical process $\{C_t\}$: $\hat{\rho} = 0.039$. This seems to suggest the hypothesis that the outcome related to a time cell is independent of the outcome of any other cell, i.e. H_0: $\rho = 0$.

In order to test this hypothesis the asymptotic standard deviation of $\hat{\rho}$ was found to be 0.11 and since the estimator is asymptotically normally distributed the hypothesis cannot be rejected by any standards. Also, a generalized likelihood-ratio test was performed yielding a statistic that under asymptotic conditions is chi-square distributed with one degree of freedom. The observed statistic was 0.126 and since $P(\chi_1^2 > 0.126) \simeq 0.733$ the null hypothesis cannot be rejected again. Of course, the two above results are only valid for "large samples". The question whether the particular sample at hand is large enough to qualify to the asymptotic expression will not be addressed in this paper.

It is interesting to notice that if indeed the time cell events are independent, then the resting periods will have a geometric distribution, which is the discrete analog of the exponential distribution. This confirms an assumption made previously by other investigators (listed by Hung). Also, the distribution of run lengths will be geometric, $P(R = r) = p^{r-1} q$, where for the particular case $\hat{q} = 0.914$. This was the expression used to calculate the theoretical values in Table 3. Taking into consideration the issue of the ill-reliability of the sample freequencies for the longer runs, mentioned in the streamflow example, the agreement between the observed and theoretical frequencies seems fair enough.

CONCLUSIONS

A model for intermittent processes has been presented. Its application to the positive increments of daily runoff is encouraging enough to support the concept that the hydrograph

Table 3

Results for the Single Particle Data

There are 1383 zero values in 1514 observations

Parameters: $\mu = -6.248$ $\sigma = 4.581$
$\rho = 0.039$ $\alpha = 1.923$

Marginal fit of $\{X_t\}$ ⟶ $\chi^2_9 = 7.216$

Class (feet)	Frequency Observed	Frequency Theoretical	Class (feet)	Frequency Observed	Frequency Theoretical
- 0.0	0.913	0.914	1.6 - 1.8	0.008	0.008
0.0 - 0.2	0.001	0.002	1.8 - 2.0	0.005	0.006
0.2 - 0.4	0.004	0.004	2.0 - 2.2	0.003	0.005
0.4 - 0.6	0.008	0.006	2.2 - 2.4	0.005	0.004
0.6 - 0.8	0.010	0.008	2.4 - 2.6	0.002	0.002
0.8 - 1.0	0.007	0.009	2.6 - 2.8	0.001	0.002
1.0 - 1.2	0.013	0.010	2.8 - 3.0	0.001	0.001
1.2 - 1.4	0.007	0.010	3.0 - 3.2	0.001	0.000
1.4 - 1.6	0.011	.009	3.2 - 3.4	0.001	0.000

Distribution of Run Lengths

Run Lengths (r)	P(R = r) Observed	P(R = r) Theoretical	No. of observed Runs
1	0.905	0.914	105
2	0.069	0.079	8
3	0.017	0.007	2
4	0.009	0.001	1

rising and lowering limbs can be represented by different conceptual models. The discharge events classified in the first category (rising limb) have low sensibility to the watershed storage characteristics, and therefore can be accordingly modeled. This is essentially what has been done in the first examples of this paper. Of course, events that fall in the second category (recession) will have to be modeled with the aid of some physical knowledge about the watershed characteristics.

The model was also used for representing the bedload particle movement. For a particular set of data, it was found that the events are serially independent. This suggests that *the power transformed truncated normal distribution*, used to fit the step length data, does not "a priori" present any advantage over some potential competitor. In fact more traditional distributions may even yield a better fit.

ACKNOWLEDGEMENT

This work was done while the author was a graduate student at Colorado State University, under the sponsorship of the "Conselho Nacional de Pesquisas" (National Research Council, Brazil). To his major professor, Dr. V.M. Yevjevich, goes the author's appreciation for the suggestion of the topic.

The discussions on the statistical topics with Dr. D.C. Boes and on the bedload particle example with Dr. C. Hung are gratefully acknowledged.

