WATER RESOURCES RESEARCH, VOL. 20, NO. 3, PAGES 379–390, MARCH 1984

Stochastic Streamflow Models for Hydroelectric Systems

M. V. F. Pereira, G. C. Oliveira, C. C. G. Costa, and J. Kelman

CEPEL, Centro de Pesquisas de Energia Elétrica

This paper describes the development of a monthly streamflow model for the Brazilian hydroelectric system. The model is based on the disaggregation of lag 1 autoregressive annual flows into monthly values. Model features include addition of new sites, nonparametric generation of monthly flows, and correction of negative values. A methodology for assessing model adequacy is described and applied in a case study comparing the proposed model and a multivariate monthly autoregressive model. The economic effect of model selection is illustrated in a realistic generation planning case study: it is shown that investment differences resulting from the application of different models may reach US $1 billion.

Introduction

Streamflow records play a critical role in the planning and operation studies of a hydroelectric system. For example, a generation expansion plan must meet some reliability constraints (risk of deficit or expected energy not supplied) that are usually estimated by the simulation of system operation over a large number of generated sequences. An "inadequate" streamflow model tends to distort these indices and thus affects the decision making.

The purpose of this paper is to describe the development of a monthly streamflow model for the Brazilian hydroelectric system. The adopted model is based on the disaggregation of annual flows into monthly values. The work is organized in three main parts: (1) theoretical and practical aspects of model development, (2) criteria for assessing the adequacy of a model, and (3) economic effect of model selection. The first part describes the multivariate annual generation scheme, based on the autoregressive model and the seasonal disaggregation of annual values. Problems discussed include the addition of new stations, nonparametric representation of marginal distributions, spatial correlation, and correction of negative flows.

The second part concerns the development of criteria for comparing alternative streamflow models. The adequacy of a model is evaluated on the basis of statistical properties that are considered relevant for power systems planning. A case study illustrates the application of these criteria.

The last part tries to assess the practical implications in terms of system reliability when different streamflow models are used. A planning study with the Brazilian hydroelectric system is used to estimate investment differences resulting from the use of alternative streamflow models.

Theoretical and Practical Aspects of Model Development

The Brazilian hydroelectric system is composed of reservoirs with large storage volumes, which gives it multiyear regulation capability. Monthly time steps are adopted for planning and operation studies. Since the dams are arranged in cascade along the rivers, it is possible to model either the total flows arriving at each site or the incremental flows corresponding to the drainage area limited by this site and upstream stations. This last option is presently recommended for planning studies of the Brazilian hydroelectric system [*Costa et al.*, 1981].

The adopted streamflow generation scheme can be summarized as follows.

1. A three-parameter lognormal distribution is fitted to the incremental annual flows.
2. The annual process (logarithms of annual flows) is generated in a multivariate manner.
3. The annual process is exponentiated to produce annual flows.
4. The annual flows are disaggregated on a unisite basis into monthly flows.
5. Negative monthly flows are corrected.
6. A summation of incremental monthly flows along the cascades is performed to obtain total flows.

Representation of the Annual Marginal Distribution

A three-parameter lognormal distribution is fitted to the annual flow X. Consequently, $W = \ln(X - \alpha) - \mu'$ is normally distributed with zero mean and standard deviation of σ'. Parameters α, μ', and σ' are estimated to preserve the moments of the flows, although other estimation procedures could be used (*Stedinger* [1980] compares several estimation options). Parameters are calculated through the following relations [*Charbeneau*, 1978]:

$$\begin{aligned} \sigma' &= (\ln(\phi))^{1/2} \\ \mu' &= \ln[\sigma/(\phi(\phi - 1))^{1/2}] \\ \alpha &= \mu - \sigma/(\phi - 1)^{1/2} \end{aligned} \tag{1}$$

where μ and σ represent the mean and standard deviation of the annual flows, ϕ is the (only) real root of the equation $\phi^3 + 3\phi^2 - (4 + \gamma^2) = 0$, and γ is the skewness coefficient of the annual flows. If the sample skewness coefficient turns out to be negative, a normal distribution is fitted to the annual flow X.

Annual Generation Scheme

Annual flows are generated by a multivariate autoregressive lag 1 model of the type described by *Matalas* [1967]:

$$W_t = FW_{t-1} + HV_t \tag{2}$$

Paper number 3W0961.
0043-1397/84/003W-0961$05.00

where

- W_t n-dimensional vector related to the annual flows as described in the previous section;
- V_t n-dimensional vector of residuals (standardized independent variables);
- F, H $n \times n$ matrices;
- n number of stations.

F and H are estimated to preserve the lag 0 and lag 1 correlations between the stations [*Matalas*, 1967]. However, planning studies usually require the joint simulation of the existing hydroelectric system with the new plants being added. The comparison of expansion alternatives is made easier if it is possible to produce synthetic streamflows only to the new sites under consideration, while the generated flows to the existing sites remain unchanged.

The Matalas model was then extended to preserve the lag 0 and lag 1 spatial correlation between the new series and the previously generated sequences for k stream sites:

$$W_t = FW_{t-1} + GR_t + HV_t \qquad (3)$$

where W_t, F, and V_t are as described previously, R_t is a k-dimensional vector of zero mean normal variables previously generated for the k stream sites by (2), and G is a $n \times k$ matrix. Applying $E(\ )$ and Cov ( ) to (3), we obtain

$$E(W_t|W_{t-1} = w, R_t = r) = [F\ G]\begin{bmatrix} w \\ r \end{bmatrix} \qquad (4)$$

$$\text{Cov}\,(W_t, W_t|W_{t-1} = w, R_t = r) = H\,\text{Cov}\,(V_t, V_t)H' = HH' \qquad (5)$$

Given that W_t and R_t are multivariate normally distributed and that Cov (W_t, W_t) is equal to Cov (W_{t-1}, W_{t-1}) (that is, the stationarity hypothesis), the conditioned moments can be derived and F, G, and H can then be calculated as the solution of

$$[F\ G] = [\text{Cov}\,(W_t, W_{t-1})\ \text{Cov}\,(W_t, R_t)] \cdot \begin{bmatrix} \text{Cov}\,(W_t, W_t) & \text{Cov}\,(W_{t-1}, R_t) \\ \text{Cov}\,(R_t, W_{t-1}) & \text{Cov}\,(R_t, R_t) \end{bmatrix}^{-1} \qquad (6)$$

$$HH' = \text{Cov}\,(W_t, W_t) - [F\ G]\begin{bmatrix} \text{Cov}\,(W_{t-1}, W_t) \\ \text{Cov}\,(R_t, W_t) \end{bmatrix} \qquad (7)$$

Equation (6) can be solved by partition of the inverse and (7) by spectral decomposition of the right-hand side.

The Seasonal Disaggregation Model

Disaggregation models in hydrology were first suggested by *Valencia and Schaake* [1973]. The VS (Valencia-Schaake) model can be described as

$$Y_t = AX_t + CV_t \qquad (8)$$

where

- Y_t 12-dimensional column vector of zero mean;
- X_t scalar of zero mean;
- V_t 11-dimensional column vector of residuals (independent standardized random variables);
- A 12×1 matrix;
- C 12×11 matrix.

X_t represents the flow during year t (minus its mean) and Y_t the monthly flows for the same year (minus their means). A and C are such that the covariance matrices S_{YY} and S_{YX} are preserved, that is, are equal to the historical sample matrices $\hat{S}_{YY}$ and $\hat{S}_{YX}$. It can also be shown that

$$DA = 1 \qquad DC = 0 \qquad (9)$$

where $D = [1/12\ 1/12\ \cdots\ 1/12]$ is a 12-dimensional row vector.

The above properties ensure that the disaggregated monthly values Y always add up exactly to the given annual value X. This is easily seen by multiplying both sides of (8) by D, which results in

$$DY_t = X_t \qquad (10)$$

In the original reference, C was represented as a 12×12 matrix $\tilde{C}$, calculated as the solution of

$$\tilde{C}\tilde{C}' = M \qquad (11)$$

where M is a 12×12 matrix obtained from historical covariance matrices.

