

## A NEW APPROACH TO IDENTIFY THE STRUCTURAL ORDER OF PAR ( $p$ ) MODELS

Fernando Luiz Cyrino Oliveira<sup>1\*</sup> and Reinaldo Castro Souza<sup>2</sup>

Received December 17, 2009 / Accepted February 4, 2011

**ABSTRACT.** The periodic autoregressive model, a particular structure of the Box & Jenkins family, denoted by PAR ( $p$ ), is employed to model the series of hydrological streamflow used for estimating the operational costs of the Brazilian hydro-thermal optimal dispatch. Recently, some aspects of this approach began to be studied and several researches on this topic are being developed. This paper focuses on the identification stage of the orders  $p$  of these models. Nowadays, the identification is based on evaluating the significance of the coefficients of the partial autocorrelation function (PACF), based on the asymptotic results of Quenouille. The purpose of this study is on the application of the computer-intensive *Bootstrap* technique to estimate the significance of such coefficients. The results show that identification via *Bootstrap* is considerably more parsimonious, leading to the identification of lower orders in most cases and corroborating some points raised in previous studies on the traditional approach.

**Keywords:** PAR ( $p$ ), Partial Autocorrelation Function, *Bootstrap*.

### 1 INTRODUCTION

Regarding electrical energy generation, Brazil has approximately 90% of its energy supplied by hydroelectric plants. This explains the high levels of investment and research done in order to guarantee and improve all the system of energy production. One of the main characteristics of the generation systems that have hydraulic predominance is the strong dependence of the hydrological regimes. Thus, it is evident the relevance of the streamflow simulation models that seek to optimize the system operations performance, the output being an increase of reliability and costs reduction.

As Hipel & McLeod (1994) showed, some historical series, among them the seasonal hydrological ones show an autocorrelation structure that depends not only on the time interval between the observations, but also on the observed period. Furthermore, according to Salas & Obeysekera

---

\*Corresponding author

<sup>1</sup>Pontifical Catholic University of Rio de Janeiro – PUC-Rio, Rio de Janeiro, RJ, Brazil. E-mail: fcyrino@ele.puc-rio.br

<sup>2</sup>Pontifical Catholic University of Rio de Janeiro – PUC-Rio, Rio de Janeiro, RJ, Brazil E-mail: reinaldo@ele.puc-rio.br

(1982), natural stochastic processes are, in general and in a broad sense, stationary, that is, first and second order movements of the probabilities distribution are not affected by variations related to the choice of the time origins, which is one of the assumptions to the application of Box & Jenkins methodology, Harvey (1981). In the periodic processes case, two models structures can be considered: PAR (periodic autoregressive) and PARMA (periodic autoregressive-moving average). PAR ( $p$ ) model fit to each period of the series an AR ( $p$ ) model. In a similar way, a PARMA ( $p, q$ ) consists of an ARMA ( $p, q$ ) model to each period analyzed. In hydrology, PAR ( $p$ ) modeling resulted from Thomas & Fiering (1962, according to Hipel & McLeod (1994)) researches.

According to Rasmussen *et al.* (1996), PAR ( $p$ ) models extrapolation to the PARMA ( $p, q$ ) models is not a trivial task and can be non-justifiable given the good performance of autoregressive models. Yet, as demonstrated in Hosking (1984), the literature describes procedures to hydrological series modeling that present a long dependence, that is, have the  $d$  parameter of ARIMA model (differentiation degree) taking fractional values. These models are known as Arfima, Trevisan, Souza & Souza (2000). The  $d$  estimation is based in general on the periodogram and smoothed periodogram function. These models will not be studied here. As said by Maceira (1989), hydrological series with time intervals shorter than a year, as monthly series, have as a feature the periodic behavior of their probabilistic proprieties, for instance, average, variance, asymmetry and autocorrelation structure. Analyses of this kind of series can be carried out through the use of autoregressive formulations whose parameters present a periodic behavior. Such procedures are called periodic autoregressive models; they are also referred to as PAR ( $p$ ), where  $p$  is the order of the model, that is, the number of model autoregressive terms. Generally,  $p$  is a vector,  $p = [p_1, p_2, \dots, p_{12}]$ , where each element provides the order of each period (month, in case of monthly series).

