A MULTIVARIATE SYNTHETIC DAILY STREAMFLOW GENERATOR

By

J. Kelman, Research Engineer
CEPEL - Electric Energy Research Center
J. M. Damazio, Research Engineer
CEPEL - Electric Energy Research Center
and
J. P. da Costa, Research Engineer
CEPEL - Electric Energy Research Center
C.P. 2754, Rio de Janeiro, Brazil

ABSTRACT

There are several stochastic models available in the literature developed to represent the one-site daily streamflow process, (Yevjevich, 1984). The one proposed by Kelman (1977, 1980) is based on the assumption that the rising and the falling limbs of daily hydrographs ought to be modelled separately because they translate different physical processes. It is presented an extension of this approach, the DIANA model (Kelman et al.,1983) which was developed for flood control and dam safety studies in Brazil. The DIANA model differs from the previous version in several aspects, the main one is that it allows a multisite generation of synthetic sequences. The runnof Q(t) on day t is considered to be the sum of two components Q(t) = U(t) + O(t), where U(t) is a seasonal, auto-correlated, non-negative intermittent process that represents the rising of hydrographs due to precipitation events. Since precipitation is not explicity considered in the model, no precipitation data is used either to calibrate the model or to generate synthetic series. U(t) is assumed to result from the censoring of a transformed normal AR(1) process, Z(t). O(t) represents the emptying of the watershed which is assumed to be O(t) = K Q(t-1) where K is a random variable upper bounded by a model parameter, which is smaller than 1. The probability distribution of K depends only on U(t). The multivariate extension is accomplished by modelling the vector Z(t) (each component corresponds to a site) by the approach suggested by Matalas (1967). The model is tested for eight cascaded gauging stations of South-Southeast Brazil.

INTRODUCTION

There are several problems in the engineering practice that demand the discretization of streamflow in a time interval as short as one day. Typical examples are the calculation of reservoir space requirements for flood control (Beard, 1968), reliability calculation for pumped storage reservoirs

(O'Connell and Jones, 1978) and dam-safety analysis (Kelman and Damazio, 1983).

The development of operational stochastic models for daily streamflow series has encountered some difficulties. In general, the adaptation of traditional time series models, as ARIMA models, in modelling of daily streamflow fails due to data large skewness, the strong seasonal effects and different behaviour of the rising and falling limbs of the hydrographs (Kelman, 1976 and O'Connell and Jones, 1978).

Kelman (1976, 1980) obtained satisfactory results by separately modelling the rising and falling limbs of the hydrographs. The rising limb was modelled similarly to precipitation as an intermittent process and the falling limbs was conceived as the result of the emptying ouflow from linear reservoirs.

In this paper it is presented an extension of this approach, the DIANA model, which was developed for flood control and dam-safety studies in Brazil. The DIANA model differs from the previous version in several aspects, the main one being the allowance to multisite generation of synthetic sequences.

MODEL'S DESCRIPTION

The runoff $Q(t)$ on day t is considered to be the sum of two components:

$$Q(t) = U(t) + O(t) \; ; \; t=1,2,\ldots \tag{1}$$

Conceptually, $U(t)$ represents external factors (e.g.: precipitation) with intermittent characteristics that affect the rising of the hydrograph. On the other hand, $O(t)$ represents the persistent emptying outflow from the watershed. As only the total flow $Q(t)$ time series is available, some rule is necessary to split $Q(t)$ in its components.

Whenever $q(t) > \lambda q(t-1)$ let us assume that $u(t) > 0$, $\lambda \in (0,1)$ being a watershed characteristic recession constant. For any λ, the following relations hold:

$$u(t) = 0 \quad \text{if} \quad q(t) \leq \lambda q(t-1) \tag{2a}$$

$$u(t) = q(t) - \lambda q(t-1) \quad \text{if} \quad q(t) > \lambda q(t-1) \tag{2b}$$

In this way, whenever $u(t) > 0$ the total flow is given by

$$q(t) = u(t) + \lambda q(t-1) \quad ; \quad u(t) > 0 \tag{3}$$

and so, by the use of (1)

$$o(t) = \lambda q(t-1) \quad ; \quad u(t) > 0 \tag{4}$$

On the other hand, whenever external factors are not active ($u(t)=0$) the total flow is only given by $O(t)$. In DIANA model $O(t)$ is considered to be equal to the outflow of a linear reservoir with stochastic behaviour. This behaviour is modelled setting $O(t)$ as a random fraction $K \leq \lambda$ of the previous flow.

$$q(t) = o(t) = k\, q(t-1),\ k \leq \lambda,\ u(t) = 0 \qquad (5)$$

The approach can then be described putting together equations (1), (3) and (5):

$$q(t) = u(t) + k\, q(t-1) \qquad (6)$$

$$u(t) = 0 \rightarrow k \leq \lambda \qquad (6a)$$

$$u(t) > 0 \rightarrow k = \lambda \qquad (6b)$$

There are two aspects to consider in the modelling of the $U(t)$ process:

i) Its marginal probability distribution (it has a probability mass p at $u(t)=0$).

ii) The external factors that govern the rising limbs of the hydrograph may result from the action of persistent meteorological processes (e.g. cold fronts). The modelling of the $U(t)$ process must be conceived as reproducing this induced time persistence.

