

<http://dx.doi.org/10.1590/2318-0331.011715186>

Monthly streamflow forecast for National Interconnected System (NIS) using Periodic Auto-regressive Endogenous Models (PAR) and Exogenous (PARX) with climate information

Previsão de vazões mensais para o Sistema Interligado Nacional (SIN) utilizando modelos Periódicos Autorregressivos Endógenos (PAR) e Exógenos (PARX) com a utilização de informações climáticas

Cleiton da Silva Silveira¹, Alan Michell Barros Alexandre², Francisco de Assis de Souza Filho²,
Francisco das Chagas Vasconcelos Junior³ and Samuellson Lopes Cabral⁴

¹Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, CE, Brazil

²Universidade Federal do Ceará, Fortaleza, CE, Brazil

³Fundação Cearense de Meteorologia e Recursos Hídricos, Fortaleza, CE, Brazil

⁴Centro Nacional de Monitoramento e Alertas de Desastres Naturais, São José dos Campos, SP, Brazil

E-mails: cleitonsilveira@unilab.edu.br (CSS), alan-michell@hotmail.com (AMBA), assis@ufc.br (FASF), juniophy@gmail.com (FCVJ), samuellsoncabral@hotmail.com (SLC)

Received: December 27, 2015 - Revised: July 12, 2016 - Accepted: November 27, 2016

ABSTRACT

This study aims to find a seasonal streamflow forecast model simultaneous to all stations of SIN using periodic autoregressive models with exogenous variables (PARX) using climate indexes. Comparing the results from PAR and PARX Models, this research analyzes the impact on forecasts by using climate information. The proposed models for streamflow forecast has been carried out using natural streamflow data from Operador Nacional do Sistema (ONS) and statistical techniques (such as multiple linear regression and stepwise method to choose explanatory variables). On 27 climate indexes utilized, 4 of them are suggested in this work. The performance analysis methodology is based on the ELECTRE method further the NASH coefficient, the mean absolute percentage error, the multi-criteria distance and correlation. Forecasts with one month lead, the PAR models present better results for most stations of SIN within seasons DJF, MAM, and JJA, while for SON season there is greater efficiency from PARX model. This kind of model shows better performance during dry season in the basins at Northern Brazil – Amazonas and Araguaia-Tocantins; Central-Eastern Brazil – Eastern Atlantic and the most rivers located in the Paraná basin.

Keywords: Monthly streamflow forecast; Climate information; National Interconnected System.

RESUMO

Este estudo propõe um modelo de previsão simultânea de vazões sazonais para todos os locais SIN através de modelos periódicos autorregressivos simples (PAR) e com variáveis exógenas (PARX) utilizando índices climáticos. Os modelos propostos de previsão de afluência utilizam os dados de vazões naturais gerados pelo Operador Nacional do Sistema (ONS) e técnicas estatísticas como as de regressão linear múltipla e o método stepwise para escolha de variáveis explanatórias. São utilizados 27 índices climáticos, dos quais 4 foram sugeridos neste trabalho. A análise de desempenho das metodologias é baseada no método ELECTRE com o uso do coeficiente de NASH, do erro médio percentual absoluto, da distância multicritério e da correlação. Para previsões com um mês de antecedência, os modelos do tipo PAR apresentam melhores desempenhos na maioria dos postos do SIN nos trimestres DJF, MAM e JJA, enquanto para o período SON a uma maior eficiência do modelo PARX. O PARX apresenta melhor desempenho no período seco das bacias do norte do Brasil - Amazonas e Araguaia-Tocantins; centro-leste brasileiro - Atlântico Leste e na maioria dos rios que formam a Bacia do Paraná.

Palavras-chave: Previsão sazonal; Índices climáticos; Setor elétrico.



INTRODUCTION

The Brazilian power grid has a great contribution in the energy from hydroelectric plants. The Brazilian hydroelectricity generation and distribution capacity is directly influenced by the hydrological variability of the streamflows and the interconnection of the hydroelectric power plants, whose forecasts and uncertainties should be considered in the system operation and planning (ALEXANDRE, 2012; ANEEL, 2002).

In general terms, the quality of the hydrological forecasts affects the performance of the system operation. Thus, methodological improvements of hydrological scenarios or inflows generation contribute to the advance of the planning process and the operation schedule of the National Integrated System (NIS).

The monthly streamflow forecasting currently used by the National System Operator (ONS) can be performed through two methods. The first one uses rain-flow models with daily time-step added to the stochastic model of weekly streamflow forecast - PREVIVAZ, which can predict up to six weeks lead time. The generation of streamflow scenarios is obtained from the processing of the GEVAZP model, which usually consider 1,000 scenarios generated from a historical series of natural streamflow with 12 previous months. The second method can be performed more directly by the use of the PREVIVAZM software, which has a calculation structure similar to that of PREVIVAZ with a monthly time-step. The PREVIVAZM model was elaborated as a tool for special studies to verify the conditions of the supply of the power demand. The models utilized in both methods use the Periodic AutoRegressive model (PAR) or the Periodic AutoRegressive model with Moving Averaging (PARMA) (ALEXANDRE, 2012).

Studies performed since the 1990s pointed out to the use of climatic variables could reduce uncertainties due to the high correlation of these variables with hydrological variables (SOUZA FILHO; LALL, 2003; FOLEY et al., 2002; BARROS et al., 2004). On global scale, can be highlighted the works of Dettinger and Diaz (2001), Amarasekera et al. (1997) carried out in the tropical zones of South America and Africa and that of Stuck et al. (2006) and Garcia and Mechoso (2005) in South America. The studies of Carriello et al. (2005), Soares et al. (2006) and Kim and Vissotto (2003) analyzed the relation between Sea Surface Temperature (SST) and the hydrography in Brazil. In terms of specific regional basins, the works of Cardoso and Silva Dias (2006) and Berri et al. (2002), focused on basins in the Brazilian southeast; Foley et al. (2002) and Barros et al. (2004) analyzed basins of the Brazilian northwest showed the good representative capacity from SST variations associated to the streamflow variability in Brazil.