Equation (11) can be solved by spectral decomposition of M, that is,

$$\tilde{C} = \tilde{P}\tilde{\lambda}^{1/2} \qquad (12)$$

where $\tilde{P}$ is the 12×12 eigenvector matrix of M and $\tilde{\lambda}$ is the 12×12 diagonal matrix of eigenvalues, ordered in decreasing magnitude. However, the fact that $D\tilde{C} = 0$ implies that $\tilde{C}$ has rank 11 and thus that the last eigenvalue λ_{12} is null. In this case the last column of $\tilde{C}$ in (12) is equal to zero and should be dropped. The solution of (11) is then rewritten as a 12×11 matrix C, defined as

$$C = P\lambda^{1/2} \qquad (13)$$

where P is the 12×11 matrix of eigenvectors (last column dropped) and λ is the 11×11 diagonal matrix of nonzero eigenvalues.

Although the VS model has many interesting properties, the representation of links between the years is not accurate. In other words, the VS model is not able to preserve the correlation between December of one year and January of the following year [*Mejia and Rousselle*, 1976]. The VS model also introduces an odd correlation between months of different years. *Kelman et al.* [1979] have shown that

$$E(Y_{t,i}Y_{\tau,j}) = E(Y_{t,j}Y_{\tau,i}) \qquad \forall\, t \neq \tau \qquad \forall\, i, j = 1, 2, \cdots, 12 \qquad (14)$$

In other words, the VS model implies that the covariance between December of year t and January of year $t + 1$ (1-month lag) is equal to the covariance between January of year t and December of year $t + 1$ (23-month lag).

In order to improve the interannual representation, *Mejia and Rouselle* [1976] introduced the following extension in the VS model:

$$Y_t = AX_t + BZ_t + CV_t \qquad (15)$$

where Y_t, X_t, V_t, A, and C are as in (8), Z_t is a p-dimensional vector of zero mean, and B is a $12 \times p$ matrix. Z_t represents the last p streamflows of the previous year $t - 1$. According to Mejia and Rouselle, A, B, and C are such that matrices S_{XZ}, S_{YZ}, S_{YX}, and S_{YY} are preserved, thus providing a correct representation of the link between years. However, *Kelman*

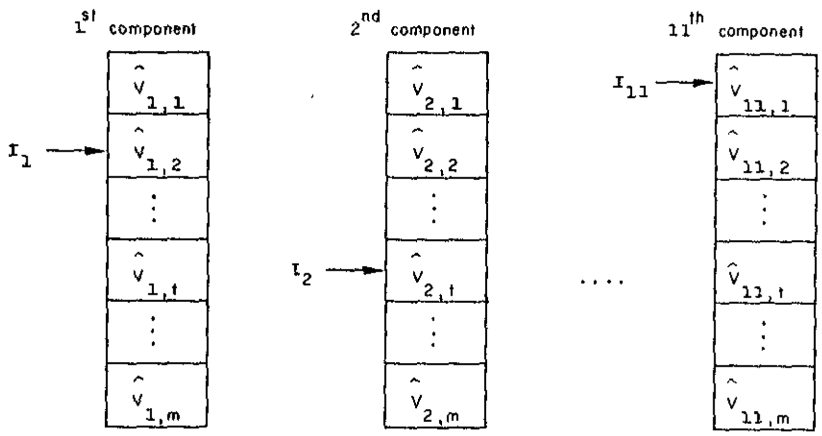


Fig. 1. Nonparametric generation of monthly flows for a historical record of M years.

et al. [1979] and *Lane* [1980] have independently pointed out that S_{yy}, S_{xz}, S_{xy}, and S_{yz} are not preserved when A, B, and C are calculated as suggested by the MR (Mejia and Rouselle) formulation. Kelman et al. derived an analytical expression for estimating the resulting S_{yy} and S_{yz} when $p = 1$. The article by Lane presents a procedure for correcting A, B, and C so as to make S_{yy} equal to $\hat{S}_{yy}$ at the expense of S_{yz}.

Experience with the analytical expression indicated that disturbances are in general much smaller than parameter uncertainty [*Kelman et al.*, 1979]. Therefore no correction schemes were used in the adopted model.

Although the VS and MR schemes can be easily extended to the n-stations case, handling of the resulting arrays is very uncomfortable. The adopted model restricts the multisite generation to the annual level, that is, the annual flows are generated by the multivariate scheme described previously and the monthly values are obtained for each station separately, using the MR scheme for $p = 1$.

Representation of the Monthly Marginal Distribution

Kelman et al. [1979] and *Todini* [1980] have independently shown that the skew coefficients of the monthly flows can be expressed as linear functions of the skewness of X_t (the annual flow) and of V_t, the vector of standardized residuals. These relations are used to calculate the skew coefficients of V_t so as to incorporate any specified skewness in the seasonal flows.

Experiments with several streamflow series have shown that in order to preserve the historical skew coefficients it is often necessary to generate residuals with very high skew values ($\gamma > 30$). Similar values were found by *Todini* [1980] for the Nile river. However, experience with synthetic traces has also shown that the sample moments of these highly skewed residuals may be distorted.

For this reason a nonparametric approach in which residuals V_t are generated from the empirical cumulative distribution of historical residuals $\hat{V}_t$ was developed. The monthly marginal distribution is thus represented indirectly by means of the residuals. The calculation of these residuals proceeds as follows:

From (15) the historical residuals $\hat{V}_t$ can be written as the solution of

$$C\hat{V}_t = \hat{Y}_t - A\hat{X}_t - B\hat{Z}_t \qquad (16)$$

Since C is 12×11, the least squares solution of (16) is given by

$$\hat{V}_t = C^{+}(\hat{Y}_t - A\hat{X}_t - B\hat{Z}_t) \qquad (17)$$

where C^{+} is a 11×12 matrix called the pseudoinverse of C. Since C has rank 11, C^{+} is obtained as

$$C^{+} = (C' C)^{-1} C' \qquad (18)$$

By substituting (13) into (18) and noting that eigenvectors are orthonormal, the calculation is simplified to

$$C^{+} = (\lambda^{1/2} P' P \lambda^{1/2})^{-1} C' = \lambda^{-1} C' \qquad (19)$$

The final expression for the residuals is obtained by substituting (19) into (17):

$$\hat{V}_t = \lambda^{-1} C'(\hat{Y}_t - A\hat{X}_t - B\hat{Z}_t) \qquad (20)$$

Residuals are generated by sampling indices for each component of the historical residual vectors. Figure 1 illustrates the process for a historical sample of m years. In the figure, $\hat{V}_{i,t}$ represents the ith component of the historical residual vector in year t ($i = 1, 2, \cdots, 11$; $t = 1, 2, \cdots, m$). This residual is indexed by I_i, which is sampled from a uniform distribution. Independent sampling of indices $I_1, I_2, \cdots, I_{11}$ produces a complete residual vector V. Therefore there will be $(m)^{11}$ different residual vectors, which is a large "population," given typical historical record lengths of $m = 30$ to 50 years.

One frequent question concerning the use of nonparametric generation schemes is whether it is possible to generate flows that are higher or lower than those found in the historical record. The answer in this case is easily seen to be positive. In fact, about half of the synthetic sequences used in the case study discussed later on in this paper presented flows that were beyond the range of historical values.

Representation of Seasonal Cross Correlation

If the disaggregation of annual values is done separately for each station, the only "source" of seasonal cross correlation in (15) comes from the annual value X_t. As a consequence, monthly cross correlation tends to be smaller than shown in historical records. However, with the nonparametric approach it is possible to generate cross-correlated residuals V_t and thus significantly improve the representation of spatial dependence of monthly flows.