The previous version of the model solved these two questions using a power transformation of $U(t)$, $U(t) = (Y(t))^{\alpha}$. The $Y(t)$ process was considered as the result of censoring at the origin an AR(1) process with $N(\mu,\sigma)$ marginal distribution. This parametric representation of $U(t)$ allowed the estimation of all parameters (including α and ρ, the lag one autocorrelation of the non-censored AR(1) process) by a pseudo maximum likelihood procedure or by the method of moments.

This parametric approach is advantageous when dealing with flood studies in flashy rivers, where it is necessary an extrapolation of the daily increments of the flow. In large basins, floods result from the coincidence of several events which are not necessarily exceptional. Indeed, a major flood in a large basin is usually shaped by the persistence of minor daily increments of flow. These increments can already be present in the sample. In this case using the empirical frequency distribution of $U(t)$ one has no need to make any assumption regarding the shape of the distribution. This is the adopted alternative in the DIANA model. Persistence in the DIANA model is also modelled through an AR(1) process with censoring. This process is mapped into $U(t)$ using a non-parametric relationship that preserves the empirical distribution $F_U(.)$.

Let $Z(t)$ be the AR(1) process defined by:

$$z(t) = \rho\ z(t-1) + \sqrt{1-\rho^2}\ \varepsilon(t) \qquad (7)$$

where $\varepsilon(t)$ is a normal distributed white-noise and ρ is the lag one autocorrelation of $Z(t)$.

$Y(t)$ results from imposing a censoring in $Z(t)$ and is defined as:

$$y(t) = z(t) \quad \text{if} \quad z(t) > \beta \qquad (8a)$$

$$y(t) = \beta \quad \text{if} \quad z(t) \leq \beta \qquad (8b)$$

$$\beta = \Phi^{-1}(p) \qquad (8c)$$

where β defines the censoring interval $(-\infty, \beta)$, $\Phi(.)$ is the c.d.f. for the standard normal distribution, and $p = P[u(t)=0]$.

The relationship between $U(t)$ and $Y(t)$ is obtained solving the equation $F_U(u(t)) = \Phi(y(t))$.

When there are $\ell \geq 2$ daily streamflow series to be modelled the spatial dependence is introduced in the model just assuming a multivariate AR(1) process:

$$Z(t) = A\ Z(t-1) + B\ \varepsilon(t) \qquad (9)$$

where $Z(t)$ is the vector $[Z_1(t), Z_2(t), \ldots, Z_\ell(t)]^T$, $Z_i(t)$ is the AR(1) process that corresponds to the i-th hydrograph, and $\varepsilon(t)$ is a vector of ℓ standard normal independent deviates.

The A and B matrices should be chosen in order to resemble the persistence in each hydrograph and the lag-zero covariance between them represented in the covariance matrix M_0 of $Z(t)$.

The generation procedure is performed in the following steps:

1. Sample the vector $Z(0) = [z_1(0), \ldots z_\ell(0)]^T$ from the standard multivariate normal distribution with M_0 as the covariance matrix.
2. Sample initial total flows $q_i(0)$; $i=1,\ldots\ell$, from an empirical multivariate distribution $F_{Q_0}(.)$.
3. Set $t = 1$
4. Sample the vector $\varepsilon(t) = [\varepsilon_1(t), \ldots, \varepsilon_\ell(t)]^T$ from the standard multivariate independent normal distribution.
5. Calculate $Z(t) = A\ Z(t-1) + B\ \varepsilon(t)$
6. For each series calculate $y_i(t) = \max[\beta_i\ , z_i(t)]$ and obtain $u_i(t)$ solving $F_{U_i}(u_i(t)) = \Phi(y_i(t))$
7. If $u_i(t) > 0$ then $q_i(t) = u_i(t) + \lambda_i q_i(t-1)$, if $u_i(t)=0$ then sample $k_i \leq \lambda_i$ from an empirical distribution $F_{k_i}(.)$ calculate $q_i(t) = k_i q_i(t-1)$.
8. Set $t = t+1$ and return to step 4.

ESTIMATION PROCEDURES

Given a multivariate daily streamflow series $\{q_i(t), t = 0,h;\ i=1, \ell\}$ it is necessary to estimate:

a) The characteristic recessions constants (λ_i)

b) The frequency distribution functions $F_{U_i}(.)$ and $F_{k_i}(.)$;

c) The censoring interval limits β_i for the $Y_i(t)$ processes and

d) The matrices A and B.

Assume that the λ_i's are known. Then equations (2a) and (2b) can be used for each series to calculate the corresponding $u_i(t)$, $t=1,2,\ldots,h$ series. Therefore $F_{U_i}(.)$ and $F_{k_i}(.)$ can be obtained in a straightforward way. Moroever, let m_i be the number of zeros in the $u_i(t)$ series. The estimate of β_i is given by:

$$\Phi(\hat{\beta}_i) = \frac{m_i}{h} \tag{10}$$

The estimation of the matrices A and B is done considering the matrices $M_0 = E[Z_t Z_t^T]$ and $M_1 = E[Z_{t+1} Z_t^T]$ given by:

$$m_0(i,j) = 1;\ i=j \tag{11a}$$

$$m_0(i,j) = \rho_{i,j}(0),\ i\neq j \tag{11b}$$

$$m_1(i,j) = \rho_{i,j}(1),\ i=j \tag{11c}$$

$$m_1(i,j) = \rho_{i,j}(0)\ \rho_{i,i}(1),\ i\neq j \tag{11d}$$

where $\rho_{i,i}(1)$ is the lag-one autocorrelation coefficient of $z_i(t)$ and $\rho_{i,j}(0)$ the lag zero cross-correlation coefficient between $z_i(t)$ and $z_j(t)$. The parameters $\rho_{i,j}(0)$ and $\rho_{i,i}(1)$ will be estimated from a realization of the censored process $Y(t)$.