In order to take advantage of the climatic information, statistical models of monthly streamflow forecasting have been proposed in several studies, proving that the use of climatological indexes as explanatory variables in mathematical models provides ability to explain regimes of intrannual inflows. Among those studies, we can highlight the ones by Uvo and Graham (1998), Souza Filho and Lall (2003, 2004), Silva and Molion (2004), Rohn et al. (2003), and Rocha et al. (2007).

The current paper has the objective of developing a simultaneous seasonal flow prediction model for all NIS sites by

Periodic Autoregressive Models with Exogenous Variables (PARX) by using climatic indexes, preserving the spatial structure through the correlation of the residuals.

METHODS

Data

The observational data base used as input for PAR and PARX models are streamflow from ONS since 1931 to 2008. Exogenous variables into PARX models are climate indexes made available by International Research Institute for Climate and Society (IRI). The period from 1931 to 1998 covers calibration time of the models, while the verification period is defined 1999-2008.

National Interconnected System (SIN)

The National Interconnected System (SIN) manages the generation and transmission of electricity in Brazil. The SIN is a hydrothermal system in large scale with prominently hydropower plants, just 1.7% of installed capacity in Brazil are isolated systems located over Amazon region (ONS, 2016).

For planning and management, the ONS adopts streamflow forecasting by using subsets of hydrological output in each basin, denominated base station. At the rest of the sites, the streamflows is predicted through monthly linear regression based on the data from the base stations to complement the forecast for the whole SIN. ONS currently works with a total of 88 Base Stations (BS), which are representative of the various regional hydrological regimes found in Brazil (ONS, 2010). These BS are listed in Figure 1.

Periodic Autoregressive Model (PAR)

The method is based on periodic autoregressive models and for the maintenance of the spatial correlation., Among the BS, the noise correlation (CRD) is utilized and defined at the moment of the calibration of the regression models. The first procedure to establish the PAR model is to choose the explanatory variables by using stepwise technique.

The standardization of the monthly series is the first stage of calculation of most hydrological series analyzes, as follows

$$Z_{t,m} = \frac{(Q_{t,m} - \mu_m)}{\sigma_m} \quad (1)$$

$Z_{t,m}$ represents standardized streamflow in month ($m=1,2,3...12$) for the time t . Since t the number of the months in the time series, then it varies from 1 to $12*N$, where N is the total of years. $Q_{t,m}$ is streamflow at a given BS in period of months t in m^3/s , μ_m is the seasonal streamflow average of m (unit m^3/s), and σ_m is the seasonal streamflow standard deviation in month m in m^3/s .

The PAR model (p) is referred by the index " p " and denominated the order or autoregressive terms from the model. In general, " p " is a vector that establish the order of each element given on a PAR period (p_1, p_2, \dots, p_{12}). The standard formulation for a standardized variable " Z " could be formally described with



Figure 1. Spatial distribution of Base Stations utilized in this analysis.

the following equation $Z_{t,m} = \varnothing_1^m (Z_{t-1,m-1}) + \dots + \varnothing_{p_m}^m (Z_{t-p_m,m-p_m}) + a_t$, where \varnothing is Autoregressive Period operator, p_m is the order of the Autoregressive Period operator “ m ”, and a_t the time series of noise, they are independent with mean zero and variance $\sigma_a^{2(m)}$.

The noises are defined at the moment of the calibration of the models, based on the difference between the expected inflows and the observed ones. The maintenance of the spatial correlation in the prediction of streamflow is performed through the use of the monthly covariance matrix of the noise between the BS to generate the random component of the error following a normal multivariate distribution.

Periodic Autoregressive Exogenous Model with noise correlated (PARX-CRD)

The PARX-CRD model generation strategy is similar to PAR-CRD, in order to consider previous months as predictors of the month to be forecasted and by maintaining the spatial correlations. However, it adds climatic variables as possible predictors, similar to Equation 2

$$Z_{t,m} = \varnothing_1^m (Z_{t-1,m-1}) + \dots + \varnothing_{p_m}^m (Z_{t-p_m,m-p_m}) + \sum_{j=1}^n \varnothing_j^{m-1} \cdot \beta_j^{m-1} + a_t \quad (2)$$

where β_j^{m-1} is the predictor climatic variable j for the month m . Monthly data were obtained from (NOAA, 2014). As criterion for choosing the available indices, all indexes that have some physical relationship to the streamflow variability in South America were used in the correlation analysis. This analysis pointed

to the possibility of using 22 indices added to the SST Index of the SST in South Atlantic at the Confluence of the Brazil-Malvinas (ASBM) defined by Cataldi (2008).

Then the possibility of creating other climate indices was analyzed. First, the predominant hydrological regime of the 88 SB was determined, with a predominant streamflows above the annual average in the first semester and streamflows below the annual average in the second semester. Data from the International Research Institute for Climate and Society (IRI) data were analyzed using data from Sea Surface Temperature (SST) and Zonal Wind (ZW).

Besides those variables, we analyze four time series from spatial mean of zonal wind at 925 hPa in the regions: ZW1 (10S-20S,30W-45W), ZW2 (5S-15S,100W-130W), and ZW3 (15N-5N,150W-180W); and spatial mean of SST on the region: SST1 (5S-20S; 90W-130W). The four regions are shown in the Figure 2.

In order to evaluate the influence of the climatic indexes and their representativeness in streamflow series, 27 indexes were considered and originated from circulation and atmospheric dynamics and oceanic conditions. Table 1 summarizes all the indexes used in this study.

Performance evaluation of the models

The evaluation procedure of this work considers the forecast with one month lead time. Due to the small number of forecasts, they will be grouped in quarterly blocks: DJF, MAM, JJA, and SON.