The implementation of this improvement is very easy: when generating streamflows for station l for a given year t,

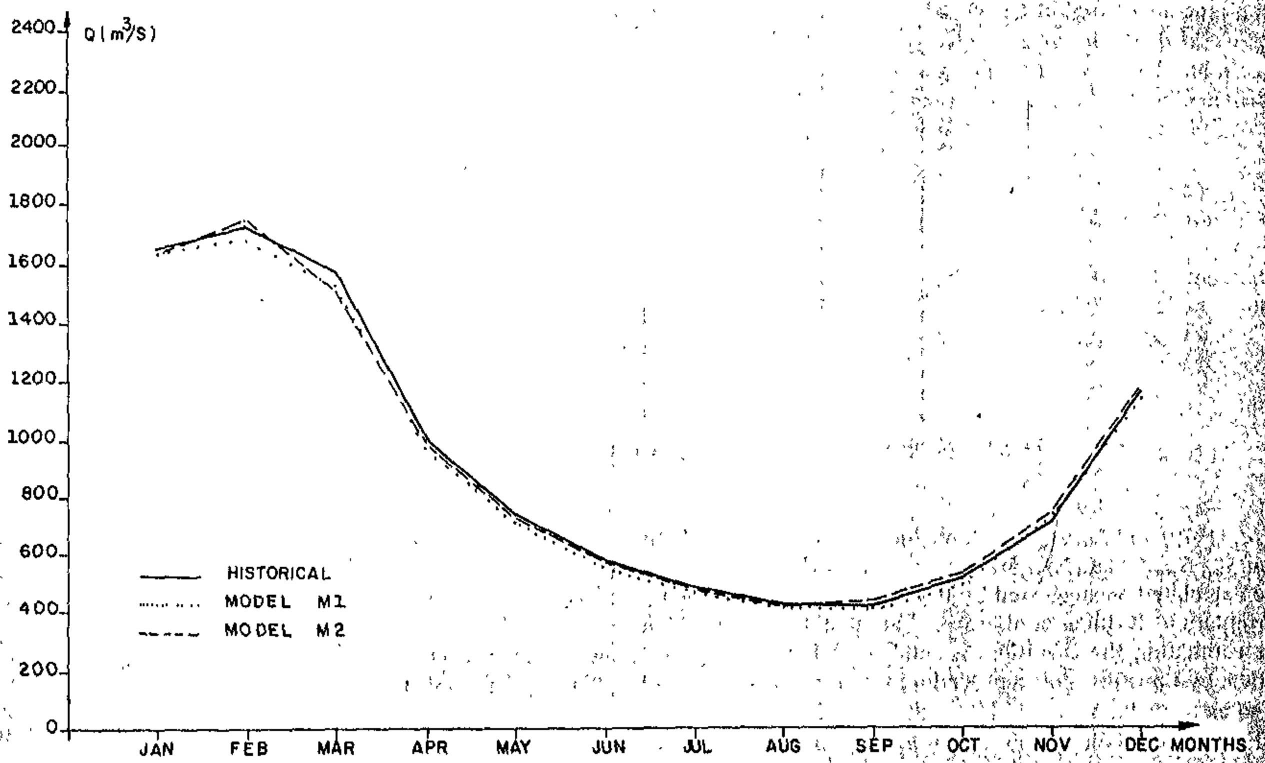


Fig. 2. Monthly means for Furnas.

it is only necessary to use the same indices $I_1, I_2, \cdots, I_{11}$ that were previously sampled for station k for the same year t. It should be noted that only the indices are the same but not the historical residual values, which will vary from station to station. It can be seen that this scheme will preserve the spatial dependence between each component of the residual vectors and thus improve the representation of the cross correlation between flows.

It is also interesting to note that in actual generation it is not even necessary to record the sequence of sampled I_i; one only needs to store the initial "seed" of the pseudorandom number generator which produced that sequence and reproduce it when needed. In this way, multivariate correlation can be represented without affecting the "univariate" character of the disaggregation process.

Correction of Negative Flows

During disaggregation of annual values, it is possible to generate negative monthly flows. This is a common problem in cascaded systems, where incremental flows can be relatively small.

A computational scheme was then developed in which the consequences of eliminating negative flows are attenuated by redistributing the correction throughout the year. The scheme involves the solution of the following problem:

$$\text{Min} \sum_{i=1}^{12} (Y_i^* - Y_i)^2/\sigma_i^2 \qquad (21)$$

subject to

$$Y_i^* \geq 0 \qquad \forall\, i = 1, 2, \cdots, 12$$

$$\sum_{i=1}^{12} Y_i^* = \sum_{i=1}^{12} Y_i$$

where

Y_i^* corrected flow;
Y_i generated flow;
σ_i monthly standard deviation.

The objective function corresponds to a weighted least squares minimization; the inequality constraints impose the nonnegativity of the incremental flows; the equality constraint preserves the yearly total. A very efficient solution algorithm is described by *Lawson and Hanson* [1974, pp. 165–167].

The application of the correction scheme may affect the moments of the generated samples. One can see that the flows in the "wet" seasons (higher σ) should be more affected due to the weighting factors.

CRITERIA FOR ASSESSING THE ADEQUACY OF A MODEL

An useful streamflow model should preserve important features of the "real" stochastic process. In practice, this adequacy is measured by the model's capability of reproducing the probability distribution of some relevant random variables. The selection of the random variables should naturally reflect the requirements of the proposed application. In the case of expansion planning for power generation, such requirements include the representation of critical periods, time and space correlation between the flows to the various reservoirs, accumulated inflow to the reservoirs during droughts, etc. [*Kelman and Pereira*, 1977].

It is important to note that any characteristic of the process that has been used in the definition of the streamflow model cannot be used to assess its adequacy. For example, if the seasonal means are parameters of a model being ana-

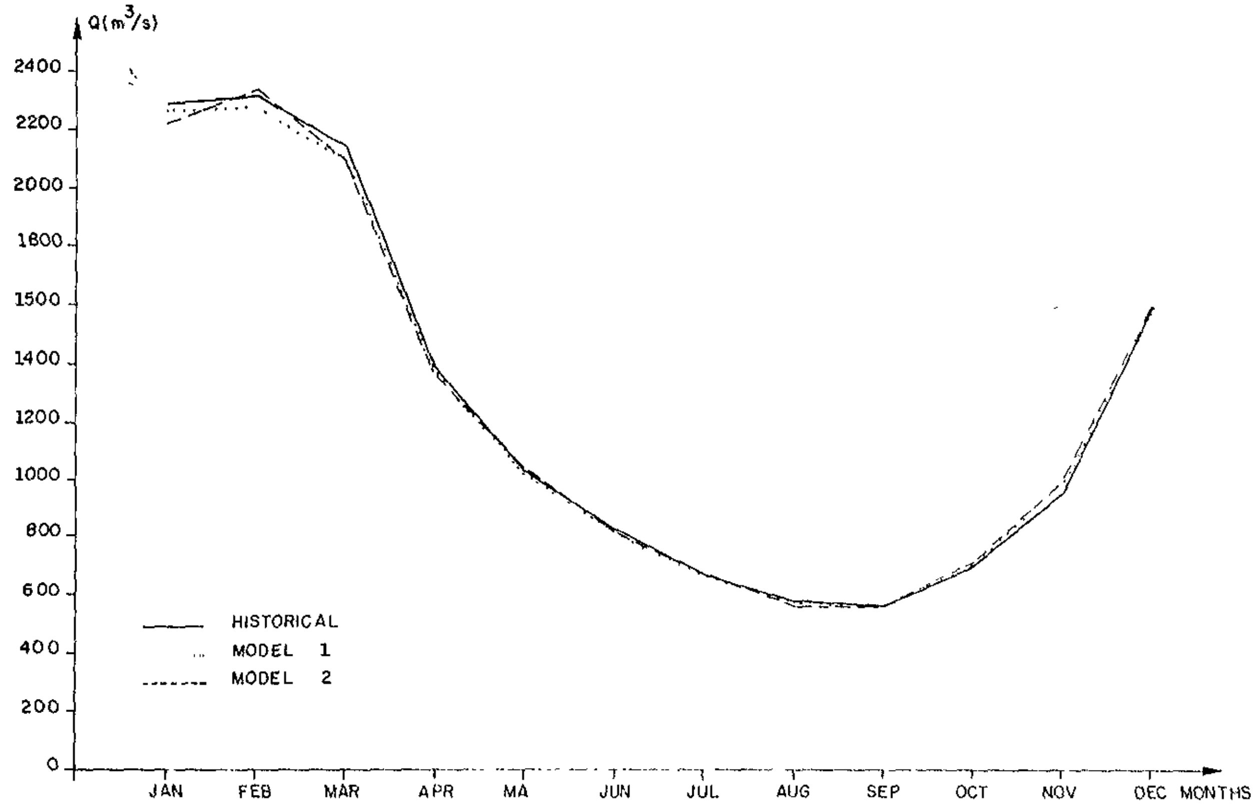


Fig. 3. Monthly means for P. Colombia.

lyzed, it is useless to compare the historical and generated means: one knows beforehand that they will be preserved. Such a comparison should only be used to verify the computer program and not to validate the model [*Stedinger and Taylor*, 1982].