Let $t_1, t_2, \ldots t_h$ be a set of indices such that $u_i(t_1) \leq u_i(t_2) \ldots \leq u_i(t_h)$. The function that relates $y_i(t)$ and $u_i(t)$ is given by $u_i(t) = g_i|\Phi(y_i(t))|$. In this function, for each $u_i(t_j) > 0$ there is an associated interval $|d_j, e_j|$ which is a function of the index j, given by:

$$\Phi(d_j) = \frac{j-1}{h},\ j \geq m+1 \tag{12a}$$

$$\Phi(e_j) = \frac{j}{h},\ j \geq m+1 \tag{12b}$$

For $u_i(tj) = 0$, there is only one associated value:

$$y_i(t_j) = \beta_i,\ j \leq m \tag{13}$$

The function g is useful in the step 6 of the generation procedure, as it gives $u_i(t) = g_i[\Phi(y_i(t)]$. On the other hand, at the estimation phase it is not possible to obtain exactly the value of $y_i(t_j)$, when $u_i(t_j) > 0$. The adopted approach consists in associating to each $u_i(t_j) > 0$ the median point of the associated interval:

$$\Phi(y_i(t_j)) = \frac{1}{2}[\Phi(d_j) + \Phi(e_j)] \; ; \; j \geq m + 1 \qquad (14)$$

With equations (12), (13) and (14) it is possible to obtain for each site i a realization of the process $Y_i(t)$.

If the null values of $u_i(t)$ are neglected, $\rho_{i,i}(1)$ and $\rho_{i,j}(0)$ can be estimated from the contiguous pairs $(y_i(t) > \beta_i; y_i(t+1) > \beta_i)$ for the estimation of $\rho_{i,i}(1)$ and $(y_i(t) > \beta_i; y_j(t) > \beta_j)$ for the estimation of $\rho_{i,j}(0)$. These pairs are considered as random samples taken from truncated (identically and non-identically) bivariate standard normal distributions. In the appendix it is described an estimation procedure which can be used in both cases. Given a set of estimated $\hat{\rho}_{i,i}(1)$ and $\hat{\rho}_{i,j}(0)$ equations (11) are used to construct M_0 and M_1. A and B are then given by:

$$A = M_1 M_0^{-1} \qquad (15)$$

$$BB^T = M_0 - M_1 M_0^{-1} M_1^T \qquad (16)$$

For the solution of (16) it is adopted the principal components solution (Matalas, 1967). The solution of (15) and (16) imposes some restrictions on $\rho_{i,i}(1)$ and $\rho_{i,j}(0)$ in order to guarantee that M_0 and $M_0 - M_1 M_0^{-1} M_1^T$ are positive definite. These conditions are assured if the matrix V given by

$$V = E\left[\begin{bmatrix} Z_{t+1} \\ Z_t \end{bmatrix} \left[Z_{t+1}^T \quad Z_t^T \right] \right] \qquad (17)$$

is positive definite. It may happen, due to sample variation, that the set of estimated correlations $\rho_{i,i}(1)$ and $\rho_{i,j}(0)$ do not present a consistent pattern of correlation. In this case matrix V can be slightly modified (Fiering, 1968) rendering it positive definite.

No formal procedures are available for the estimation of the characteristic recession constants and they have been selected based on trial-and-error. Convergence is accepted when some statistics taken respectively from the historical and the synthetic series are sufficiently "close".

It is important to note that in the case of the non-parametric modelling of the $u_i(t)$ marginal distributions, the chosen values for λ_i define the maximum possible generated total flow, $Q_{max,i}$. It can be demonstrated, using (2), that:

$$Q_{max,i} = \frac{u_{max,i}(\lambda_i)}{1 - \lambda_i} \qquad (18)$$

where $u_{max,i}(\lambda_i)$ is the greatest value in the sample of $u_i(t)$'s.

Example

The allocation of reservoir space for flood control in a cascaded system of reservoirs can be done with the help of synthetic multivariate daily local inflow series. The study of

the reservoir system of the Paraná River (southeast of Brazil) was done by the multivariate generation of 1000 "flood seasons" (October 1st, to April,30th) of local daily inflows to eight reservoirs (see figure 1). In order to consider the seasonal effects, the model's parameters were estimated using seasons of 14 days, with the exception of the characteristic recessions which were assumed constant along the seasons. The historical local flows were obtained by taking the differences between total flows, using representative time-lags. Results for only three reservoirs will be shown.

The confidence on the results produced by any model is based on its capacity to produce synthetic series that are statistically indistinguishable from the correspondent historical series. This validation should be done by comparing relevant properties of the generated series. The set of relevant properties should be chosen in view of the application one intends to do with the generated series.