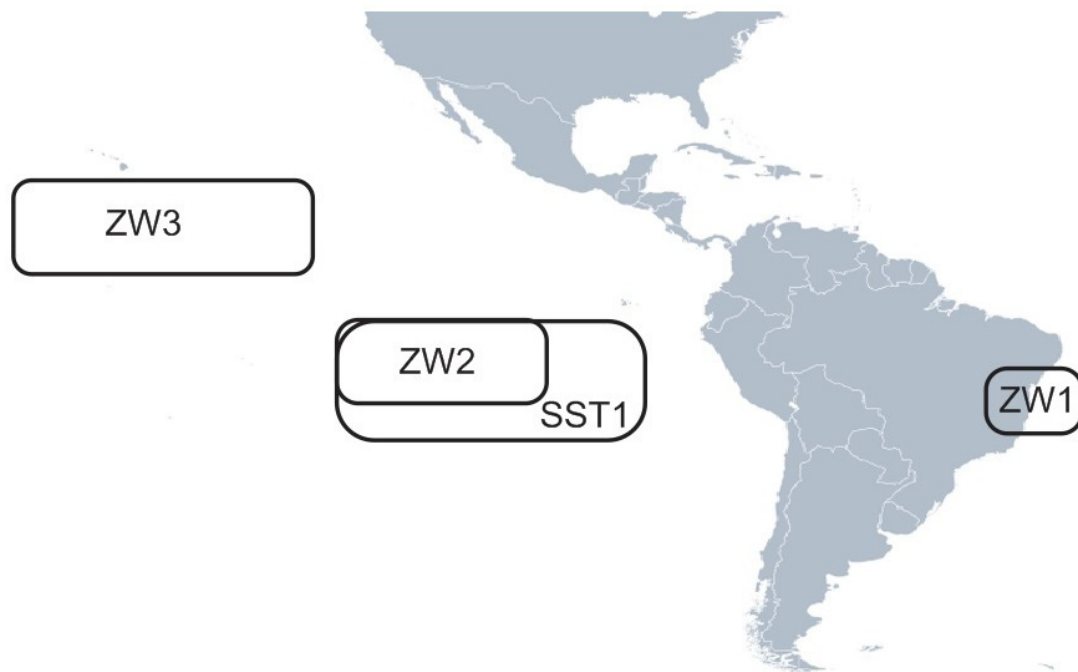


Figure 2. Localization of the regions where suggested climate index were extracted.

Table 1. Climate Indexes as input into PARX model.

Acronym	DESCRIPTION
AMO	Atlantic Multi-decadal Oscillation
ASBM	Atlantic South Confluence Brasil-Malvinas – SST anomaly
BEST	Bivariate ENSO time serie
CAR	Caribbean SST index
MEI	Multivariate ENSO Index
NAO	North Atlantic Oscillation
Nino ₁₊₂	Nino1+2 SST index (0-10S, 90W-80W)
Nino ₃	Nino3 SST index (5N-5S;150W-90W)
Nino _{3,4}	Nino3.4 SST index (5N-5S;170W-120W)
Nino ₄	Nino4 SST index (5N-5S;160E-150W)
NOI	North Oscillation Index
NP	North Pacific Pattern (30N-65N;160E-140W)
NTA	North Tropical Atlantic SST Index (6N-18N;60W-20W) (6N-10N;20W-10W)
ONI	Oceanic NINO Index
PDO	Pacific Decadal Oscillation – SST Anomaly
PNA	Pacific North America Index
PW	Pacific warming 1° EOF os SST (15S-15N; 60E-170E)
QBO	Quase-biannual Oscillation – Equatorial Zonal Wind (30hPa)
SOI	Southern Oscillation Index
TNA	Tropical North Atlantic – SST anomaly (5.5N-23.5N;15W-57.5W)
TNI	NINO Trans Index – NINO
TP	Tropical Pacific 1° EOF of SST (20N-20S;120E-60W)
TSA	Tropical South Atlantic Index – SST anomaly (0-20S;10E-30W)
ZW1	ZW1 Zonal Wind at 925hPa (10S-20S,30W-45W),
ZW2	ZW2 Zonal Wind at 925hPa (5S-15S,100W-130W)
ZW3	ZW3 Zonal Wind at 925hPa (15N-5N,150W-180W)
SST1	SST index (5S-20S,90W-130W)

Source: NOAA, IRI, adapted from Cataldi (2008).

In order to evaluate the performance of estimated (calibration) and predicted (validation) streamflows in relation to the observed streamflows for hydroelectric power plants, underlying metrics applied by ONS (2010) are used. The results are presented in Table 2 in relation to the median of the generated scenarios.

The NASH coefficient expresses the efficiency of making better predictions with the value 1 (one), representing 100% efficiency. Negative values indicate that the performance of the model is worse than the performance of the forecast model considering the value of the monthly historical average only.

The mean absolute error (EMPA) represents the differences between the observed and simulated standardized values in relation to the observed streamflow. Values close to zero indicate near-perfect prediction. By rising the difference term, the EMPA tends to give greater weight to the large discrepancies between the observed and predicted fields.

The correlation (CORREL) can assume values between -1 and 1 therefore perfect anticorrelation and perfect correlation. Moreover, there is also the total absence of correlation verified with a result equal to zero. This index has the capacity to detect phase correspondence between the series. For the calculation of this index, the median of the forecasts was grouped by seasons and by station, then calculation was made in relation to the observed series.

ONS (2010) developed an index called Multicriteria Distance - DM, which uses the NASH and EMPA indicators, according to Table 2.

Electra

One of the tools widely used to aid in decision problems is Multi-criteria or Multi-objective Decision Analysis. This approach suggests that there is usually no alternative that is the best across

Table 2. Index for evaluation of the models.