PAR ( $p$ ) model is mathematically described as follows:

$$\begin{aligned} \left( \frac{Z_t - \mu_m}{\sigma_m} \right) &= \varphi_1^m \left( \frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \varphi_2^m \left( \frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}} \right) + \dots \\ &\dots + \varphi_{p_m}^m \left( \frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \end{aligned} \tag{1}$$

- $Z_t$  = Seasonal series of period  $S$ .
- $S$  = Number of periods ( $S = 12$  to monthly series).
- $T$  = Time index,  $t = 1, 2, \dots, SN$ , function of the  $T$  year ( $T = 1, 2, \dots, N$ ) and the  $m$  period ( $m = 1, 2, \dots, S$ ).
- $N$  = Number of years.
- $\mu_m$  = Seasonal average of the period  $m$ .
- $\sigma_m$  = Seasonal standard deviation of the period  $m$ .
- $\varphi_i^m$  =  $i$ -th autoregressive coefficient of the period  $m$ .
- $p_m$  = Order of the autoregressive operator of the period  $m$ .
- $a_t$  = Series of independent noises with average zero and variance  $\sigma_a^{2(m)}$ . In order to simplify the notation, the noises will be treated only as  $a_t$ .

In the context of the medium range energy operation planning of Brazilian electrical system, the model used is called Newave (developed by CEPEL, the Brazilian Electrical Energy Research Center), in which the planning is represented by a problem of multi-stage stochastic linear programming, whose objective is the total operation costs minimization (fuel costs of the thermal units plus the meet demand penalty) during the planning horizon. According to Marcato (2002), Newave model is based on the optimization technique SDDP (Stochastic Dual Dynamic Programming) and, once an aggregate representation of the hydroelectric park is given, it is possible to consider several interlinked subsystems. It also enables the static or the dynamic representation of the system configuration, proper load discretization, representation of the provision cuts in the electric energy market, and also the access of various affluent natural energy (ANE) scenarios, obtained through periodic autoregressive models. As mentioned by Oliveira (2010), the ANE is calculated from the natural streamflows and the productiveness equivalent to the storage of 65% of the hydroelectric plants reservoirs useful value.

That said, PAR ( $p$ ) model is the one adopted in modeling and simulation of Brazilian electric sector affluences thus an autoregressive model of order  $p$  is fitted to each of the historical series periods (months) of streamflows and/or ANE of each of the four subsystems that enter into the composition of the National Interconnected System (Southeast/Midwest, South, Northeast and North).

This work aims at the investigation of the method of identification of the orders  $p$  of each subsystems series – stage that is being the target of some authors’ critiques, as bad chosen orders can generate unreliable results and inconsistent synthetic scenarios – and the proposition of a new approach to improve on this stage using the intensive computation technique *Bootstrap*.

## 2 THE ORDERS “ $p$ ” IDENTIFICATION

According to Maceira (1989), the traditional identification of orders  $p$  of the PAR ( $p$ ) models is carried out by the analysis of autocorrelation function (ACF) and partial autocorrelation function (PACF).

Taking  $\rho_k^m$  as the correlation between  $Z_t$  and  $Z_{t-k}$ , such that  $t$  corresponds to the period  $m$ :

$$\rho_k^m = E \left[ \left( \frac{Z_t - \mu_m}{\sigma_m} \right) \left( \frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \tag{2}$$

The set of  $\rho_k^m$  autocorrelation functions of the periods  $m = 1, \dots, s$ , describes the series time dependence structure. These functions are given by:

$$\begin{aligned} \rho_k^m = E \left[ \left( \frac{Z_t - \mu_m}{\sigma_m} \right) \left( \frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] &= \varphi_1^m E \left[ \left( \frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \left( \frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + \dots \\ &\dots + \sigma_{p_m}^m E \left[ \left( \frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \left( \frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + E \left[ a_t \left( \frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \end{aligned} \tag{3}$$

Once the parameters of a PAR (p) model are known, the  $\rho_k^m$  functions are given by solving (3) and can be expressed by a combination of exponential decays and/or sinusoidal waves. This way, each  $\rho_k^m$  tends to zero as k increases, Box, Jenkins & Reinsel (1994).