As in this example one is concerned with flood control the selected properties are:

a) daily flows first and second moments seasonal variation;
b) annual maximum flows frequency distribution;
c) mean, variation, skewness and kurtoses of daily and annual maximum flows;
d) flood volume required to regulate the mean historical annual maximum flow;
e) daily flow correlograms.

Figures 2, 3 and 4 show the seasonal evolution of the mean and of the standard deviation for the daily local flows. Figures 5, 6 and 7 show for each of the reservoirs three hydrographs:

1- The historical hydrograph with the maximum daily flow.
2- A generated hydrograph with the same maximum daily flow as in 1.
3- The generated hydrograph with the maximum generated daily flow.

Figures 8, 9 and 10 show the correlograms of historical and generated series. The results are generally good with the exception of the São Simão series. Table 1 compares the historical and generated lag-zero cross correlations between local flows in different sites. It can be seen that the model is still underestimating the spatial correlation matrix, although the results obtained are reasonable approximations. Figures 11,12 and 13 show the empirical generated and historical annual maximum daily flow frequency distribution. The goodness of fit may be confirmed through the chi-squared statistic obtained with 5 discretization intervals (1.3,3.1 and 3.6). The hypothesis that the historical series of maximum flows are random samples of the probability distribution function defined by the model can not be rejected. Tables 2,