REFERENCES

(1) COHEN, A.C. Jr. 1959, Simplified Estimators for the Normal Distribution when Samples are Singly Censored or Truncated, Technometrics, Vol. 1, No. 3, pp. 217-237.
(2) YOUNG, C.K., 1968, Discussion of "Mathematical Assessment if Synthetic Hydrology" by N.C. Matalas, Water Resources Research, Vol. 4, No. 3, pp. 681-682.
(3) GREEN, N.M.D., 1973, A Synthetic Model for Daily Streamflow, Journal of Hydrology, 20, pp. 351-364.
(4) QUIMPO, R.G., 1967, Stochastic Model of Daily River Flow Sequences, Hydrology Paper, 18, Colorado State University.
(5) SALDARRIAGA, J., and YEVJEVICH, V., 1970, Application of Run-Lengths to Hydrology Series, Hydrology Paper, 40, Colorado State University.
(6) HUNG, C., 1975, Stochastic Analysis of Bedload Particle Movement, Unpublished Ph.D. dissertation, Colorado State University.

APPENDIX

First and Second Derivatives of the Log-Likelihood Function

Define

$$T(\nu,i,j,k) = \sum^{\ell} \left[\frac{\phi(\xi)}{\Phi(\xi)}\right]^{i} \nu^{j\alpha} (\log \nu)^{k},$$

where ν is a dummy variable that can represent "x_i", "y_i", x_iy_i", and "z_i"

$$\ell = \begin{cases} n_3, & \text{if } \nu \equiv z \\ n_2, & \text{otherwise} \end{cases}$$

$$\xi = \frac{-\rho\nu^{\alpha} - \mu(1-\rho)}{\sigma\,(1-\rho^2)^{1/2}}$$

and

$$I(i,j,k) = \int_{-\infty}^{\infty} \phi(t)\; t^{i}\left[\phi(\varepsilon)\right]^{j}\left[\Phi(\varepsilon)\right]^{k} dt$$

where

$$\varepsilon = \frac{-\mu-(\rho\,\sigma t)^{1/2}}{\sigma\,(1-\rho)^{1/2}}$$

and

$$\beta(i) = I(i,1,1)\left[\frac{2I(0,1,1)}{I(0,0,2)} - \frac{\mu}{\sigma(1-\rho)^{1/2}}\right] - \left(\frac{\rho}{1-\rho}\right)^{1/2} I(i+1,1,1) - I(i,2,0)$$

$i = 0,1.$

$$\frac{\partial LL}{\partial \mu} = \frac{-2n_1\, I(0,1,1)}{\sigma(1-\rho)^{1/2} I(0,0,2)} + \frac{T(x;0,1,0) + T(y;0,1,0) - 2n_2\mu}{\sigma^2(1+\rho)}$$

$$- \frac{(1-\rho)^{1/2}\, T(z;1,0,0)}{\sigma\,(1+\rho)^{1/2}} + \frac{T(z;0,1,0) - n_3\mu}{\sigma^2}$$

$$\frac{\partial LL}{\partial \sigma} = \frac{2n_1\mu\; I(0,1,1)}{\sigma^2(1-\rho)^{1/2} I(0,0,2)} - \frac{2n_2 + n_3}{\sigma}$$

$$+ \frac{T(x;0,2,2) + T(y;0,2,0) - 2\rho T(xy;0,1,0)}{\sigma^3(1-\rho^2)}$$

$$+ \frac{2\mu\left[n_2\mu - T(x;0,1,0) - T(y;0,1,0)\right]}{\sigma^3(1+\rho)}$$

$$+ \frac{\rho T(z;1,1,0) + \mu(1-\rho)T(z;1,0,0)}{\sigma^2\,(1-\rho^2)^{1/2}}$$

$$+ \frac{T(z;0,2,0) - 2\mu T(z;0,1,0) + n_3\mu^2}{\sigma^3}$$

$$\frac{\partial LL}{\partial \rho} = \frac{-n_1\mu\; I(0,1,1)}{\sigma\,((1-\rho)^3)^{1/2} I(0,0,2)} - \frac{n_1\; I(1,1,1)}{(\rho(1-\rho)^3)^{1/2} I(0,0,2)} + \frac{n_2}{(1-\rho^2)}$$

$$+ \frac{\mu\left[n_2\mu - T(x;0,1,0) - T(y;0,1,0)\right]}{\sigma^2(1+\rho)^2} - \frac{T(z;1,1,0)}{\sigma\left\{(1-\rho^2)^3\right\}^{1/2}}$$