Fixing m and varying k from 1 to  $p_m$  in (3) it is obtained for each period a set of equations usually called Yule-Walker equations. To any period m:

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \rho_2^{m-1} & \dots & \rho_{p_m-1}^{m-1} \\ \rho_1^{m-1} & 1 & \rho_1^{m-2} & \dots & \rho_{p_m-2}^{m-2} \\ \rho_2^{m-1} & \rho_1^{m-2} & 1 & \dots & \rho_{p_m-3}^{m-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{p_m-1}^{m-1} & \rho_{p_m-2}^{m-2} & \rho_{p_m-3}^{m-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1^m \\ \varphi_2^m \\ \varphi_3^m \\ \vdots \\ \varphi_{p_m}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_2^m \\ \rho_3^m \\ \vdots \\ \rho_{p_m}^m \end{bmatrix} \tag{4}$$

Denoting  $\varphi_{kj}^m$  the j-th autoregressive parameter of a process of order k,  $\varphi_{kj}^m$  is the last parameter in this process. The Yule-Walker equations to each period m can be rewritten as follows, Souza & Camargo (2004):

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \rho_2^{m-1} & \dots & \rho_{k-1}^{m-1} \\ \rho_1^{m-1} & 1 & \rho_1^{m-2} & \dots & \rho_{k-2}^{m-2} \\ \rho_2^{m-1} & \rho_1^{m-2} & 1 & \dots & \rho_{k-3}^{m-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{k-1}^{m-1} & \rho_{k-2}^{m-2} & \rho_{k-3}^{m-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1}^m \\ \varphi_{k2}^m \\ \varphi_{k3}^m \\ \vdots \\ \varphi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_{k2}^m \\ \rho_{k3}^m \\ \vdots \\ \rho_{kk}^m \end{bmatrix} \tag{5}$$

Omitting the period m notation in order to facilitate the matters, the set of values  $\varphi_{kk}$ ,  $k = 1, 2, \dots$ , are called *partial autoregressive function* of the period m. This particular set is another way of representing the stochastic process dependence structure during the time. In an autoregressive process of order  $p_m$ , the partial autocorrelation function  $\varphi_{kk}$  will be different from zero for k lower than or equal to  $p_m$  and zero for k higher than  $p_m$ .

Thus, the models “classical” identification consists of determining the most appropriate orders of the autoregressive operators of each period  $p_m, m = 1, \dots, s$ . This can be done by obtaining estimations  $\varphi_{kk}, k = 1, \dots, N/4$  and replacing in (5) the autocorrelations by the respective sample values. If the autoregressive operator order for any period m is  $p_m$ , then  $\varphi_{kk}$  for  $k > p_m$  has a distribution approximately Gaussian with zero mean and variance  $1/N$ , Morettin & Toloj (2006). For a period m, the procedure looks for a largest order i so that  $\hat{\varphi}_{kk}^m$  for  $k > i$  are not significant. This form of identification is the used in Newave and the maximum order accepted is six. It is argued that elevated orders show bigger chances of having negative autoregressive coefficient that can eventually produce undesirable positive coefficients in the *Benders’ cuts* (optimization phase), Maceira & Damázio (2004). It is relevant to stress that the PAR (p) model is applied to historical series modeling and simulation, not with purposes of prediction.

### 3 ASYMPTOTICAL RESULTS TO THE PROBABILITY DISTRIBUTION OF THE ACF AND PACF

The probability distributions of these random variables are extremely complex. In Anderson (1942) *apud* Neto (1991), it is shown that, if the estimated parameter,  $\rho_k$ , is null and the series size is relatively large, then the  $\rho_k$  estimator has Gaussian distribution:  $\rho_k \sim N[0, V(\rho_k)]$ .