In practice, however, moments are verified in search of inconsistencies or deviations. For example, monthly means and variances of incremental flows are parameters of the model and should thus be exactly preserved. On the other hand, variance of total flows depend on the spatial covariance between incremental flows upstream of each site. Since this spatial covariance is only indirectly represented in the

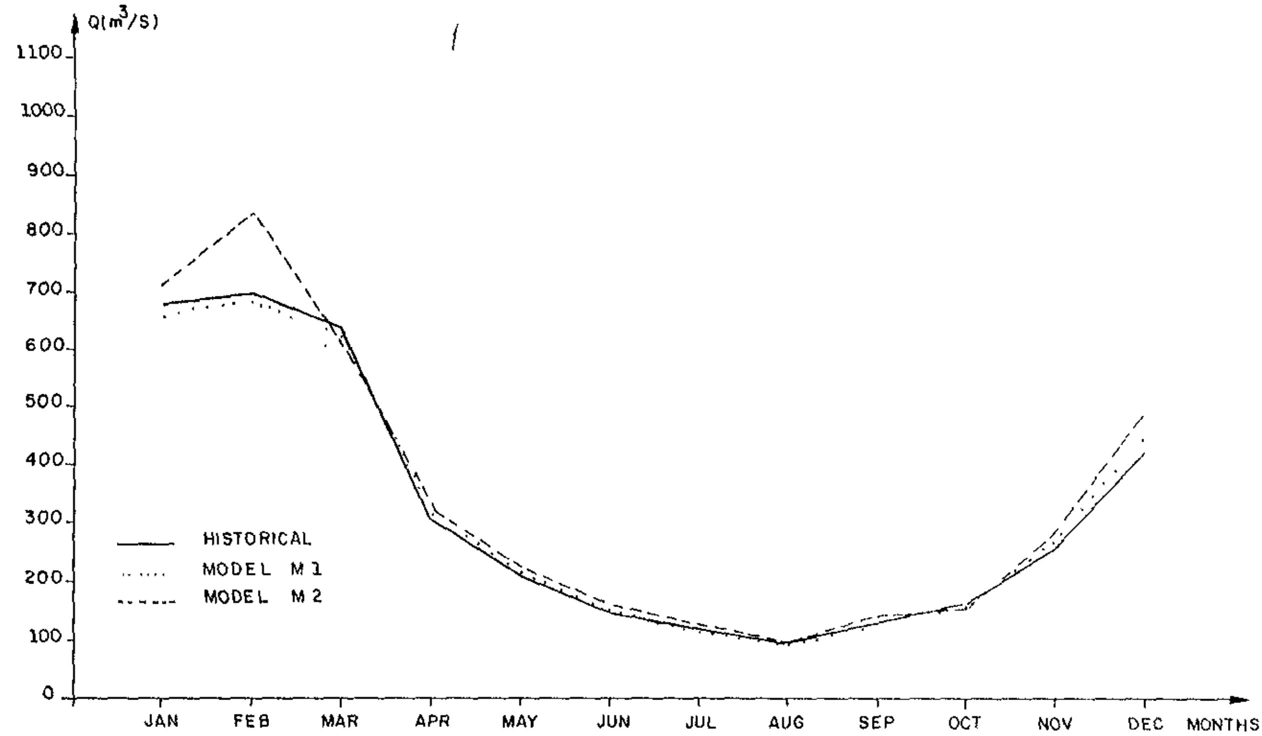


Fig. 4. Monthly standard deviations for Furnas.

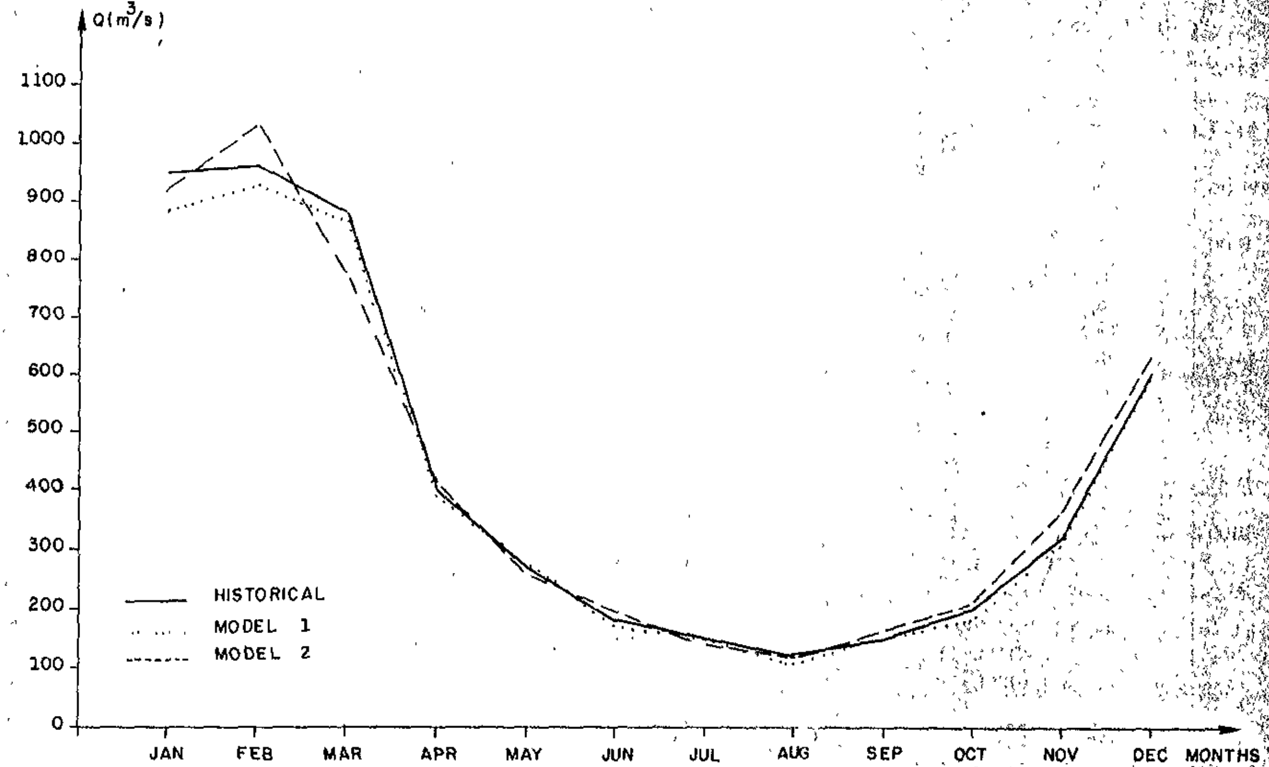


Fig. 5. Monthly standard deviations for P. Colombia.

nonparametric approach, variance of total flows is not expected to be exactly reproduced. It should also be remembered that the correction of negative flows may affect the moments of the generated sequences and that covariance matrices S_{yy} and S_{yz} are not exactly preserved.

The selection of the relevant random variables is a complex matter. The subject was discussed with engineers from major utilities and from Eletrobrás, the holding company for power generation in Brazil [*Kelman and Pereira*, 1977]. It was generally accepted that the random variables to be selected should be associated with the concept of run: run sum, run length, and average run intensity [*Yevjevich*, 1972]. The analysis of the model with regard to these variables was made through goodness of fit tests between the probability