$$+ \frac{(1+\rho^2)T(xy;0,1,0) - \rho T(x;0,2,0) + T(y;0,2,0)}{\sigma^2(1-\rho^2)^2} + \frac{\mu T(z;1,0,0)}{\sigma(1-\rho^2)^{1/2}(1+\rho)}$$

$$\frac{\partial LL}{\partial \alpha} = \frac{\mu(1-\rho)\left[T(x;0,1,1)+T(y;0,1,1)\right] - T(x;0,2,1) - T(y;0,2,1)}{(1-\rho^2)\sigma^2} +$$

$$+ \rho T(xy;0,1,1)/(1-\rho^2)\,\sigma^2$$

$$+ \frac{\mu T(z;0,1,1) - T(z;0,2,1)}{\sigma^2} - \frac{\rho T(z;1,1,1)}{\sigma\,(1-\rho^2)^{1/2}} + \frac{2n_2+n_3}{\alpha}$$

$$+ T(xy;0,0,1) + T(z;0,0,1)$$

$$\frac{\partial^2 LL}{\partial\mu^2} = \frac{-2n_1}{\sigma^2(1-\rho)}\frac{\beta(0)}{I(0,0,2)}\ \frac{-2n_2}{\sigma^2(1+\rho)}\ \frac{-n_3}{\sigma^2} + \frac{\rho(1-\rho)^{1/2}}{\sigma^3(1+\rho)^{3/2}}T(z;1,1,0)$$

$$+\left\{\left(\frac{1-\rho}{1+\rho}\right)^3\right\}^{1/2}\frac{\mu}{\sigma^3}T(z;1,0,0) - \frac{(1-\rho)}{\sigma^2(1+\rho)}T(z;2,0,0)$$

$$\frac{\partial^2 LL}{\partial\mu\partial\sigma} = \frac{2n_1}{\sigma^3(1-\rho)}\frac{\beta(0)}{I(0,0,2)} + \frac{2n_1}{\sigma^2(1-\rho)^{1/2}}\frac{I(0,1,1)}{I(0,0,2)}$$

$$+\frac{2\left[2n_2\mu-T(x;0,1,0)-T(y;0,1,0)\right]}{\sigma^3(1+\rho)} + \frac{2\left[n_3\mu-T(z;0,1,0)\right]}{\sigma^3}$$

$$+\frac{(1-\rho)^{1/2}\left[(1+\rho)\sigma^2-(1-\rho)\mu^2\right]T(z;1,0,0)-2\mu\rho\,(1-\rho)^{1/2}\,T(z;1,1,0)}{\sigma^4(1+\rho)^{3/2}}$$

$$+\frac{\mu(1-\rho)T(z;2,0,0)}{\sigma^3(1+\rho)} - \frac{\rho^2T(z;1,2,0)}{\sigma^4\left\{(1+\rho)^3(1-\rho)\right\}^{1/2}} + \frac{\rho T(z;2,1,0)}{\sigma^3(1+\rho)}$$

$$\frac{\partial^2 LL}{\partial\mu\partial\rho} = \frac{-n_1}{\sigma(1-\rho)^{3/2}}\frac{I(0,1,1)}{I(0,0,2)}\left[\frac{\mu\beta(0)}{\sigma} + \frac{\beta(1)}{\sigma^{1/2}}\right] - \frac{n_1}{\sigma(1-\rho)^{3/2}}\frac{I(0,1,1)}{I(0,0,2)}$$

$$+\frac{2n_2\mu-T(x;0,1,0)-T(y;0,1,0)}{\sigma^2(1+\rho)^2} + \frac{\left[\sigma^2(1+\rho)-\mu^2(1+\rho)\right]T(z;1,0,0)}{\sigma^3\left\{(1+\rho)^5(1-\rho)\right\}^{1/2}}$$

$$+\left\{\frac{(1-\rho)}{(1+\rho)^5}\right\}^{1/2}\frac{\mu}{\sigma^3}T(z;1,1,0) + \frac{\rho T(z;1,2,0)}{\sigma^3\left\{(1+\rho)^5(1-\rho)^3\right\}^{1/2}}$$

$$+\frac{\mu T(z;2,0,0)}{\sigma^2(1+\rho)^2} - \frac{T(z;2,1,0)}{\sigma^2(1+\rho)^2(1-\rho)}$$