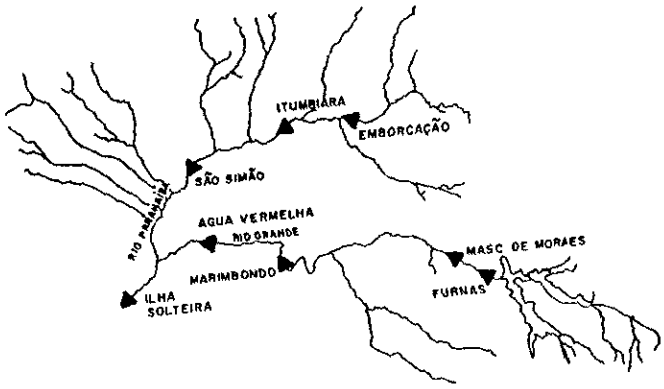

Figure 1 Reservoir System

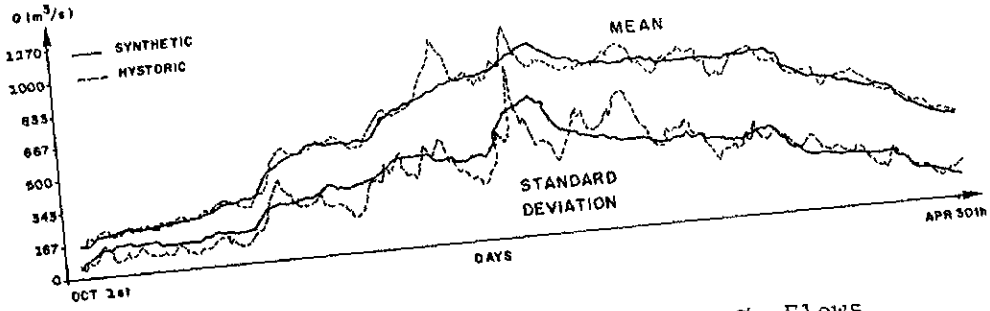

Figure 2 Seasonal Effects for Emborcação Flows

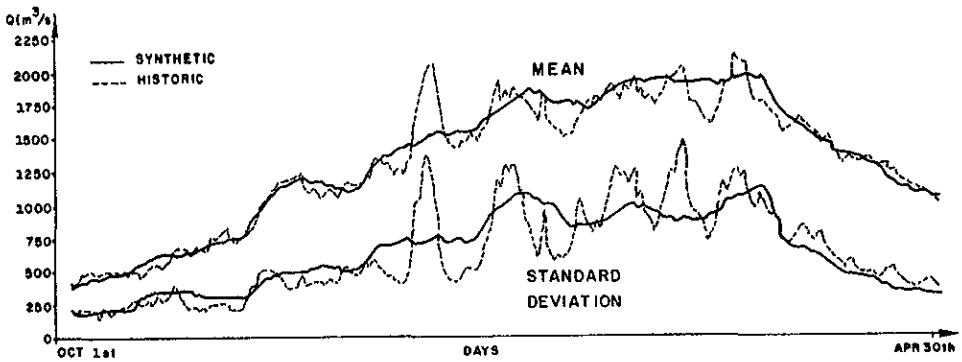

Figure 3 Seasonal Effects for Itumbiara Local Flows

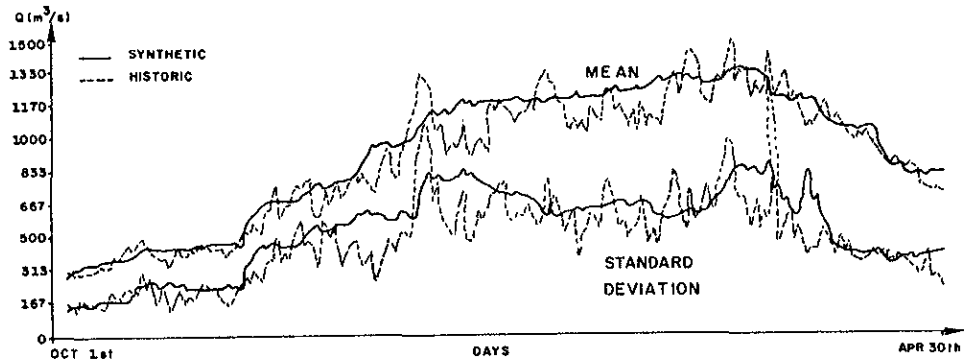

Figure 4 Seasonal Effects for São Simão Local Flows

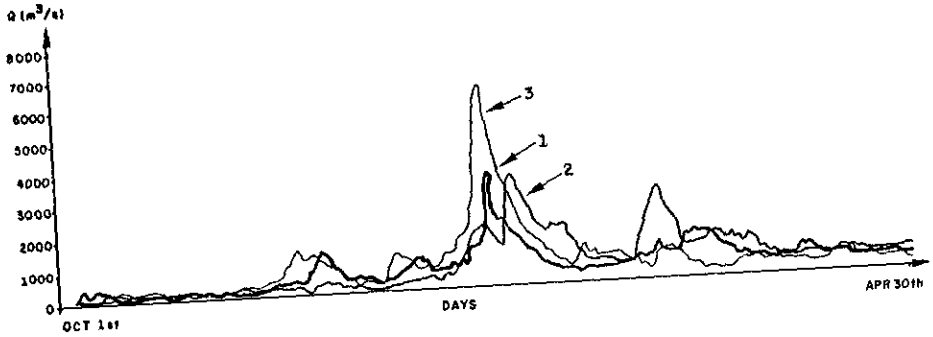

Figure 5 Hydrographs for Emborçacão

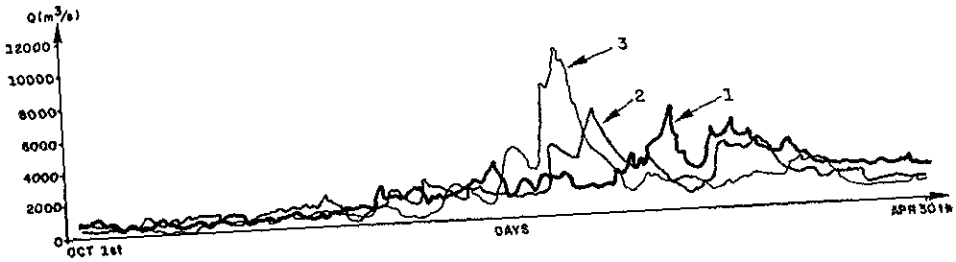

Figure 6 Hydrographs for Itumbiara

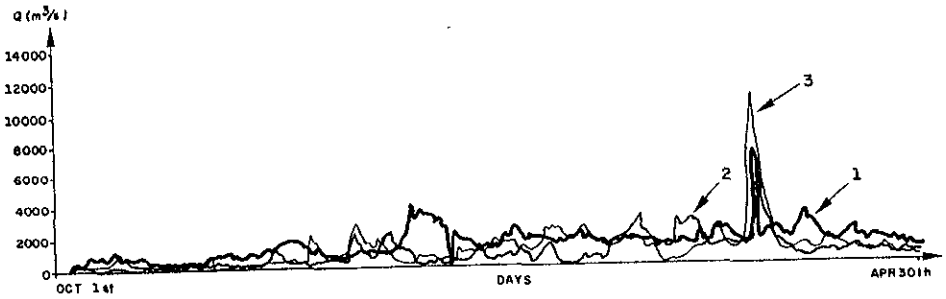

Figure 7 Hydrographs for São Simão

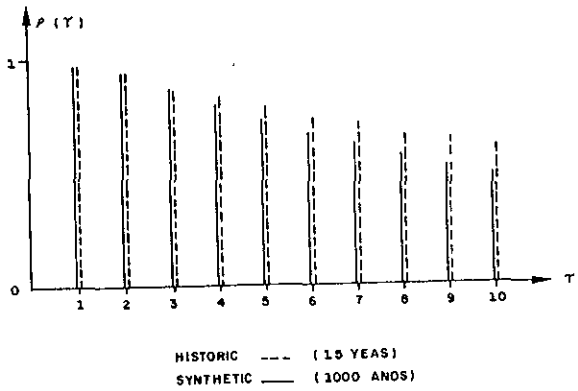

Figure 8 Correlogram for Emborçacão