Indexes	Equacions
Mean absolute percentage error	$EMPA = \frac{1}{nt} \sum_{i=1}^{nt} \frac{ QO_i - QP_i }{QO_i} \times 100$
Nash-Sutcliffe	$NASH = 1 - \frac{\sum_{i=1}^{nt} (QP_i - QO_i)^2}{\sum_{i=1}^{nt} (QO_i - \overline{QO_i})^2}$
Multi-criterion distance	$DM = \sqrt{(1 - NASH)^2 + (EMPA)^2}$

QP_i – predicted streamflow in time interval t ;
 QO_i – observed streamflow in time interval t ;
 nt – number of time interval;
 $EMPA$ – mean absolute percentage error;
 $NASH$ – Nash Sutcliffe coefficient (NASH; SUTCLIFFE, 1970);
 $\overline{QO_i}$ – mean observed streamflow in time interval t ;
 DM – multi-criterion distance.

all criteria. In order to define the best model for the electric sector, the method used will be Electre, as mentioned by Braga and Gobetti (1997). The evaluation criteria NASH, EMPA, CORREL and DM will be exploited for this approach.

Electre I

The central idea of the Electre I method is to separate the total set of alternatives from those that are preferred in most evaluation criteria and do not cause an unacceptable level of dissatisfaction in the other criteria. This method is mainly applied in the treatment of discrete alternatives evaluated qualitatively, for a partial ordering of the alternatives. Two indexes are used that measure the advantage and the disadvantage of each alternative in relation to the others. They are: $C(a,b)$ referring to the concordance index with the affirmative aSb and $D(a,b)$ corresponding to the discordance index with the affirmative aSb . The agreement between two alternatives a and b represents the willingness to choose the alternative a in place of b . The concordance index can be understood as a weighted percentage of the criteria for which alternative a is preferred to alternative b .

The first of those indexes is given by the equation

$$C(a,b) = \frac{\sum(W^+ + 0,5W^-)}{\sum(W^+ + W^- + W^-)} \tag{3}$$

where W^+ is the sum of weights of the criteria for condition a is preferred to b , Wc is the sum of the weights of the criteria for condition $a = b$, and W^- corresponding to the sum of weights of the criteria for the b is the preferred to a .

The concept of disagreement is complementary to that of agreement and represents the discomfort experienced in choosing the alternative to over alternative b , and is determined by:

$$D(a,b) = \max_{K \in I^-} \frac{\left[\frac{\bar{z}(bK) - \bar{z}(aK)}{R^*} \right]}{R^*} \tag{4}$$

where R Z_k^* Z_k^- e Z_{ak} the evaluation of a in relation to criterion k . Z_k^* is the o best degree of evaluation obtained for the criterion k , and Z_k^- the worse degree evaluation obtained for criterion k .

Table 3. Criteria for discordance.

Indexes	D
EMPA	0.01
NASH	0.05
DM	0.05
CORREL	0.02

The relation the preference is defined to establish value of threshold (p, q) , between 0 and 1, such as the alternative a is preferred to alternative b , being verified that aSb if, only if, $\begin{cases} C(a,b) \geq p \\ D(a,b) \leq q \end{cases}$. To this analysis, it was utilized: $p=0.70$ and q varying according to the selected indexes, as Table 3.

Electra II

The procedure of ordering of the method Electra II is built in two stages: progressive classification and regressive classification. The final classification is given through average of classification obtained in each stage. The progressive classification could be described as following. $G_f(t)$ is a subset of G_p , where $G_f(0) = G_p$; and $G_f(t)$ a subset of G_p where $G_f(0) = G_p$ fixes $t=0$. Therefore, chosen the nodes do not dominated in $G_f(t)$. This set is denoted by C . The non-dominated C nodes C are also selected. This set is denoted by $A(t)$. For each $x \in A(t)$ is obtained a classification from $v'(x) = t + 1$. In this way, the sets $G_f(t)$ and $G_f(t)$ are reduced from $G_f(t+1) = G_f(t) - A(t)$ and $G_f(t+1) = G_f(t) - A(t)$.

If $G_f(t+1)$ for empty set, the process stops. Otherwise, $t=t+1$ and the procedure is repeat from step 2. In the regressive classification are reverted the directions of the arrows in G_p and G_f obtained an classification $a(x)$, the same way as $v'(x)$. The rating is reset by doing $v''(x) = I + a_{max} - a(x), \forall x \in X$. The final classification of alternatives, $\bar{v}(x)$ is calculated by defining the function $m(x)$ as $m(x) = \frac{v' + v''}{2}, \forall x \in X$. The values are ordered following $m(x)$.

RESULTS

In the determination of the monthly regressions for the 88 SB under study, standardized streamflow data between 1 and 11 months lead time (lag1 to 11) were tested as explanatory variables. It should be mentioned that the streamflow used in this study were standardized according to Equation 1.

By using the stepwise, we selected the explanatory variables, as presented in Table 4. In the PAR model there is a major influence of the lag1 autoregressive variable, present in most (87.7%) of the SB regressions through all months. The streamflow with lag2 appears in 17.4% of the equations, especially in the months of July and August on the Grande Basin. While for the streamflow with lag3 the frequency is of 12.9% and for the streamflows with lag4 and 5 they have the frequency of 9.1%.

As shown in Table 4, January, March, April, July, and August have the second most commonly used explanatory variables, which are formed by streamflow in two months lead time, lag2. However, for the months of May, September, and October, the variables with the second highest level of explanation are those formed by streamflows with lag3; Variables with lag4 and lag5 appear more

frequently in December and June, respectively. February has the second explanatory variable most used in the regressions in lag11 flows, with 11 months lead time.

The monthly regressions of SB are built for 6 variables. In 44.1% of these regressions only one variable is required to explain the variability of the dependent term on the equation. According to Table 5 for the equation with two and three variables the percentage is 32.7% and 14.6%. It is estimated that 31 equations, 2.9% of the total, do not have variables, so they do not have definite regressions. Therefore, they will always forecast the average at the station.