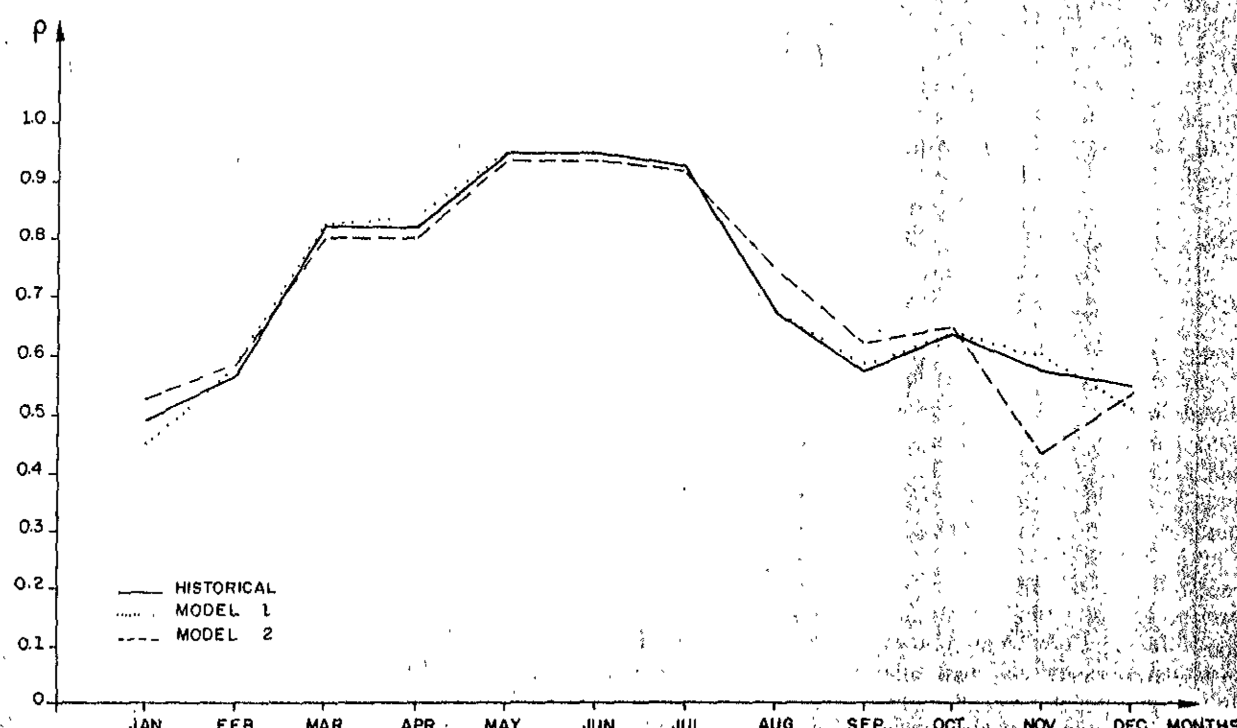


Fig. 6. Monthly autocorrelations for Furnas.

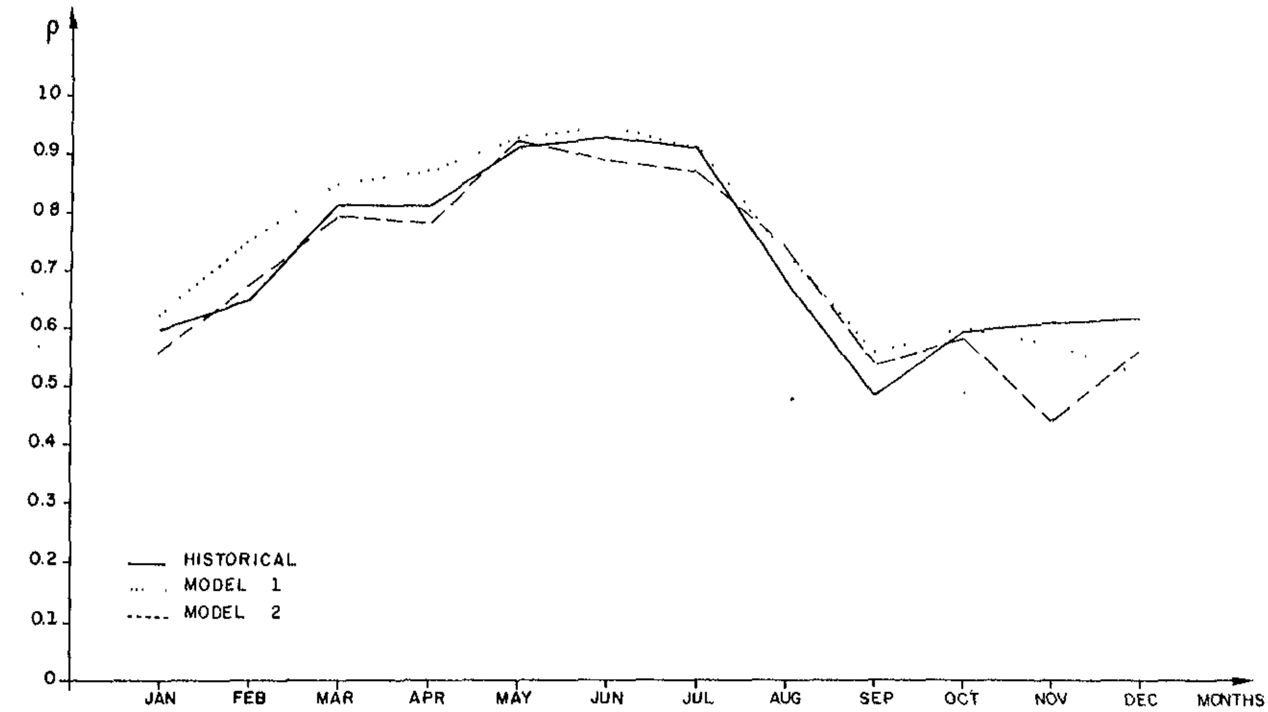


Fig. 7. Monthly autocorrelations for P. Colombia.

distributions obtained from the historical record and the ones obtained from generated samples.

The maxima of these variables are also relevant to energy planning. The critical period, for example, corresponds to the worst hydrological situation in the historical record. The importance of maximum values is particularly relevant for systems with large storage capacity. These systems are insensitive to the smaller short-term perturbations.

Maximum values can be computed for each generated sample of the same length as the historical record. Since the historical record itself produces only one maximum value, the adequacy of the model is measured by the proportion of generated values smaller or bigger than the historical. A very small proportion indicates that the model may be inadequate, since it "considers" the historical sample as an atypical realization of the stochastic process.

One variable, the maximum deficit, was selected as the most relevant to generation expansion planning. The maximum deficit variable represents the minimum reservoir storage necessary to meet a prestablished regulated outflow q^* for a given streamflow sequence [*Gomide*, 1978]. It is easy to see that the maximum deficit is directly related to the representation of the worst drought in the period.

Experience has shown that the maximum run sum, run length, and intensity are highly correlated with the maximum deficit, that is, the additional information brought by these

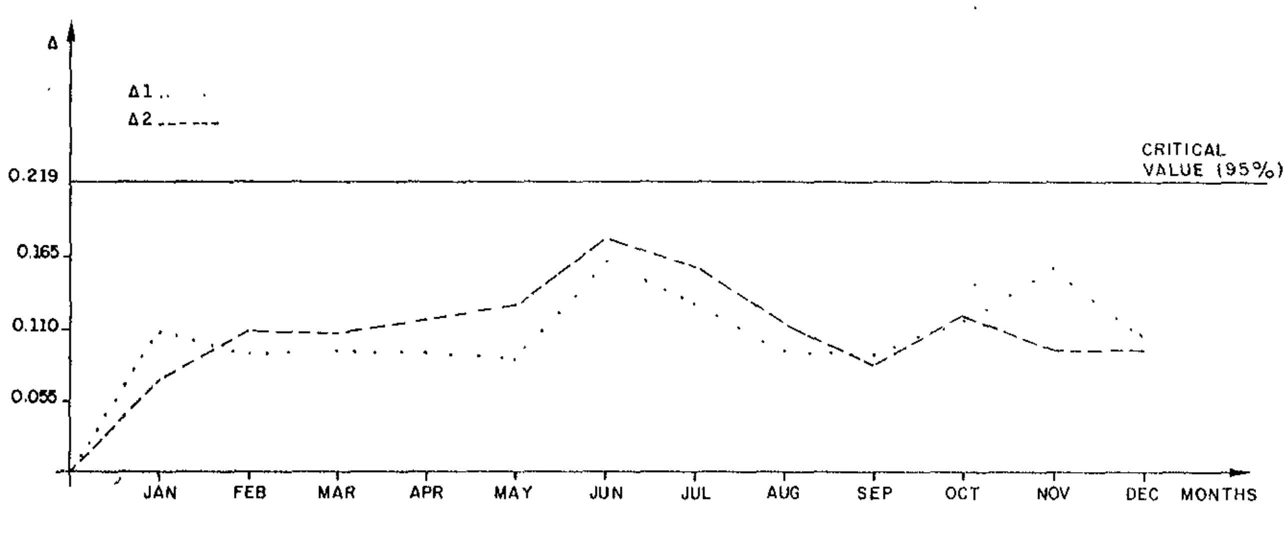
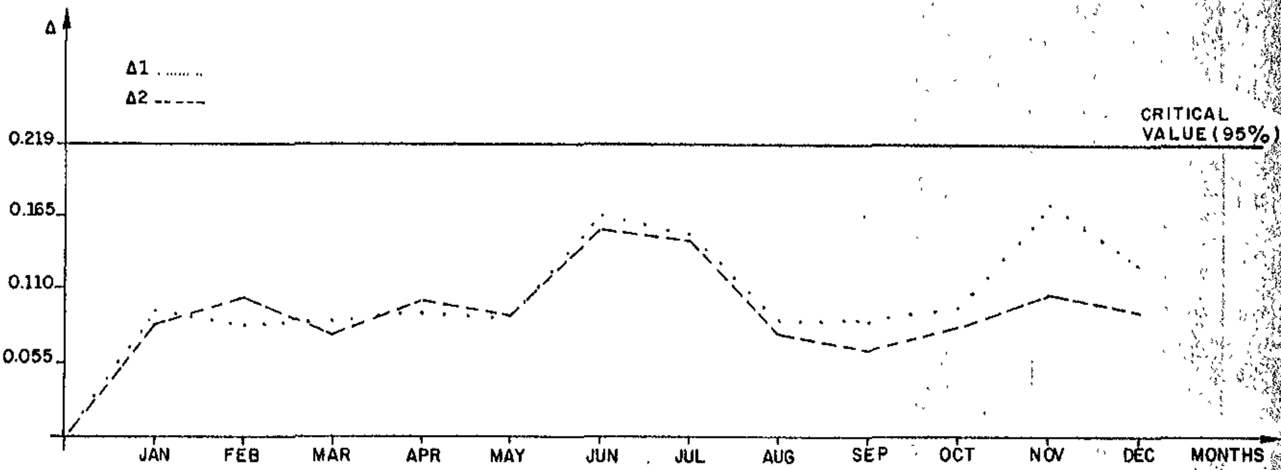