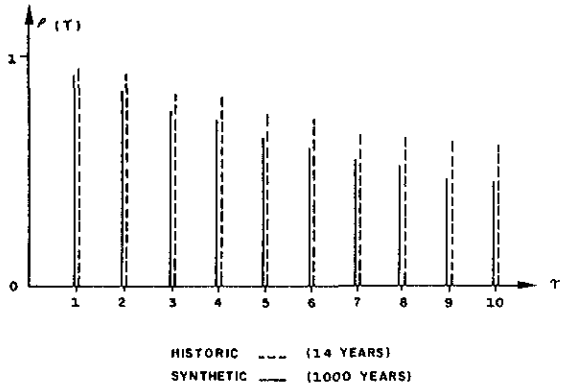

Figure 9 Correlogram for Itumbiara

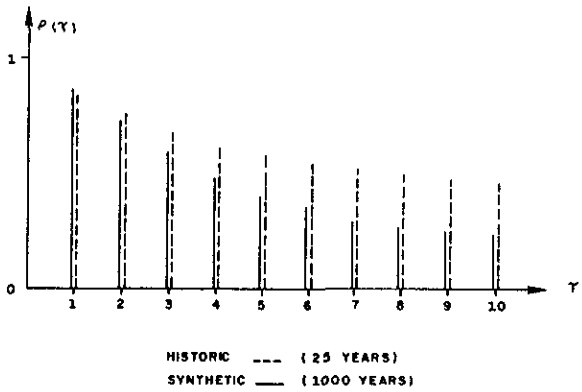

Figure 10 Correlogram for São Simão

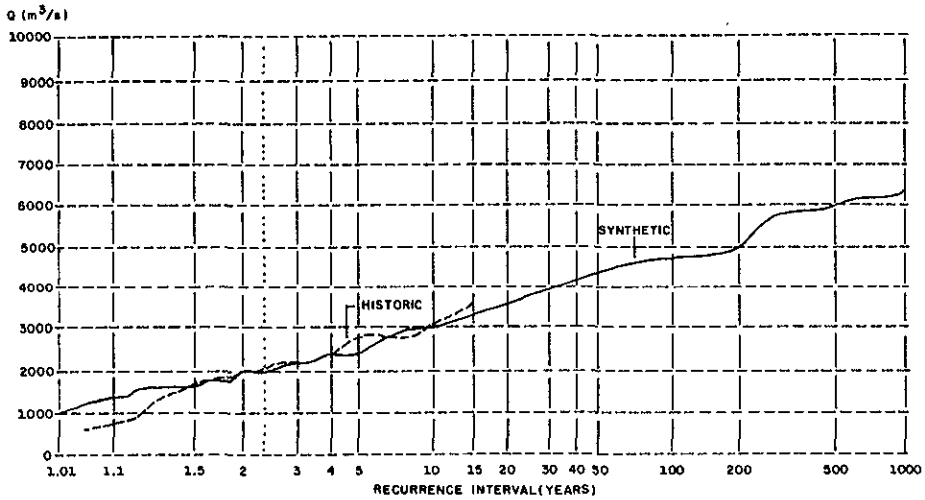

Figure 11 Flood Frequency Curve for [redacted]rçacão

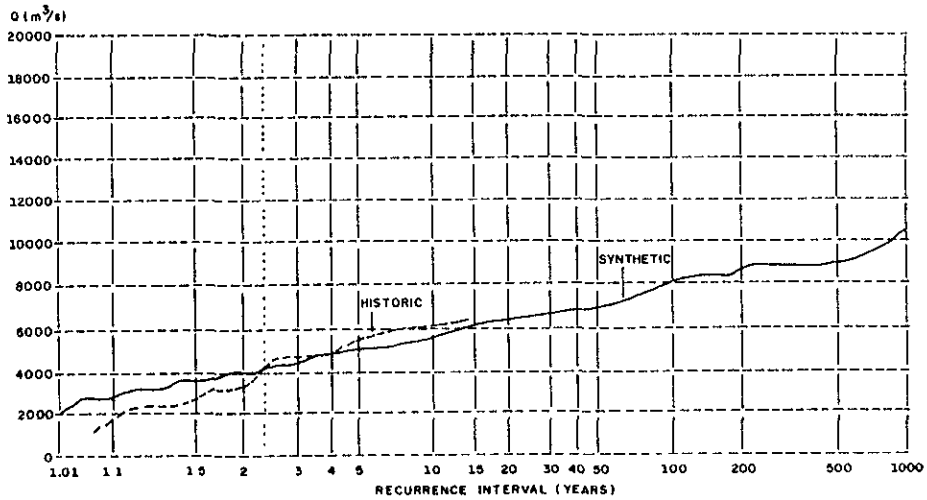

Figure 12 Flood Frequency Curve for Itumbiara

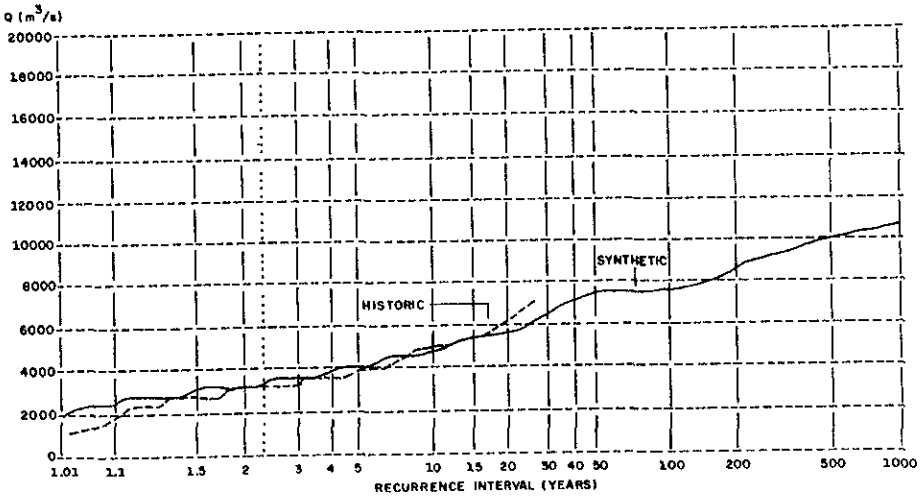

Figure 13 Flood Frequency Curve for São Simão

3 and 4 compare some statistics obtained from historical and synthetic sequences. It can be seen that only in the second reservoir the seventh statistic shows significant disagreement between the generated and the historical series.