$$\frac{\partial^2 LL}{\partial\mu\partial\alpha} = \frac{T(x;0,1,1)+T(y;0,1,1)}{(1+\rho)\ {}^2} + \frac{T(z;0,1,1)}{\sigma^2}$$

$$+\frac{\rho\left[\rho T(z;1,2,1)+\mu(1-\rho)T(z;1,1,1)\right]}{\sigma^3\left\{(1+\rho)^3(1-\rho)\right\}^{1/2}} - \frac{T(z;2,1,1)}{\sigma^2(1+\rho)}$$

$$\frac{\partial^2 LL}{\partial\sigma^2} = \frac{-2n_1\mu^2}{\sigma^4(1-\rho)}\frac{\beta(0)}{I(0,0,2)} - \frac{4n_1\mu}{\sigma^3(1-\rho)^{1/2}}\frac{I(0,1,1)}{I(0,0,2)} + \frac{2n_2+n_3}{\sigma^2}$$

$$+\frac{3\left[2\rho T(xy;0,1,0)-T(x;0,2,0)-T(y;0,2,0)\right]}{\sigma^4(1-\rho^2)}$$

$$+\frac{3\left[2\mu T(z;0,1,0)-T(z;0,2,0)-\rho_3\mu^2\right]}{\sigma^4}$$

$$+\frac{6\mu\left[T(x;0,1,1)+T(y;0,1,0)-n_2\mu\right]}{\sigma^4(1+\rho)}$$

$$+\frac{(1-\rho)^{1/2}\left[\mu^3(1-\rho)-2\mu\sigma^2(1+\rho)\ T(z;1,0,0)\right]}{\sigma^5\left\{(1+\rho)^3\right\}} + \frac{\rho^3T(z;1,3,0)}{\sigma^5\left\{(1-\rho^2)^3\right\}}$$

$$+\frac{\rho\left[3\mu^2(1-\rho)-2\sigma^2(1+\rho)\right]T(z;1,1,0)+3\mu\rho^2T(z;1,2,0)}{\sigma^5\left\{(1+\rho)^3(1-\rho)\right\}}$$

$$-\frac{\mu\left[(1-\rho)\mu T(z;2,0,0)+2\rho T(z;2,1,0)\right]}{\sigma^4(1+\rho)} - \frac{\rho^2T(z;2,2,0)}{\sigma^4(1-\rho^2)}$$

$$\frac{\partial^2 LL}{\partial\sigma\partial\rho} = \frac{n_1\mu}{\sigma^2(1-\rho)^2\ I(0,0,2)}\left[\frac{\mu\beta(0)}{\sigma} + \frac{\beta(1)}{\rho^{1/2}}\right] + \frac{n_1\mu}{\sigma^2\left\{(1-\rho)^3\right\}^{1/2}}\frac{I(0,1,1)}{I(0,0,2)}$$

$$+\frac{2\{\rho\left[T(x;0,2,0)+T(y;0,2,0)\right]-(1+\rho^2)T(xy;0,1,0)\}}{\sigma^3(1-\rho^2)^2} +$$

$$+\frac{2\mu\left[T(x;0,1,0)+T(y;0,1,0)-n_2\mu\right]}{\sigma^3(1+\rho)^2}$$

$$+\frac{\mu\left[\mu^2(1-\rho)-\sigma^2(1+\rho)\right]T(z;1,0,0)}{\sigma^4(1+\rho)^2\,(1-\rho^2)^{1/2}}$$

$$+\frac{\left[\sigma^2(1+\rho)-\mu^2(2\rho^2-3\rho+1)\right]T(z;1,1,0)-\mu\rho(2-\rho)T(z;1,2,0)}{\sigma^4\left\{(1+\rho)^5(1-\rho)^3\right\}^{1/2}}$$

$$+\frac{\mu\left[T(z;2,1,0)-\mu T(z;2,0,0)\right]}{\sigma^3(1+\rho)^2}-\frac{\rho^2T(z;1,3,0)}{\sigma^4\left\{(1-\rho^2)^5\right\}^{1/2}}+\frac{\rho T(z;2,2,0)}{\sigma^3\left\{(1-\rho^2)^2\right\}^{1/2}}$$

$$\frac{\partial^2LL}{\partial\sigma\partial\alpha}=\frac{-2\{\mu(1-\rho)\left[T(x;0,1,1)+T(y;0,1,1)\right]-T(x;0,2,1)-T(y;0,2,1)}{(1-\rho^2)\sigma^3}+$$