Fig. 8. Goodness of fit test between the historical and generated cumulative distributions (Δ is the maximum difference between the empirical probability distribution of the historical and generated samples) for Furnas.



RESULT OF THE TWO-SAMPLE SMIRNOV TEST

Fig. 9. Goodness of fit test between the historical and generated cumulative distributions (Δ is the maximum difference between the empirical probability distribution of the historical and generated samples) for P. Colombia.

indices was small [*Costa et al.*, 1981]. For this reason, and for the sake of simplicity, they were not included in the case study discussed in the following section.

Application of Assessment Criteria: A Case Study

The application of the criteria for adequacy assessment will be illustrated in a case study with two cascaded streamsites in Rio Grande, Brazil: Furnas and Porto Colombia. We generated 25 samples of 40 years each, corresponding to the historical record, for the two stations, using two alternative models: the adopted streamflow model (M_1) and a seasonal autoregressive model (M_2). In this model, each monthly flow has a two-parameter lognormal distribution:

$$\ln (Y_{i+1,t}) = \mu_{i+1}' + \rho_i' \frac{\sigma_{i+1}'}{\sigma_i'} (\ln (Y_{i,t}) - \mu_i') + \sigma_{i+1}'(1 - \rho_i'^2)^{1/2}\eta_{i+1,t} \qquad (22)$$

where

- $Y_{i,t}$ inflow for month i, year t;
- i indexes the months, equal to 1, 2, $\cdots$, 12; $i + 1 = 13$ implies $i + 1 = 1$;
- t indexes the years, equal to 1, 2, $\cdots$;
- μ_i' monthly mean of inflow logarithms;
- σ_i' monthly standard deviation of inflow logarithms;
- ρ_i' monthly autocorrelation coefficient of inflow logarithms;
- $\eta_{i+1,t}$ standardized residual.

More details about parameter estimation and multivariate generation scheme can be found, for example, in works by *Matalas* [1967] and *Salazar et al.* [1977].

The verification phase of the assessment procedure concerns the preservation of sample moments. Figures 2 through 7 show the monthly means, standard deviations, and lag one autocorrelation coefficients estimated from the historical record and from the synthetic samples produced by both models. Figures 8 and 9 show the results of goodness of fit tests between the historical and synthetic cumulative distributions. The performance of both M_1 and M_2 can be considered equally good in all these aspects.

The validation phase starts with goodness of fit tests for the run sum, run length, and average intensity statistics. The results are summarized in Table 1 and indicate that both models are very adequate with regard to these variables.

However, a sharp difference appears in the calculation of the maximum deficit variable. It can be seen in Table 2 that about half of the values produced by model M_1 are higher than the historical. Since the maximum deficit is closely related to the worst drought in the period, this means that half of the sequences generated by M_1 produced more severe droughts than found in the historical record. Model M_2, on the other hand, considers the historical drought as a very unlikely event, since none of the 25 values obtained from the generated sequences exceeded the historical one.

In terms of model choice, the above results indicate that model M_1 should be preferred to M_2. This choice also favors security, since model M_2 would tend to produce less severe droughts than those found in the historical record.

The differences observed in the maximum deficit tests seem surprising, given the equally good performance of both models in the previous tests. A partial explanation can be found by looking at the annual moments of the generated sequences.

It was verified that the annual standard deviation and autocorrelation produced by M_2 were smaller than the historical values, as can be seen in Table 3. As a consequence, annual droughts produced by M_2 tend to be less severe than those produced by M_1. M_2 considers the 5-year worst drought found in the historical record as a very unlikely event. A very detailed analysis of these aspects is found in the work by *CEPEL* [1978].

The actual choice for the Brazilian system involved 34 stations covering the main Brazilian basins and confirmed the simple conclusions of this case study. Further details can be found in the work of *Kelman et al.* [1981].

It remains to be seen whether model choice has any practical effect in generation planning studies. This subject will be discussed in the following section.

TABLE 1. Goodness of Fit Tests For Run Sum, Run Length, and Run Intensity

Station	Source of Data	Run Sum, 10^9 M^3 Mean	Run Sum Standard Deviation	Run Sum Δ	Run Length, months Mean	Run Length Standard Deviation	Run Length χ^2	Run Intensity, 10^9 M^3/month Mean	Run Intensity Standard Deviation	Run Intensity Δ
Furnas	Historical records	3.0	6.7		4.8	7.2		0.4	0.4	
	M_1	2.9	5.2	0.06	4.3	5.0	8.0	0.5	0.5	0.10
	M_2	3.0	4.3	0.12	4.9	5.2	5.3	0.5	0.4	0.10
P. Colombia	Historical records	3.7	8.7		4.4	6.7		0.6	0.5	
	M_1	3.8	7.6	0.09	4.6	5.9	3.7	0.6	0.6	0.08
	M_2	3.6	5.3	0.10	4.5	4.9	2.0	0.6	0.5	0.10

Δ is the maximum difference between the empirical probability distributions of the historical and generated samples. For the above cases, the critical values of the two-sample Smirnov test for the Furnas and P. Colombia stations are 0.192 and 0.184, respectively (5% significance level). The χ^2 measures the goodness of fit between two discrete distributions (multinomial test). The critical value for the above cases at 5% significance level is 9.49.

THE ECONOMIC EFFECT OF STREAMFLOW MODEL CHOICE

It has been argued in the literature [*Klemeš et al.*, 1981] that the maximum deficit index tends to "inflate" the differences between alternative streamflow models. In other words, models that would be very different in terms of required capacity for a fixed reliability turn out to be very similar when viewed in terms of reliability for a fixed capacity. Since this last measure is what really matters in practical terms, the actual impact of model choice in planning could thus be very reduced.

However, one of the main conclusions of our experience with the application of streamflow models is that model choice does have practical implications in terms of system reliability and planning investments. This will be illustrated with an example in generation planning.

The generation reliability problem is to evaluate the ability of a system to supply the load demand, taking into account load fluctuations and random equipment outages. In generation systems that are predominantly hydro, the limits on the peak capacity of the hydro plants have two different causes: (1) loss of head due to reservoir depletion and (2) equipment outages.

The loss of head effect is particularly severe in the Brazilian generation system, where 90% of the generation comes from hydro units. For example, the loss of available power due to reservoir depletion in the South/Southeast hydro subsystem planned for 1987 may reach 5000 MW, about 12% of the total installed capacity of 40,000 MW. For other systems this loss may may go to up to 20%. Random equipment outages, in turn, reduce the number of working units in a given period, thus decreasing even more the system generating capacity.