CONCLUSIONS

Stochastic daily streamflow models have seldom been reported as useful in flood studies. Few exceptions could be mentioned, as for example Bulu (1979) and Kelman and Damazio (1983). Perhaps the lack of popularity of daily streamflow models is due to the skepticism about the capability of these models to produce synthetic sequences with the same statistical properties as the single observed time series. The case study of the present paper is an example that the skepticism may not be fair. In fact the writers have been successfully performing several flood studies in Brazil (for example, Kelman et al. 1985, this Symposium) with the help of daily synthetic sequences. However it must be noted that this experience is still restricted to large basins, where exceptional floods can be seen as the joint occurence of events that individually are not exceptional.

APPENDIX

Consider a standard bivariate normal distribution with correlation coefficient ρ, truncated at $x = h$ and $y = k$. Then the following equation holds (Rosenbaum, 1961):

$$(h+k)\rho^2-\{(h+k)m_{11}-h\,k(m_{10}+m_{01})\}\rho-(h+k)-hk(m_{10}+m_{01})+km_{20}+hm_{02}=0 \qquad \text{(A1)}$$

where

$$m_{io}=\int_h^\infty\int_k^\infty \frac{x^i \exp[-\frac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)]}{2\pi\sqrt{1-\rho^2}\;L(h,k,\rho)}\,dxdy$$

$$m_{oi}=\int_h^\infty\int_k^\infty \frac{y^i \exp[-\frac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)]}{2\pi\sqrt{1-\rho^2}\;L(h,k,\rho)}\,dxdy$$

$$m_{11}=\int_h^\infty\int_k^\infty \frac{xy \exp[-\frac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)]}{2\pi\sqrt{1-\rho^2}\;L(h,k,\rho)}\,dxdy$$

$$L(h,k,\rho)=\int_h^\infty\int_k^\infty \frac{\exp[-\frac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)]}{2\pi\sqrt{1-\rho^2}}\,dxdy$$

Also:

$$m_{10}L(h,k,\rho)=Z(h)Q\left[\frac{k-\rho h}{\sqrt{1-\rho^2}}\right]+\rho Z(k)Q\left[\frac{h-\rho k}{\sqrt{1-\rho^2}}\right] \qquad \text{(A.2)}$$

$$m_{01}L(h,k,\rho)=\rho Z(h)Q\left[\frac{k-\rho h}{\sqrt{1-\rho^2}}\right] + Z(k)Q\left[\frac{h-\rho k}{\sqrt{1-\rho^2}}\right] \quad (A.3)$$

$$m_{20}L(h,k,\rho)= L(h,k,\rho) + hZ(h)Q\left[\frac{k-\rho h}{\sqrt{1-\rho^2}}\right] +$$

$$+ \rho^2 k Z(k)Q\left[\frac{h-\rho k}{\sqrt{1-\rho^2}}\right] + \frac{\rho\sqrt{1-\rho^2}}{\sqrt{2\pi}} Z\left[\frac{\sqrt{(h^2-2\rho hk+k^2)}}{(1-\rho^2)}\right] \quad (A.4)$$

$$m_{02}L(h,k,\rho) = L(h,k,\rho) + \rho^2 hZ(h)Q\left[\frac{k-\rho h}{\sqrt{1-\rho^2}}\right] +$$

$$+ kZ(k)Q\left[\frac{h-\rho k}{\sqrt{1-\rho^2}}\right] + \frac{\rho\sqrt{1-\rho^2}}{\sqrt{2\pi}} Z\left[\sqrt{\frac{(h^2-2\rho hk+k^2)}{1-\rho^2}}\right] \quad (A.5)$$

where:

$$Z(x) = \frac{1}{2\pi}\exp\left(\frac{-x^2}{2}\right)$$

$$Q(x) = \int_x^{\infty} z(t)dt$$

The adopted estimation procedure for ρ is:

a) Obtain a estimate of m_{11} using $\frac{1}{n}\Sigma x_i y_i$
b) Arbitrarily choose ρ
c) Use equations (A.2) to (A.5) to obtain m_{10}, m_{01}, m_{20} and m_{02}
d) Solve (A.1) and obtain the roots ρ_1 and ρ_2
Make $\rho^*=\rho_1$ if $|\rho-\rho_1| < |\rho-\rho_2|$ or $\rho^*=\rho_2$ else.
e) If $\rho^*\neq\rho$ set $\rho=\rho^*$ and go to step (c). If $\rho^*=\rho$ stop.

Observation: Due to sample variation, equation (A.1) may not have real roots or have roots $|\rho_i| > 1$, i=1,2. In these cases it must be used another starting point (step b).

ACKNOWLEDGMENTS

This research was supported by ELETROBRAS.

REFERENCES

Beard, L.R., 1968. "Simulation of Daily Streamflow", Hydrology Engineering Center, U.S. Army, Technical Paper No-6.

Costa, G. C.; Oliveira, C. G.; Pereira, M. V. F. and Kelman,J., 1981. "Selection of monthly flow generation model for Energy Planning". VI National Seminar on Production and Transmission

of Electric Energy, Camboriu, Brazil. (in portuguese).