$$+\rho T(xy;0,1,1)/(1-\rho^2)\ \sigma^3\}$$

$$+\frac{2\left[T(z;0,2,1)-\mu T(z;0,1,1)\right]}{\sigma^3}+\frac{\rho\left[\sigma^2(1+\rho)-\mu^2(1-\rho)\right]T(z;1,1,1)}{\sigma^4\left\{(1+\rho)^3(1-\rho)\right\}^{1/2}}$$

$$+\frac{\rho\left[\rho T(z;2,2,1)+\mu(1-\rho)T(z;2,1,1)\right]}{\sigma^3(1-\rho^2)}$$

$$+\frac{\rho^2\left[\rho T(z;1,3,1)+2\mu(1-\rho)T(z;1,2,1)\right]}{\sigma^4\left\{(1-\rho^2)^3\right\}^{1/2}}$$

$$\frac{\partial^2LL}{\partial\rho^2}=\frac{-n_1\mu}{2\sigma(1-\rho)^3\ I(0,0,2)}\left[\frac{\mu\beta(0)}{\sigma}+\frac{\beta(1)}{\rho^{1/2}}\right]$$

$$-\frac{-n_1}{2\rho^{1/2}(1-\rho)^3\ I(0,0,2)}\left[\frac{\mu\beta(1)}{\sigma}+\frac{1}{\rho^{1/2}}\left[\frac{2\ I(1,1,1)}{I(0,0,2)}\right]^2-\right.$$

$$\left.\left.-\frac{\mu I(2,1,1)}{\sigma\ (1-\rho)^{1/2}}-\frac{\rho}{1-\rho}\ I(3,1,1,1)-I(2,2,0)\right]\right]$$

$$-\frac{-3n_1\mu}{2\sigma\ \left[(1-\rho)^5\right]^{1/2}}\frac{I(0,1,1)}{I(0,0,2)}+\frac{n_1(1-4\rho)}{2\left\{\rho^3(1-\rho)^5\right\}^{1/2}}\frac{I(1,1,1)}{I(0,0,2)}$$

$$+\frac{n_2(1+\rho^2)}{(1-\rho^2)^2}+\frac{2\mu\left[T(x;0,1,0)+T(y;0,1,0)-n_2\mu\right]}{\sigma^2(1+\rho)^3}$$

$$+\frac{2\rho(3+\rho^2)T(xy;0,1,0)-(1+3\rho^2)\left[T(x;0,2,0)+T(y;0,2,0)\right]}{\sigma^2(1-\rho^2)^3}$$

$$+\frac{\mu\left[\sigma^2(2\rho-1)(1+\rho)^2+\mu^2(1-\rho)\right]\ T(z;1,0,0)}{\sigma^3\left\{(1+\rho)^7(1-\rho)^3\right\}^{1/2}}$$

$$+\frac{\mu(1-2\rho)T(z;1,2,0)-\left[3\sigma^2\rho(1+\rho)+\mu^2(1-\rho)(2-\rho)\right]T(z;1,1,0)}{\sigma^3\left\{(1+\rho)^7(1-\rho)^5\right\}^{1/2}}$$

$$-\frac{\mu^2T(z;2,0,0)}{\sigma^2(1+\rho)^3(1-\rho)}+\frac{2\mu T(z;2,1,0)}{\sigma^2(1+\rho)^3(1-\rho)^2}+\frac{\rho T(z;1,3,0)}{\sigma^3\left\{(1-\rho^2)^7\right\}^{1/2}}\frac{T(z;2,2,0)}{\sigma^2\left\{(1-\rho^2)^3\right\}^{1/2}}$$

$$\frac{\partial^2LL}{\partial\rho\partial\alpha}=\frac{-\mu\left[T(x;0,1,1)+T(y;0,1,1)\right]}{\sigma^2(1+\rho)^2}+\frac{(1+\rho^2)T(xy;0,1,1)}{\sigma^2(1-\rho^2)^2}$$