The probabilistic evaluation of power deficits in a hydrosystem therefore requires a specific methodology that takes into account the joint effect of reservoir depletion and equipment outage. This methodology is described in detail by *Cunha et al.* [1982] and is illustrated in Figure 10.

The system data includes the detailed description of the hydroelectric plants in the system (configuration, the inflow sequences arriving in each period to each reservoir (inflow sequences) and information about the system depletion policy such as priorities and rule curves (operating rules). The load model represents the monthly load duration curves discretized into equiprobable intervals. The system opera-

TABLE 2. Calculation of Maximum Deficit For Furnas and P. Colombia

Regulated outflow, % of mean	Source of Data	Maximum Deficit, 10^9 m^3 Mean	Maximum Deficit Standard Deviation	Number of Sequences With Values Above Historical
	Furnas Station			
70%	historical records	11.3		
	M_1	15.2	7.9	15
	M_2	7.8	1.3	none
75%	historical records	17.4		
	M_1	18.9	10.3	13
	M_2	10.4	2.1	none
80%	historical records	23.8		
	M_1	23.6	12.9	11
	M_2	13.7	2.9	none
	P. Colombia Station			
70%	historical records	16.2		
	M_1	18.2	9.8	14
	M_2	9.1	1.6	none
75%	historical records	24.8		
	M_1	23.4	13.1	10
	M_2	12.4	2.4	none
80%	historical records	33.6		
	M_1	30.1	17.0	10
	M_2	16.8	3.7	none

TABLE 3. Annual Standard Deviation and Lag One Autocorrelation For Furnas and P. Colombia

Source of Data	Annual STD DEV (m^3/s)	Annual Autocorrelation
	Furnas Station	
Historical records	238	0.33
M_1	238	0.37
M_2	218	0.11
	P. Colombia Station	
Historical records	333	0.41
M_1	337	0.44
M_2	276	0.13

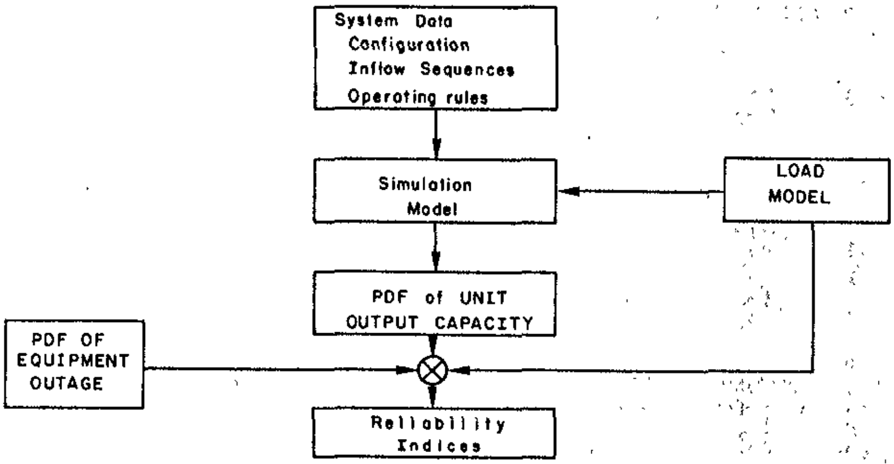


Fig. 10. Methodology for reliability evaluation in a hydroelectric system.

tion is then simulated in order to calculate the available unit output capacity of each plant in each month for a given streamflow sequence. The simulation is repeated for many sequences, thus producing samples of the multivariate probability distribution of unit output capacities.

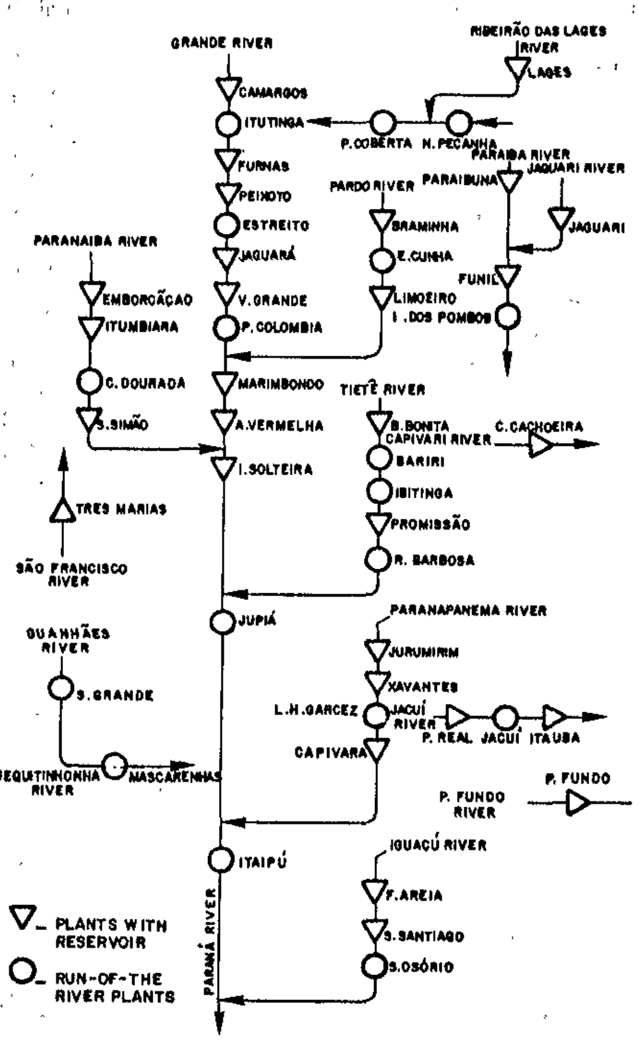


Fig. 11. Schematic representation of the South/Southeast Hydroelectric System. Note that run of the river plants may have large reservoirs which are usually not depleted due to the operation rules.

The available unit capacities in each sample i are then convoluted with the probability distribution of unit outages to produce the conditioned system generating capacity G_i. The conditioned loss of load probability (CLOLP) is defined as

$$\mathrm{CLOLP}_i = P\,[G_i < L] \tag{23}$$

where L is the monthly load and G_i is the system generating capacity given the ith sample. For n equiprobable samples, the system loss of load probability, or LOLP, is estimated as

$$\mathrm{LOLP} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{CLOLP}_i \tag{24}$$

In generation expansion planning the system installed capacity is determined by the target LOLP values, which give a measure of the desired system reliability. Since LOLP evaluation involves the simulation of the system operation over many streamflow sequences, the difference between streamflow models may be measured in practice by the difference in the resulting LOLP values.

Figure 11 shows a schematic diagram of part of the Brazilian hydroelectric system planned for 1987. This configuration corresponds to the south and southeast regions, where the main load centers are located. Total installed capacity is 39,464 MW. The monthly energy load is 16,400 GWh (22,780 average MW). Peak load is 37,000 MW with a loading factor of 0.62. In contrast with other countries, the load in the Brazilian system does not present significant seasonal variations. For this reason, monthly load distribution was considered constant along the year. Further information about the system can be found in Table 4.

The system operation was simulated twice for 1000 years using monthly data from the two different streamflow models discussed in the previous case study. Table 5 shows the resulting LOLP values. It can be seen that the LOLP associated with the monthly autoregressive model M_2 is much smaller than the LOLP corresponding to the proposed model M_1. This indicates that reservoir levels are higher when simulation uses streamflow sequences from model M_2.