Fiering, M. B., 1968. Schemes for Handling Inconsistent Matrices, Water Resources Research, Volume 4, No. 2, pp 291-297.

Kelman, J., 1976. "Stochastic Modelling of intermittent Daily Hydrologic Series", Ph.d thesis, Colorado State University, Fort Collins, Colorado Fall.

Kelman, J., 1980. "A Stochastic model for daily streamflow", Journal of Hydrology, 47 pg. 235-249.

Kelman et al.,1985. "Flood Control in a Multireservoir System", Fourth International Hydrology Symposium, Fort Collins, Colorado, USA.

Matalas, N. C., 1967. "Mathematical Assessment of Synthetic Hydrology", Water Resources Research, Volume 3, No. 4.

O'Connell, P. E.; Jones, D. A., 1979. "Some Experience with the development of models for the stochastic simulation of daily flows", in Inputs for Risk Analysis in Water Systems, Water Resources Publication.

Rosenbaum, S., 1961. "Moments of a truncated Bivariate Normal Distribution", Journal of the Royal Statistical Society, Series B, 23.

Bulu, A., 1979. "Flood Frequency Analysis Based on a Mathematical Model of Daily Flows", in Modelling Hydrologic Processes, Water Resources Publications.

Table 1 . Lag-zero cross correlation comparison
(Above principal diagonal: synthetic estimates)
(Under principal diagonal: historical estimates)

	Emborcação	Itumbiara	São Simão
Emborcação	1,00	0,68	0,50
Itumbiara	0,87	1,00	0,59
São Simão	0,71	0,63	1,00

Table 2. Comparison between statistics of 66 synthetic sequences and one historical sequence, each one of them of 15 flood seasons. In parenthesis statistics obtained with 1000 synthetic sequences. Emborcação.

	DAILY FLOW STATISTICS				
	mean	std.dev.	skew	kurtosis	record
historical	645. (642.)	445. (448.)	1.53 (1.86)	6.09 (9.87)	3520. (6380.)
synthetic minimum	573.	378.	1.07	4.00	2550.
synthetic average	642.	446.	1.75	8.84	3790.
synthetic maximum	707.	575.	3.08	23.65	6380.
P[SYNT>HIST]	0.455	0.424	0.621	0.727	0.530

	ANNUAL MAXIMUM STATISTICS				Regulation
	mean	std.dev.	skew	kurtosis	Volume
historical	1960. (2080.)	754. (740.)	0.18 (1.70)	2.51 (7.14)	6550. (38200.)
synthetic minimum	1720.	384.	-0.67	1.70	1380.
synthetic average	2080.	680.	1.02	3.91	11600.
synthetic maximum	2730.	1300.	2.66	9.59	38200.
P[SYNT>HIST]	0.667	0.333	0.909	0.773	0.727

Table 3. Comparison between statistics of 71 synthetic sequences and one historical sequence, each one of them of 14 flood season. In parenthesis statistics obtained with 1000 synthetic sequences. Itumbiara

DAILY FLOW STATISTICS					
	mean	std.dev.	skew	kurtoses	record
historical	1350. (1360.)	867. (856.)	1.70 (1.55)	7.23 (7.18)	6480. (10400.)
synthetic minimum	1190.	700.	0.93	3.88	4240.
synthetic average	1360.	850.	1.48	6.59	[illegible]440.
synthetic maximum	1510.	1040.	2.32	13.06	10400.
P[SYNT>HIST]	0.648	0.380	0.155	0.197	0.352

ANNUAL MAXIMUM STATISTICS					Regulation
	mean	std.dev.	skew	kurtoses	Volume
historical	3830. (4180.)	1550. (1160.)	0.15 (0.95)	1.85 (4.45)	14200. (54200.)
synthetic minimum	3450.	495.	-0.94	1.51	1130.
synthetic average	4180.	1080.	0.53	2.98	15100.
synthetic maximum	4940.	2100.	2.59	9.06	54200.
P[SYNT>HIST]	0.887	0.028	0.732	0.873	0.479

Table 4 . Comparison between statistics of 40 synthetic sequences and one historical sequence, each one of them of 25 flood seasons. In parenthesis statistics obtained with 1000 synthetic sequences. São Simão

	DAILY FLOW STATISTICS				
	mean	std.dev.	skew	kurtoses	record
historical	900. (932.)	606. (653.)	1.66 (1.67)	8.37 (9.07)	7150. (10900.)
synthetic minimum	886.	592.	1.13	4.89	4320.
synthetic average	932.	652.	1.63	8.66	6750.
synthetic maximum	963.	715.	2.34	18.10	10900.
P[SYNT>HIST]	0.950	0.950	0.425	0.425	0.500

	ANNUAL MAXIMUM STATISTICS				Regulation
	mean	std.dev.	skew	kurtoses	Volume
historical	3220. (3540.)	1180. (1110.)	1.42 (2.19)	5.98 (9.72)	5210. (20700.)
synthetic minimum	3070.	577.	-0.26	1.82	1720.
synthetic average	3540.	1040.	1.41	5.69	8600.
synthetic maximum	4050.	1650.	3.16	14.00	20700.
P[SYNT>HIST]	0.900	0.375	0.525	0.400	0.750