$$-\frac{2\rho\left[T(x;0,2,1)+T(y;0,2,1)\right]}{\sigma^2(1-\rho^2)^2}$$

$$-\frac{\left[\sigma^2(1+\rho)+\mu^2\rho(1-\rho)\right]T(z;1,1,1)}{\sigma^3\left\{(1+\rho)^5(1-\rho)^3\right\}^{1/2}}+\frac{\rho^2T(z;1,3,1)}{\sigma^3\left\{(1-\rho^2)^5\right\}^{1/2}}$$

$$+\frac{\rho\left[\mu(1-\rho)T(z;2,1,1)-T(z;2,2,1)\right]}{\sigma^2(1-\rho^2)^2}+\frac{\rho\mu T(z;1,2,1)}{\sigma^3\left\{(1+\rho)^5(1-\rho)\right\}^{1/2}}$$

$$\frac{\partial^2 LL}{\partial \alpha^2} = \frac{\mu\left[T(x;0,1,2)+T(y;0,1,2)\right]}{\sigma^2(1+\rho)}$$

$$+ \frac{\rho T(xy;0,1,2)-2\left[T(x;0,2,2)+T(y;0,2,2)\right]}{\sigma^2(1-\rho^2)}$$

$$+ \frac{\mu T(z;0,1,2)-2T(z;0,2,2)}{\sigma^2} - \frac{\rho^2 T(z;2,2,2)}{\sigma^2(1-\rho^2)}$$

$$+ \frac{\rho^2\left[\rho T(z;1,3,2)+\mu(1-\rho)T(z;1,2,2)\right]}{\sigma^3\left\{(1-\rho^2)^3\right\}^{1/2}} - \frac{2n_2+n_3}{\alpha^2}$$

DISCUSSION

E. Hansen (General Reporter): The author introduces a different and interesting approach for synthesizing intermittent data. The basic assumption is that the intermittent process may be considered to result from censoring a non-intermittent normal autoregressive lag-one model. No physical justification of the underlying non-intermittent process is attempted. In fact, the main objective for introducing the imaginary process is a hope that this is easier to model than the intermittent process. The marginal distribution of the intermittent signal is modelled with a power-transformed truncated normal distribution containing 3 parameters and therefore some flexibility for fitting purposes.

For obtaining estimates of the model parameters on the basis of a given time series of some intermittent process the maximum likelihood procedure is applied. The likelihood function for n observations evaluated in a straightforward way would include an n-fold integral of the multivariate normal p.d.f. This integral is not analytically available for $n > 3$. In order to avoid this difficulty, the author assumes that the pairs of observations (x_1, x_2), (x_3, x_4), (x_5, x_6), ... are independent, and the corresponding likelihood is to be maximized. The author claims that his experience with generated time series, i.e. in situations when the population values are known, supports the accuracy of the results yielded by the above simplifying approximation. It might, however, be questioned whether the relevance of the assumption does not to some extent depend on the structure of the involved processes? The assumption might f.i. be expected to be increasingly violated the larger the mean run-length of the involved intermittent process, or what amounts to the same the larger the time persistence. It is felt that this point need a little more clarification, the more so that the processes treated in the examples are all characterized by small mean run-lengths in the range of 1 to 2 time steps.

The proposed method for modelling intermittent processes has the advantage that it admits immediate known extensions to multivariate situations. Theoretical results pertaining to this situation are given in the paper. As was the case in the single-station approach, some simplifying assumptions are necessary with respect to the estimation procedure for the multivariate case. Independence of the pairs of observations (x_{1j}, x_{1k}), (x_{2j}, x_{2k}), ... pertaining to station j and k are assumed. The relevance of this assumption need to be investigated further with respect to the influence of the time persistence properties of the involved processes. In this connection, it is pointed out by the author that for the proposed multivariate approach, the higher than lag-zero cross-correlations of the historical processes will not be reproduced by the synthetic traces.

In the paper, only stationary processes are considered. However, many processes are non-stationary e.g. due to seasonal variations. It would be of interest to know if it is possible to extend the suggested approach to the type of processes very often encountered in nature, viz. processes composed of periodic and stochastic components.

J. Kelman: Thank you professor Hansen for the review as well as for the interesting comments.

I would like now to answer, the questions raised by the General Reporter.

The first one deals with the estimation procedure. Professor Hansen points out that the assumption of *independence among* successive pairs of values might lead to wrong estimates in case of strong dependence in the time series. I agree with that. However, the question of how important this problem might be depends also on the sample size. I have been applying this model to rainfall data, which is in general characterized by long samples, quite successfully. Also in the case of positive increments of daily streamflow the results were encouraging. An example of this last case is given in the paper. I do not dispute the fact that in case of small samples with high time persistence the model might not apply.