January and February are the months in which most of the regressions have only one variable, 69% and 58%, respectively. It is verified for September only 15% of the regressions are defined with one variable only and 43% of the regressions require two variables, thus showing an exception to the rule.

In PARX-CRD model, 10 most used variables in the regression equations (according to Table 6) are: streamflow lag1 – 45.9%, streamflow lag2 – 15.2%, streamflow lag3 – 11.8%, PB63₍₋₁₎ – 9.5%, PB70₍₋₁₎ – 9.5%, PB19₍₋₁₎ – 8.6%, streamflow lag6 – 8.5%, streamflow lag5 – 8.1%, streamflow lag4 – 8.0%, and PB49₍₋₁₎ – 7.9%.

Among the numbers of variables in the regression for the different months, it is verified that the month of May is where the largest number of regressions requires a group of six variables, 18.3% of the 12-month regressions, followed by July

with 14.7%; For the group with five variables, the months with the highest percentages are those of June with 12.8%, October and April, both with 10.1%; In the equations with four variables, October and April have 12.5% and 11.9%, respectively; While for the regressions with three variables, the month of February is the one with the highest percentage (16.1%) and October the second highest with 10.6%; In the equations with two variables, March and January have 16.3% and 13.7%, respectively; Regressions with one variable only are observed with higher frequencies in January (20.3%) and November (15.9%). There is an influence on the time series of streamflow in the Atlantic and Pacific SST months represented by the PDO, AMO, NAO, NINO 3 and NINO 1.2 indexes. In the FMA months the TNI variable stands out, as in the season JJA, AMO and TNI are the most representative exogenous variables. In January there is the inclusion of the variable ZW3 in 18 equations, this index is proposed in this paper.

Comparison between PARX-CRD and PAR-CRD

It is observed in Figure 3 that the mean monthly value of the EMPA of the series of estimated values of 30% ranging between 8 and 135%. For the values in the periods of DJF, MAM, JJA, and SON as averages are 34%, 30%, 23%, and 34%, respectively.

For all the periods under analysis, there is an increase of the Mean Absolute Error for SB between 70 and 88, referring to the basins of the Iguaçú River (Paraná basin) and Uruguay, as

Table 4. Frequency of time that variable was utilized during such month for PAR model.

	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
lag(-11)	1	11	7	7	10	4	16	11	9	4	8	4
lag(-10)	2	2	1	11	7	1	8	7	12	3	4	5
lag(-9)	1	2	1	7	5	12	2	8	6	3	0	5
lag(-8)	5	1	1	16	4	3	19	4	3	9	4	8
lag(-7)	8	4	5	3	8	6	6	1	5	6	10	1
lag(-6)	5	4	2	7	6	9	11	2	12	10	12	5
lag(-5)	4	10	1	8	15	30	10	9	1	4	3	1
lag(-4)	6	5	13	7	12	5	9	9	3	5	8	14
lag(-3)	1	2	6	10	32	9	20	6	24	17	5	4
lag(-2)	10	6	33	17	17	12	34	16	14	9	10	6
lag(-1)	63	81	72	80	81	86	82	83	75	61	81	81

Table 5. Number of variables by equation in PAR model.

	Number of variables by equation %						
	0	1	2	3	4	5	6
JAN	9	69	15	6	1	0	0
FEB	2	58	32	8	0	0	0
MAR	2	49	36	10	2	0	0
APR	1	41	28	20	8	1	0
MAY	1	32	27	27	8	3	1
JUNE	0	38	38	14	10	0	1
JULY	0	15	43	27	10	5	0
AUG	1	45	32	18	3	0	0
SEPT	2	38	35	22	3	0	0
OCT	8	49	35	2	6	0	0
NOV	3	44	40	9	3	0	0
DEC	5	52	31	11	1	0	0

Table 6. Explanatory variables utilized in most equation of monthly regression with PARX-CRD for 12 months in each 88 Principal Components.

	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
lag(-11)	1	4	6	8	9	5	14	10	7	4	7	5
lag(-10)	0	0	1	9	7	0	9	6	12	4	4	5
lag(-9)	2	1	1	9	6	11	4	7	2	4	0	8
lag(-8)	2	1	4	8	3	0	15	4	3	7	3	4
lag(-7)	7	3	3	1	5	7	6	1	4	12	9	2
lag(-6)	2	4	3	6	2	8	11	3	11	9	12	7
lag(-5)	4	8	2	7	14	31	13	7	0	6	3	2
lag(-4)	6	2	9	5	10	5	15	9	4	6	6	14
lag(-3)	1	2	10	12	36	12	17	7	20	17	2	4
lag(-2)	6	7	33	17	19	13	34	17	17	11	4	5
lag(-1)	65	81	74	77	73	86	81	78	75	51	79	81
SOI	3	4	0	1	1	0	3	5	2	1	9	1
CAR	2	0	15	0	25	6	5	7	4	3	2	5
NTA	3	2	17	9	0	9	1	1	2	0	1	21
PNA	3	0	12	3	1	0	4	1	2	1	7	1
NAO	4	4	1	5	4	6	0	1	14	2	2	1
MEI	4	3	0	0	3	0	0	1	0	0	3	0
NINO3	1	0	0	2	19	1	6	2	0	0	14	1
NINO1.2	0	9	7	2	5	0	3	9	3	18	2	2
NINO3.4	0	3	1	1	3	0	2	0	0	4	2	1
NINO4	1	1	0	1	1	1	1	2	1	0	1	0
NP	0	3	3	8	3	1	1	7	9	1	11	7
ONI	0	3	1	0	0	2	0	0	0	2	2	0
TNA	0	1	2	6	3	1	4	9	3	1	3	4
TSA	5	4	2	5	2	17	6	12	4	6	5	4
QBO	3	23	0	1	2	8	1	4	2	2	9	7
PDO	2	11	3	1	6	1	1	4	25	1	2	1
BEST	2	8	0	0	1	0	1	1	0	0	5	0
PW	15	6	4	8	10	2	1	5	1	9	2	0
TP	3	2	0	4	2	2	0	0	0	2	3	0
NOI	3	1	1	2	4	5	2	3	2	13	0	4
TNI	9	1	19	12	3	20	3	14	1	14	2	2
AMO	2	0	3	2	8	9	25	5	3	35	10	9
ASBM	1	2	2	0	4	3	7	1	5	1	6	0
ZW1	6	6	1	0	4	2	5	4	0	2	0	1
ZW2	3	2	3	12	3	1	6	2	7	6	6	12
ZW3	18	6	2	0	9	6	2	11	19	2	5	4
SST1	0	1	5	0	4	2	16	3	0	1	6	4