Bartlett (1946) *apud* Neto (1991) proposed an approximate expression for the variance of  $\rho_k$ :

$$V(\rho_k) \cong \frac{1}{N} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2) \tag{6}$$

For processes in which  $\rho_k = 0, k > q$ , Bartlett approximation can be used, and is given by:

$$V(\rho_k) \cong \frac{1}{N} \left[ 1 + \sum_{i=1}^q (\rho_i^2) \right] \tag{7}$$

For the sample partial autocorrelations, there is the result given by Quenouille (1948) *apud* Neto (1991), showing that in the hypothesis of an order  $p$  autoregressive process, the approximate  $\varphi_{kk}$  variance is given by:

$$V(\varphi_{kk}) \cong \frac{1}{N} \tag{8}$$

Hence, the procedures for hypothesis testing are:

$$H_{01}: \rho_k = 0 \qquad H_{02}: \varphi_{kk} = 0$$

And, from the results previously shown, it is possible to build  $(1 - \alpha)$  confidence level intervals for the parameters of interest:

$$I_{\rho_k} = \left[ \rho_k \pm Z_{(1-\alpha/2)}\sqrt{V(\rho_k)} \right] \tag{9}$$

$$I_{\varphi_{kk}} = \left[ \varphi_{kk} \pm Z_{(1-\alpha/2)}\sqrt{\frac{1}{N}} \right] \tag{10}$$

Thus, ACF and PACF and the previous intervals are used as identification tools to find the most appropriate model structure, be comparing the behaviors of the estimated and the theoretical function.

However, there are models in the structures whose ACF and PACF values, though not null, are accepted as belonging, respectively, to the confidence intervals. In this case, the traditional technique points to the acceptance of the processes as white noise, when, in fact, they are models with parameters situated in specific regions of the parametric space that have low  $\rho_k$  and  $\varphi_{kk}$  values. Such models are known in the literature as “Quasi-White Noise”, Neto & Souza (1996).

Corroborating the notion that the classical form of orders  $p$  identification can not be the most adequate, taking in account the results previously shown, Stedinger (2001) proposes some reflections about this stage and the following criterion to the selection of the significant lags: for each period  $m$ , it is sought the highest order  $i$ , such that all the estimations  $\varphi_{kk}$  for  $k < i$  are significant. This criterion does not admit non-significant intermediary lags, something that happens in the classical form. The maximum order considered is six.

Stedinger (2001) also states that a modeling procedure that considers each month individually can not produce the best or an adequate set of models. Without further mathematical details and given the kind of natural phenomena involved, the author argues that it makes no sense to consider, for instance, that the February streamflows depend on those of January, December, November, October, September and August ( $p = 6$ ), and that, to predictions referring to January, depends only on the December value ( $p = 1$ ).

Stedinger (2001) concludes by stating that the solution to the modeling challenge faced would be the adoption of a broader perspective of the decisions model that should be guided by the hydrological reasonableness of the selected models that recognize the intrinsic link among the chosen models for different periods.

Therefore, given the critical observations and analyses about the classical form of identification of the orders  $p$ , this study proposes the use of *Bootstrap* as an alternative form for the estimation of the probabilities distribution of the ACF and PACF coefficients, and subsequent analysis of the significance of the lags in order to improve on the orders  $p$  identification.

#### 4 *BOOTSTRAP* IN THE IDENTIFICATION OF ORDERS “ $p$ ”

The term *Bootstrap* has its origin in the expression “lift oneself by pulling his/her *bootstrap*”. What lies behind this expression is the fact that, through the procedure, it is possible to obtain proprieties of big samples from a reduced number of observations.

*Bootstrap*, introduced by Efron & Tibshirani (1993), is a non parametric computationally intensive stochastic technique that enables the assessment of the variability of estimators based on data of only one existent sample.