The second question deals with the fact that the stochastic processes one wants to model are in general non-stationary. The General Reporter asks how the model can be generalized *to this* case. For the study of daily rainfall as well as streamflow I have used the same structure of the model, assuming, however, that the parameters are periodic functions of time, rather than constants. The results were satisfactory.

L.V. Tavares: Most methods developed to generate daily streamflows do not work well and I believe one of the reasons for this is using the same process to simulate the rising limb and the decreasing one of the generated hydrographs. In your paper you present an application of your model to simulate the hydrographs rising limbs. Perhaps if you build up a second model to generate the hydrographs decreasing limbs you will achieve a good daily streamflow simulation model. Are you planning to do this?

J. Kelman: I am glad professor Tavares raised this question because it gives me the opportunity to discuss a point that I consider of relevance in the modeling of streamflow sequences with short time intervals. As mentioned in the text, the reason for studying the positive increments of daily streamflow stems from the conception that this intermittent stochastic process resembles closely the physical process of direct-run-

off. Professor Tavares raises the question of how the negative increments should be modeled. My reasoning is that the negative increments, or equivalently the lowering limbs of the hydrograph must translate the physical process of the emptying of the watershed storage. Therefore a way of hypothesizing a model of the negative increments is by stating which are the characteristics of the watershed storage. Elsewhere I have assumed the conceptual representation of the watershed as two linear reservoirs. The joint use of the two intermittent and alternating processes as a dual model for daily streamflow yielded satisfactory results.

Reference:
Stochastic Modeling of Intermittent Daily Hydrologic Series, Jerson Kelman, Ph.D. dissertation, Colorado State University, 1976.

K.W. Hipel: When determing a stochastic model for a particular data set, it is recommended to follow the identification, estimation and diagnostic check stages of model development. (Box and Jenkins, 1970; Box and Tiao, 1973). By fixing the one component to be Markovian (i.e. autoregressive lag one process), the author precludes the necessity of having to identify a particular autoreqressive-moving average (ARMA) model from the general family of ARMA models. Can the author's method be generalized so that the best ARMA model is employed in place of the Markov component?

The ξ_t white noise term is assumed to be independent, normally distributed, and *homoscedastic. Applied statisticians stress* the importance of satisfying the independence assumption. In the applications presented in the paper, were diagnostic checks performed to determine whether or not the independence assumption was satisfied? Although the normality and homoscedastic properties are generally not as important as the independence criterion, did the author also check whether or not these assumptions for the white noise term were reasonably well satisfied?

References:
Box, G.E.P, Jenkins, G.M., Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, California, USA, 1970.
Box, G.E.P, and Tiao, G.C., Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Mass., USA, 1973.

J. Kelman: Thank you Dr. Hipel for the two stimulating questions. The first one deals with the problem of model identification. This stage was skipped in the present study due to the lack of competitor models for the intermittent process. The main purpose of the paper is to introduce a new model, rather than selecting the "best" model for a particular set of data. However, as Dr. Hipel points out, suitable competitors could be found by assuming different characteristics for the M_t-process, like for example the referred ARMA model. In fact there is no impediment in the intermittent model for the use of any linear model for the M_t-process. Therefore, the method can be generalized in the direction suggested by Dr. Hipel.

The second question is related to the properties of the time series $\{\xi_t\}$. Here the same symbol is used either representing the stochastic process or the corresponding time series. I must clarify the point that it is not possible to get the time series $\{M_t\}$ as well as $\{\xi_t\}$ given the raw sequence $\{X_t\}$. The fact is that whenever an outcome of X_t is equal to the truncation value, one only knows that the corresponding value of M_t is smaller than the truncation value. How much smaller one does not know. In other words, figure 1 ilustrates the generation procedure, but the arrows can not be inverted to find final product the white noise, given as input the observed intermittent process. Therefore the procedure suggested by Dr. Hipel for the testing of the fitness of the model can not be applied. Instead, I have compared the sample distributions of several functionals obtained from the historic series with their counterparts obtained from generated sequences. The latter were produced by the proposed model. Although somehow deficient, this procedure gives a measure of the adequacy of the model to produce results of practical significance.