EMPA for regions of 53% for the forecasting series, DJF (44%), MAM (45%), JJA (63%), and SON (59%).

During MAM months, the forecast capacity increases subtly to a larger area belonging to the Paraná basin, except for its northwest region. The regions along Amazonas, Tocantins-Araguaia, Southeast Atlantic and North Atlantic Rivers remain with very low predictability, with Nash below zero in some of these stations.

In JJA the change is quite drastic; however, a very disparity remains in the different regions according to the NASH coefficients (Figure 4). Thus, good results are observed in the Central-Eastern region of the river basin Paraná, Southeast Atlantic and northern Uruguay Basins. While northwestern and northern of the Paraná basin, the Amazon region, and Araguaia-Tocantins remain with NASH values below zero.

In the period of SON the predictability begins to fall again. The spatial distribution of the NASH coefficients remains the same as in the JJA period, but the predictive capacity along these SB decreases considerably.

The highest NASH values can be observed in PB63 - Santo Antônio (287), PB61 - Curuá Una (277), and PB62 - Samuel (279) - Amazonas Basin; E PB59 - Tucuruí (275) - Tocantins-Araguaia Basin. (Paraná Basin) and PB74 - Foz do Areia (74) - Paraná Basin (PB) - Paraná Basin (PB) - Paraná Basin (PB) - Iguaçú Basin (Paraná Basin).

In Figures 3 and 4 are shown low predictability from regression, with NASH lower than 0.5 in most base stations. The exception is found for stations PB27 - Promissão (240), PB28 - Nova Avanhandava (242), and PB29 - Três Irmão (243), located in Tietê river basin (Paraná Basin).

Monthly streamflow forecast for National Interconnected System (NIS) using Periodic Auto-regressive Endogenous Models (PAR) and Exogenous (PARX) with climate information

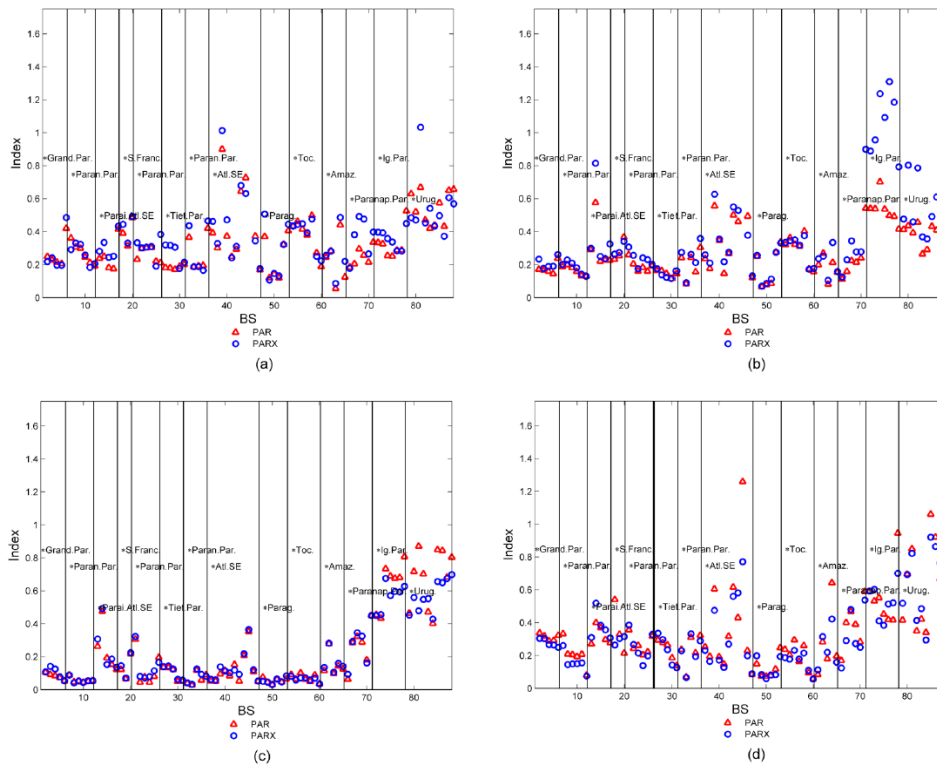


Figure 3. EMPA values by season (a) DFJ, (b) MAM, (c) JJA, and (d) SON for PAR and PARX models. X-axis indicates number of base stations and Y-axis indicates the indexes values. Grande Paran (Grand.Par.), Paranaiba-Paran (Paran.Par.), Paraba do Sul-Atlntico SE (Parai. Atlant.SE), So Francisco (S.Franc.), Tiet-Paran (Tiet.Par.), Atlntico Sudeste (Atlant.SE), Paraguai (Parag.), Tocantins (Toc.), Amazonas (Amaz.), Paranapanema-Paran (Paranap.Par.), Iguau-Paran (Ig.Par.) and Uruguai (Urug.).