The technique is recommended to problems in which it is difficult to deploy the conventional statistical procedures. In general, it shows some advantages if used in situations of short or big samples, inasmuch as it provides results close to the ones obtained through usual asymptotic ways or superior to reduced samples.

Operationally, the technique consists of a sampling with replacement of the elements of a random sample, generating a “*Bootstrap* sample”, the same size of the original sample. It is extracted a sufficient number of samples, in order to obtain the “*Bootstrap* distribution” of any statistics of interest to the analysis. This way, the set of *Bootstrap* observations corresponds to an estimation of the real distribution sampling of the statistics at stake. As shown in Efron & Tibshirani (1993), as the sample size tends to the infinite, the *Bootstrap* distribution converges to the real statistics distribution.

To Oliveira (2010), in the time series context, there are basically two ways of use *Bootstrap*: in the residues and the so-called “moving blocks” method. In the former, guaranteeing the hypothesis of residues independence, these are used to generate new “*Bootstrapped*” series. In the case of “moving blocks”, no model is adjusted a priori. From the data, blocks of  $M$  size are built from the original series. Then,  $k$  blocks are selected randomly with replacement, aggregating them to form the *Bootstrap* sample. This step is repeated  $B$  times, generating  $B$  new *Bootstrap* series, Souza (2004).

**4.1 Bootstrap distribution of  $\rho_k$  and  $\varphi_{kk}$**

To the construction of  $\rho_k$  and  $\varphi_{kk}$  *Bootstrap* distribution, it is necessary to have an algorithm that preserves the series autocorrelation structure. From the original series, one obtain  $\rho_k^{*l}$ , the  $\rho_k$  estimation in the  $l$ -th *Bootstrap* replication by sampling, with replacement,  $(n - k)$  pairs of the original sample of pairs and building with them the *Bootstrap* sample. With this generated *Bootstrap* sample, the ACF is estimated by the usual method. Repeating the process  $B$  times, one arrives at the *Bootstrap* estimator of  $\rho_k$ :

$$\rho_k^* = \sum_{l=1}^B \rho_k^{*l} / B \tag{11}$$

The set of observation pairs with lag  $k$  provide information to the estimation of the  $\rho_k$  parameter. The estimation of this autocorrelation, obtained from this set, is an element of the classical estimator sampling distribution. From this set, it is possible to obtain a *Bootstrap* sample, with  $l$  index, and the respective estimation of the  $\rho_k$  parameter. A very large set of  $\rho_k^{*l}$  estimations constitutes an approximation of the  $\rho_k$  sampling distribution. By the Law of Large Numbers, it is possible to state that:

$$\rho_k^* = \left[ \sum_{l=1}^B \rho_k^{*l} / B \right] \rightarrow E[\rho_k] \tag{12}$$

Having the *Bootstrap* distribution of the  $\rho_k$  estimator, it is possible to build the  $\varphi_{kk}$  *Bootstrap* distribution in relation to the *Bootstrap* autocorrelation of lag  $k$  and the previous ones in the usual way.

With the *Bootstrap* distribution of  $\rho_k$  and  $\varphi_{kk}$  obtained as described above, a precision measurement of these statistics can be calculated. The  $\rho_k$  and  $\varphi_{kk}$  *Bootstrap* standard errors are, respectively:

$$s^*(\rho_k) = \sqrt{\frac{\left[ \sum_{l=1}^B (\rho_k^{*l} - \rho_k^*)^2 \right]}{B}} \tag{13}$$

$$s^*(\varphi_{kk}) = \sqrt{\frac{\left[ \sum_{l=1}^B (\varphi_{kk}^{*l} - \varphi_{kk}^*)^2 \right]}{B}} \tag{14}$$

This way, *Bootstrap* confidence intervals can be built without the hypothesis of normality, for both  $\rho_k$  and  $\varphi_{kk}$ , given by (see Oliveira (2010) *apud* Neto (1991) for details):

$$I_{\rho_k} = [\rho_k \pm t_{gl, (1-\alpha)^{S^*}}(\rho_k)] \tag{15}$$

$$I_{\varphi_{kk}} = [\varphi_{kk} \pm t_{gl, (1-\alpha)^{S^*}}(\varphi_{kk})] \tag{16}$$

The next section shows the model used and the results obtained through the use of the technique in the specific case of periodic modeling.