The reliability levels coming out of the M_2 simulations correspond roughly to the target levels used in actual planning. In other words, if M_2 is the "correct" model, the installed capacity of 39,464 MW is adequate to supply a peak

TABLE 4. Plant Data For The South/Southeast Systems

Name	Storage Capacity, 10^6 m^3	Mean Inflow, m^3/s	Average Head, m	Nominal Power, MW	Number of Units	Outage Rate, %
Camargos	792	134	23	45	2	1.3
Itutinga	12	134	29	50	3	1.3
Furnas	23000	912	86	1280	8	1.3
Peixoto	4080	1013	43	477	10	1.3
Estreito	1340	1033	63	1104	6	1.3
Jaquara	450	1044	44	680	6	1.3
V. Grande	2150	1127	26	400	4	1.3
P. Colombia	1450	1257	24	320	4	1.3
Graminha	555	52	93	80	2	1.3
E. Cunha	14	85	88	108	4	1.3
Limoeiro	25	85	23	28	2	1.3
Marimbondo	6150	1714	61	1440	8	1.3
A. Vermelha	11000	1929	53	1380	6	1.3
Emborcação	17946	431	127	1000	4	1.3
Itumbiara	17027	1515	81	2100	6	2.0
C. Dourada	660	1580	32	443	8	1.3
S. Simão	12500	2241	66	1608	6	1.3
B. Bonita	3160	295	20	140	4	1.3
Bariri	544	329	23	143	3	1.3
Ibitinga	985	401	19	131	3	1.3
Promissão	7400	509	26	264	3	1.3
R. Barbosa	2700	581	29	300	3	1.3
I. Solteira	21166	4962	45	3200	20	1.3
Jupiá	3680	5737	22	1414	14	1.3
Jurumirim	6520	193	32	98	2	1.3
Xavantes	8705	250	72	416	4	1.3
L. N. Garcez	48	380	17	70	4	1.3
Capivara	10570	912	45	640	4	1.3
Itaipu	29000	9040	119	12600	18	2.0
F. Areia	5945	520	125	1251	3	2.0
S. Santiago	6750	878	96	1332	4	2.0
S. Osório	1275	926	67	1050	6	1.3
P. Fundo	1560	48	244	220	2	1.3
P. Real	3646	181	40	140	2	1.3
Jacui	29	181	93	168	6	1.3
Itauba	620	259	90	500	4	1.3
Capivari Cachoeira	179	17	740	250	4	1.3
Cubatão	31	97	690	870	14	1.3
Jaguari	1238	32	53	28	2	1.3
Paraibuna	4740	72	81	88	2	1.3
Funil/Paraiba	870	227	70	216	3	1.3
I. Pombos	8	605	31	164	5	1.3
N. Peçanha	33	160	311	378	6	1.3
Lages	601	17	314	142	10	1.3
P. Coberta	22	17	37	96	2	1.3
Sto. Grande	78	178	89	104	4	1.3
Mascarenhas	39	969	17	120	3	1.3
Três Marias	19180	1453	47	388	6	1.3

load of 37,000 MW at a mean risk of 3.8×10^{-4} (0.28 h/month).

On the other hand, if M_1 is a more adequate representation of the streamflow process, it will be necessary to increase the installed capacity to reach the target LOLP values. The additional capacity necessary can be roughly estimated as follows: by iterative calculations, it is possible to determine that the maximum peak load that can be supplied by model M_1 at a risk level of 3.8×10^{-4} is 34,500 MW. Therefore the reserve capacity requirements associated with M_1 are

$$\frac{39464 - 34500}{34500} \approx 14\%$$

Assuming that the same percentage reserve is required to meet a load of 37,000 MW, the extra installed capacity would be

$$1.14 \times 37000 - 39464 = 2716 \text{ MW}$$

TABLE 5. Monthly LOLP Values For The South/Southeast Systems

Model	Loss of Load Probability (LOLP), 10^{-4}											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
M_1	60.3	38.4	21.9	17.8	20.5	23.3	27.4	35.6	43.8	54.8	63.0	65.8
M_2	3.6	3.0	2.7	2.7	2.7	3.0	3.6	4.1	4.9	5.5	5.2	4.4

Scale is 10^{-4}.

Given typical investment costs of US $0.35 million per MW of installed peak capacity, the 2716 MW represent US $950 million of extra investment. This value gives a measure of the practical impact of model choice in planning. It should be stressed that in actual planning the investment decision is much more complex and might even result in a change of the target reliability levels. The point to be made is that significant amounts of money may or may not be invested as a consequence of streamflow model choice.

Conclusions

The development of a stochastic streamflow model comprises many aspects, ranging from the analysis of the theoretical properties of the model to practical problems such as how to avoid the generation of negative inflows or how to handle the addition of new streamflow sites of interest. Since there is no standard way of treating these problems, a "customized" solution rather than an "off the shelf" model had to be developed and tested.

Efficient model-selection procedures are essential for the successful application of stochastic models in planning studies. The adopted validation scheme was able to show sharp differences between models that were apparently very similar. A realistic case study of generation planning indicates that model choice has very important practical implications both in terms of system reliability and investment decisions.

Acknowledgments. This research received financial support from Eletrobrás. The thoughtful comments and suggestions of the anonymous referees are appreciated and have resulted in substantial improvements in the paper.

References

CEPEL, Synthetic streamflow generation (in Portuguese), *Tech. Rep. 334* (supplement), Rio de Janeiro, 1978.

Charbeneau, R., Comparison of the two- and three-parameter lognormal distributions used in streamflow synthesis, *Water Resour. Res., 14*(1), 149–150, 1978.

Costa, C. C. G., G. C. Oliveira, M. V. F. Pereira, and J. Kelman, Selection of monthly streamflow models for generation expansion (in Portuguese), paper presented at the Sixth National Seminar on Production and Transmission of Electrical Energy (VI SNPTEE), Eletrobrás, Florianopolis, Brazil, 1981.

Cunha, S. H. F., F. B. B. Gomes, G. C. Oliveira, and M. V. F. Pereira, Reliability evaluation in hydrothermal generating systems, *IEEE Trans. Power Appar. Syst., PAS-101*(12), 4665, 4673, 1982.

Gomide, F. L. S., Range and deficit analysis using Markov chains, *Hydrol. Pap. 4*(79), Colo. State Univ., Fort Collins, 1978.

Kelman, J., and M. V. F. Pereira, Criteria for the assessment of models of hydrologic series (in Portuguese), paper presented at the Fourth National Seminar on Production and Transmission of Electrical Energy (IV SNPTEE), Eletrobrás, Rio de Janeiro, Brazil, 1977.

Kelman, J., G. C. Oliveira, and M. V. F. Pereira, Synthetic streamflow generation by disaggregation (in Portuguese), paper presented at the Fifth National Seminar on Production and Transmission of Electrical Energy (V SNPTEE), Recife, Brazil, 1979.

Kelman, J., G. C. Oliveira, M. V. F. Pereira, and C. C. G. Costa, Hydrologic series model: Methodology manual, *Tech. Rep. 113/81-A*, CEPEL, Rio de Janeiro, 1981.

Klemeš, V., R. Srikanthan, and T. A. McMahon, Long-memory flow models in reservoir analysis: What is their practical value?, *Water Resour. Res., 17*(3), 737–751, 1981.

Lane, W. L., Corrected parameter estimates for disaggregation schemes, paper presented at the International Symposium on Rainfall-Runoff Modelling, Miss. State Univ., University, 1980.

Lawson, C., and R. Hanson, *Solving Least-Square Problems*, Prentice-Hall, Englewood Cliffs, N. J., 1974.

Matalas, N. C., Mathematical assessment of synthetic hydrology, *Water Resour. Res., 3*(4), 937–945, 1967.

Mejia, J. M., and J. Roussele, Disaggregation models in hydrology revisited, *Water Resour. Res., 12*(2), 185–186, 1976.

Salazar, P. G., M. V. F. Pereira, and J. Kelman, Monthly streamflow generation for energy studies (in Portuguese), paper presented at the Fourth National Seminar on Production and Transmission of Electrical Energy (IV SNPTEE), Eletrobrás, Rio de Janeiro, Brazil, 1977.

Stedinger, J. R., Fitting lognormal distributions to hydrologic data, *Water Resour. Res., 16*(4), 481–490, 1980.

Stedinger, J. R., and Taylor, M. R., Synthetic streamflow generation, 1, Model verification and validation, *Water Resour. Res., 18*(4), 909–918, 1982.

Todini, E., The preservation of skewness in linear disaggregation schemes, *J. Hydrol., 47*, 199–214, 1980.

Valencia, D. R., and Schaake, J. C., Disaggregation processes in stochastic hydrology, *Water Resour. Res., 9*(3), 580–585, 1973.

Yevjevich, V., *Stochastic Processes in Hydrology*, Water Resources Publications, 1972.

C. C. G. Costa, J. Kelman, G. C. Oliveira, and M. V. F. Pereira, CEPEL, Centro de Pesquisas de Energia Elétrica, Cidade Universitaria, Ilha do Fundão, Caixa Postal 2754, 20.000 Rio de Janeiro, RJ Brasil.

(Received January 6, 1982;
revised March 1, 1983;
accepted May 31, 1983.)