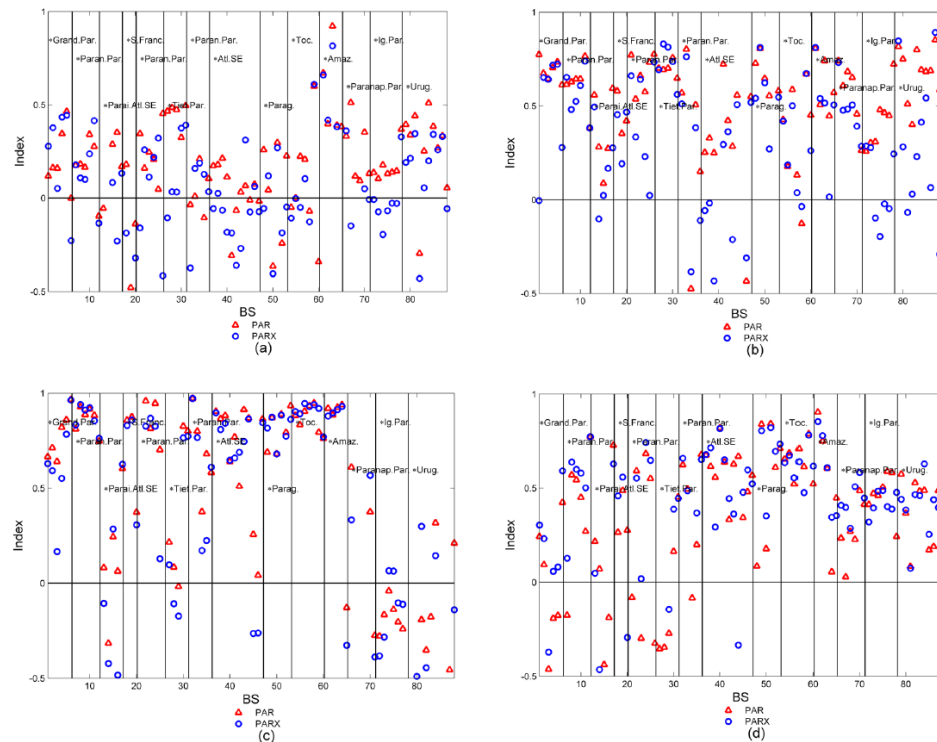


Figure 4. NASH values by season (a) DFJ, (b) MAM, (c) JJA, and (d) SON for PAR and PARX models. X-axis indicates number of base stations and Y-axis indicates the indexes values. Grande Paran (Grand.Par.), Paranaiba-Paran (Paran.Par.), Paraba do Sul-Atlntico SE (Parai. Atlant.SE), So Francisco (S.Franc.), Tiet-Paran (Tiet.Par.), Atlntico Sudeste (Atlant.SE), Paraguai (Parag.), Tocantins (Toc.), Amazonas (Amaz.), Paranapanema-Paran (Paranap.Par.), Iguau-Paran (Ig.Par.) and Uruguai (Urug.).

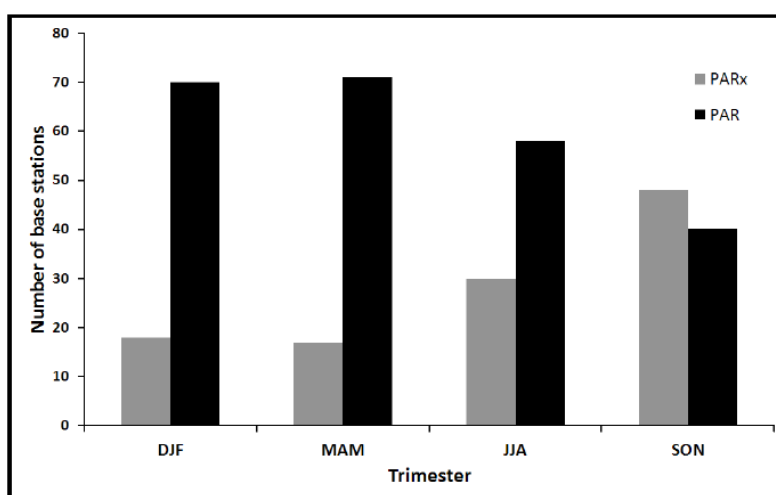


Figure 5. Comparison between PAR and PARX using the Electre method, the y-axis indicates the score obtained by each technique when considering all the base stations. In those stations where Electre did not identify the best methodology 0.5 score were recorded for PAR and PARX.

For the MAM season the prediction capacity improves for a larger area of the Paraná Basin, except for the northwest region of the basin. The regions of the Amazonas, Tocantins-Araguaia, Southeast Atlantic and North Atlantic Rivers remain very low predictability.

Figure 5 shows that the PAR model is in the last quarters DJF, MAM and JJA, while PARX is better in the SON quarter.

CONCLUSION

The analysis of the prediction models for 1 lead month showed a better performance in general of PAR models than the PARX ones at the periods of December, January and February (DJF); March, April and May (MAM) and June, July, and August (JJA), based on the Electre II multi-criteria test. For the period of September, October and November (SON) the PARX model presents better predictions than the PAR in most of the stations.

Studies pointed out to the improvement of the performance of streamflow forecast models by the insertion of explanatory variables from climatic indexes (CATALDI, 2008; SOUZA FILHO; LALL, 2003). However, it was observed that the process described is not confirmed for all regions and all months of the year. This is probably due to the strong signal of the persistence in the PAR model, forecasts with more antecedence may have different results than those shown in this paper. Nevertheless, the PARX model shows good performance (based on the Electra II multicriteria test) for the dry period of the basins in the north of Brazil, Amazonas and Araguaia-Tocantins; and in the central east of Brazil in the Eastern Atlantic and most of the rivers in the Paraná Basin.