### 5 PROPOSED MODEL AND RESULTS OBTAINED

In this study, the series used are the monthly ANA’s, running from January/1931 to December/2008 available from the Operator of the National Electric System (ONS).

In the model used in this study, to each of the periods of each of the subsystems, the algorithm creates monthly “pairs” with lags ranging from 1 to  $k$ . These “pairs” are sampled with replacement, forming  $B$  *Bootstrap* samples of the same size of the original sample. From each one of the samples formed, the algorithm estimates  $B$  *Bootstrap* ACF coefficients to every period for all subsystems.

Once all the  $\rho_k^m$  were obtained, the following step consists on the estimation of the PACF, whose calculations are based on the equation (4). The following example is for  $k = 1, 2$  and period  $m$ :

$$\varphi_{11}^{m(B_i)} = [1]^{-1} \rho_1^{mB(i)} \tag{17}$$

$$\begin{bmatrix} \varphi_{21}^{mB(i)} \\ \varphi_{22}^{mB(i)} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1^{m-1B(i)} \\ \rho_1^{m-1B(i)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1^{mB(i)} \\ \rho_1^{mB(i)} \end{bmatrix} \tag{18}$$

And this follows to the others lags for each period  $m$ . The index  $B(i)$  represents the  $i$ -th sampled with replacement obtained by the algorithm. The proposed approach here is the following: to each  $i$  *Bootstrap* sampling of the partial autocorrelation function, all the indices that appear on the right hand side of the matrix equation (18) should also refer to the  $i$  sample ACF. This way, it is expected that the  $\varphi_{kk}^{mB(i)}$  estimate sampling distribution capture all the randomness inherent to the procedure.

After running the algorithm, it is obtained  $B$  *Bootstrap* estimates for each one of the  $\rho_k$  and  $\varphi_{kk}$  for all the periods and all four subsystems. Therefore, accurate metrics can be calculated; for instance, the estimated standard errors (equations (13) and (14)). Thus, confidence intervals can be built and the real significance of the ACF and PACF coefficients can be established, following equations (15) and (16).

In order to guarantee the *Bootstrap* estimations convergence (through the Law of Large Numbers), 10.000 simulations were used to each one of the  $\rho_k$  and  $\varphi_{kk}$  for each period. Obviously, a better or worse approximation depends fundamentally on the sample quality. In the



the construction of the confidence intervals as given by (15) and (16), the 5% level of significance was adopted and this was compared to the corresponding ones obtained from the intervals constructed from the 2.5% and 97.5% percentiles. The values obtained did not show any significant differences.

The results obtained for the orders of each period of the Brazilian subsystems are shown next. For both forms of identification, the  $p$  values found in accordance to the established criteria are shown. The objective, therefore, is to compare the results obtained through the two approaches and check on their proposed approach applicability. From Oliveira's (2010), the criteria proposed were:

**Criterion 1:** for a period  $m$ , the procedure looks for the largest order  $i$  so that all  $\hat{\phi}_{kk}^m$  estimations for  $k > i$  are not significant, with maximum order equals to six (following the guide used by Newave).

**Criterion 2:** for a period  $m$ , the procedure looks for the largest order  $i$  so that all  $\hat{\phi}_{kk}^m$  estimations for  $k < i$  are significant. This is the criterion that does not admit non-significant intermediary lags. The maximum order considered is six.

Analyzing the results shown in Tables 1 and 2, it is possible to note that, from the 48 periods considered (12 months of each of the subsystems), in only nine cases the orders pointed by the Criterion 2 of the Classical identification and the *Bootstrap* Criteria 1 and 2 present some differences. For each one of these cases, the confidence interval of the parameters of the possible adjusted model was simulated also through *Bootstrap* procedures.

**Table 1 – Results of the orders identification for the Southeast/Midwest and South subsystems.**

	Southeast/Midwest				South			
	Classical		<i>Bootstrap</i>		Classical		<i>Bootstrap</i>	
	Criterion 1	Criterion 2	Criterion 1	Criterion 2	Criterion 1	Criterion 2	Criterion 1	Criterion 2
Jan	5	1	1	1	1	1	1	1
Feb	6	1	1	1	1	1	1	1
Mar	1	1	1	1	4	1	1	1
Apr	2	2	2	2	5	1	1	1
May	3	1	3	1	1	1	1	1
Jun	1	1	1	1	5	1	1	1
Jul	3	3	3	3	4	2	1	1
Aug	1	1	1	1	1	1	1	1
Sep	1	1	1	1	1	1	1	1
Oct	3	3	1	1	1	1	1	1
Nov	1	1	1	1	6	1	1	1
Dec	4	1	1	1	1	1	1	1

The cases in which the orders pointed by the Criterion 2 of the Classical identification were higher than the ones through *Bootstrap* (October in Southeast/Midwest and North subsystems, and July in South subsystem), the simulations showed that the model parameters that used the Criterion 2 of the Classical identification were not significant, excepts for the first autoregressive parameter, corroborating the order identified by the *Bootstrap* procedure in these periods.

**Table 2 – Results of the orders identification for the Northeast and North subsystems.**

	Northeast				North			
	Classical		Bootstrap		Classical		Bootstrap	
	Criterion 1	Criterion 2	Criterion 1	Criterion 2	Criterion 1	Criterion 2	Criterion 1	Criterion 2
Jan	1	1	1	1	1	1	1	1
Feb	4	2	2	2	4	2	2	2
Mar	1	1	1	1	1	1	1	1
Apr	1	1	1	1	1	1	1	1
May	1	1	3	1	1	1	1	1
Jun	1	1	5	1	1	1	1	1
Jul	1	1	1	1	3	3	3	3
Aug	1	1	1	1	1	1	2	2
Sep	1	1	3	1	5	1	5	2
Oct	3	1	1	1	2	2	1	1
Nov	2	2	2	2	5	1	1	1
Dec	5	1	1	1	1	1	1	1

For the May model in Southeast/Midwest subsystem and May, June and September models in Northeast subsystem, in which the orders pointed by the Criterion 1 of the *Bootstrap* were higher than the ones of the Criterion 2 of the Classical identification, simulations showed that for the 95% confidence level of only the first and the last parameters of the models were significant, *i.e.*, the intermediate parameters are not. In these cases, it is recommended the use of more parsimonious models, resulting in the inferior orders pointed out by the Criterion 2 of the *Bootstrap* identification.

In the case of the August model in North subsystem, for both the Criteria 1 and 2 of the *Bootstrap* identification, the orders selected were higher than the Criteria 2 of the Classical one. The simulations conducted pointed out that, at the 95% confidence level, the two parameters of the model are significant, but, at the 99% confidence level, only the first one would remain in the model.

Finally, to the September model in North subsystem, the simulations showed that, in the model of order 5, only the first, the second and the fifth parameters were significant. Following the parsimony principle, one should opt for the order 2, *i.e.*, the model that was pointed out by the Criterion 2 of the *Bootstrap* identification.

**6 CONCLUSIONS AND FINAL REMARKS**

The use of PAR ( $p$ ) to model hydrological streamflows and/or affluent natural energy series has been applied to the Brazilian integrated energy system, in order to provide possible ways to improve the existing approach, which has been openly criticized.

The use of *Bootstrap* to estimate the autocorrelation and partial autocorrelation functions coefficients lead to the identification of lower orders in most of the periods for all the subsystems, making the models substantially more parsimonious, one of the main recommendations of Box & Jenkins approach (see Box, Jenkins & Reinsel (1994)).

In addition to that, the use of Criterion 2 to orders selection seems to be more accurate, confirming the previous results of Oliveira (2010) and Stedinger (2001), who criticized the classical form of identification and suggested that the solution to the challenge faced in the modeling procedure would be the adoption of a broader perspective of the decisions model that must be guided by the hydrological reasonableness of the models selected. In general, the orders identified by the Classical Criterion 2 and the *Bootstrap* Criteria 1 and 2 were identical, except for the ones whose significance of the parameters models was checked and, by using the parsimony principle, the *Bootstrap* Criterion 2 procedure seems to be the most accurate.

The *Bootstrap* technique was efficient in the identification of the orders of the models of each period (mainly in relation to Criterion 2), approximating the Criterion 2 of the classical approach in 90% of the cases. This way, the alternative recommended as the most accurate is the use of the intensive computation technique during the critical stage of identification of orders  $p$ , as well as, as a starting change, the use of the Criterion 2 replacing the Criterion 1, in current use.

## REFERENCES

- [1] BOX GEP, JENKINS GM & REINSEL GC. 1994. *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, New Jersey.
- [2] EFRON B & TIBSHIRANI RJ. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- [3] HARVEY AC. 1981. *Time Series Models*. Philip Allan. London.
- [4] HIPEL KW & MCLEOD AI. 1994. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier. Amsterdam.
- [5] HOSKING JRM. 1984. Modelling Persistence in Hydrological Time Series Using Fractional Differencing. *Water Resource Research*, **20**: 1898–1908.
- [6] MACEIRA MEP. 1989. Operação Ótima de Reservatórios com Previsão de Afluências. M.Sc. Dissertation, COPPE/UFRJ, Rio de Janeiro, Brasil.
- [7] MACEIRA MEP & DAMÁZIO JM. 2004. The use of PAR ( $p$ ) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the Brazilian hydropower system. In: 8<sup>th</sup> International Conference on Probabilistic Methods Applied to Power Systems, Iowa State University, Ames, Iowa.
- [8] MARCATO ALM. 2002. Representação híbrida de sistemas equivalentes e individualizados para o planejamento da operação de médio prazo de sistemas de potência de grande porte. Ph.D Dissertation, Department of Electrical Engineering, PUC-Rio, Brazil.
- [9] MORETTIN PA & TOLOI CMC. 2006. *Análise de séries Temporais*. Editora Blucher, São Paulo.
- [10] NETO AC. 1991. *Bootstrap em séries temporais*. Ph.D Dissertation, Department of Electrical Engineering, PUC-Rio, Brazil.
- [11] NETO AC & SOUZA RC. 1996. A *Bootstrap* Simulation Study in ARMA ( $p, q$ ) Structures. *Journal of Forecasting*, **15**: 343–353.
- [12] OLIVEIRA FLC. 2010. Nova abordagem para geração de cenários de afluências no planejamento da operação energética de médio prazo. M.Sc. Dissertation, Department of Electrical Engineering, PUC-Rio, Brazil.

- [13] ONS – OPERADOR NACIONAL DO SISTEMA ELÉTRICO. Disponível em <<http://www.ons.com.br/home/>>. Acesso em 2009.
- [14] RASMUSSEN RF, SALAS JD, FAGHERAZZI L, RASSAM JC & BOBEE R. 1996. Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resource Research*, **32**: 3151–3160.
- [15] SALAS JD & OBEYSEKERA JTB. 1982. ARMA model identification of hydrologic time series. *Water Resource Research*, **18**: 1011–1021.
- [16] SOUZA RC & CAMARGO ME. 2004. Análise e Previsão de Séries Temporais: os modelos ARIMA. Sedigraf, Ijuí.
- [17] STEDINGER JR. 2001. Report on the evaluation of CEPEL's PAR models. Cornell University. School of Civil and Environmental Engineering. Ithaca, New York.
- [18] THOMAS HA & FIERING MB. 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: Design of Water Resource Systems edited by A. Maass *et al.*, Harvard Univ. Press, Cambridge, Massachusetts, 459–493.
- [19] TREVISAN ES, SOUZA LR & SOUZA RC. 2000. Estimação do parâmetro “ $d$ ” em modelos ARFIMA. *Pesquisa Operacional*, **20**: 73–82.