Considering the performance of the climatic forcing in the dynamics of the precipitation processes, the information added shown by the PARX for a month lead time in the months of SON suggests that the investigation of climatic indexes can be an alternative for the reduction of the uncertainties associated to the forecast for the SIN. It should be highlighted that the SON is the transition period between the dry and rainy season on most

parts of Brazil, so the ability of models based on persistence may not be enough to capture this onset, requiring information from the large scale characteristics of the atmosphere-ocean and their impacts on the hydrological cycle.

REFERENCES

- ALEXANDRE, A. M. B. *Previsão de vazões mensais para o sistema interligado nacional utilizando informações climáticas* 2012. 293 f. Tese (Doutorado em Engenharia Civil) - Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza, 2012.
- AMARASEKERA, K. N.; LEE, R. F.; WILLIAMS, E. R.; ELTAHIR, E. A. B. ENSO and natural variability in the flow of tropical rivers. *Journal of Hydrology*, v. 200, n. 1-4, p. 24-39, 1997. [http://dx.doi.org/10.1016/S0022-1694\(96\)03340-9](http://dx.doi.org/10.1016/S0022-1694(96)03340-9).
- ANEEL - AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. *Atlas de Energia Elétrica do Brasil*. Brasília: ANEEL, 2002. 153 p.
- BARROS, R. C.; ROLIM, P. A. M.; ALMEIDA, R. M. B.; MELO, A. P. Influência do ENOS na variabilidade da vazão, na bacia do alto rio Negro. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 13., 2004, Fortaleza. *Anais...* Fortaleza: Sociedade Brasileira de Meteorologia, 2004. Available from: <<http://www.cbmet.com/index.php>>. Access on: 30 mar. 2012.
- BERRI, G. J.; GHIETTO, M. A.; GARCÍA, N. O. The Influence of ENSO in the Flows of the Upper Paraná River of South America over the Past 100 Years. *Journal of Hydrometeorology*, v. 3, n. 1, p. 57-65, 2002. [http://dx.doi.org/10.1175/1525-7541\(2002\)003<0057:TI OEIT>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2002)003<0057:TI OEIT>2.0.CO;2).
- BRAGA, B.; GOBETTI, L. Análise multiobjetivo. In: PORTO, R. L. (Ed.). *Técnicas quantitativas para o gerenciamento de recursos hídricos*. Porto Alegre: Ed. UFRGS/ABRH, 1997. p. 361-420 .

- CARDOSO, A. O.; SILVA DIAS, P. L. The relationship between ENSO and Paraná River flow. *Advances in Geosciences*, v. 6, p. 189-193, 2006. <http://dx.doi.org/10.5194/adgeo-6-189-2006>.
- CARRIELLO, F.; SOARES, J. V.; FERREIRA, N. J. A resposta hidrológica das sub-bacias brasileiras e sua relação com o fenômeno ENSO-El Niño/Oscilação Sul. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia. *Anais...* Goiânia: Instituto Nacional de Pesquisas Espaciais -INPE, 2005. p. 2479-2486. Available from: <<http://urlib.net/ltid.inpe.br/sbsr/2004/11.19.19.09>>. Access on: 30 mar. 2012.
- CATALDI, M. *Estudo Numérico da Influência das Anomalias da TSM do Atlântico Sul Extratropical e do Pacífico Equatorial no Regime Hidrometeorológico das Regiões Sul e Sudeste do Brasil*. 2008. 255 f. Tese (Doutorado em Engenharia Civil) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- DETTINGER, M. D.; DIAZ, H. F. Global characteristics of streamflow seasonality and variability. *Journal of Hydrometeorology*, v. 1, n. 4, p. 289-310, 2000. [http://dx.doi.org/10.1175/1525-7541\(2000\)001<0289:GCOSFS>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2).
- FOLEY, J. A.; BOTTA, A.; COE, M. T.; COSTA, M. H. El Niño–Southern oscillation and the climate, ecosystems and rivers of Amazonia. *Global Biogeochemical Cycles*, v. 16, n. 4, p. 1132, 2002. <http://dx.doi.org/10.1029/2002GB001872>.
- GARCIA, N. O.; MECHOSO, C. R. Variability in the discharge of South American rivers and in climate. *Hydrological Sciences Journal*, v. 50, n. 3, p. 459-478, 2005. <http://dx.doi.org/10.1623/hysj.50.3.459.65030>.
- KIM, I. S.; VISSOTTO, S. Relações entre variações das vazões nas bacias hidrográficas do Brasil e os fenômenos El Niño e La Niña. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 15., 2003, Curitiba. *Anais...* Curitiba: ABRH, 2003. p. 1-5.
- NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models, part I – a discussion of principles. *Journal of Hydrology*, v. 10, n. 3, p. 282-290, 1970. [http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6).
- NOAA - NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION. *Climate Indices: monthly atmospheric and ocean time series*. Washington, 2014. Available from: <<http://www.esrl.noaa.gov/psd/data/climateindices/list>>. Access on: 22 mar. 2014.
- ONS - OPERADOR NACIONAL DO SISTEMA ELÉTRICO. *Operação do Sistema Interligado Nacional: Relatório Anual de Avaliação das Previsões de Vazões – 2010*. Brasília: ONS, 2010. 230 p.
- ONS - OPERADOR NACIONAL DO SISTEMA ELÉTRICO. *O que é o SIN - Sistema Interligado Nacional*. Rio de Janeiro: ONS, 2016. Available from: <http://www.ons.org.br/conheca_sistema/o_que_e_sin.aspx>. Access on: 25 apr. 2016.
- ROCHA, E. J. P.; ROLIM, P. A. M.; SANTOS, D. M. Modelo estatístico hidroclimático para previsão de níveis em Altamira-PA. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 17., 2007, São Paulo. *Anais...* São Paulo: Associação Brasileira de Recursos Hídricos, 2007. p. 1-8.
- ROHN, M. C.; KAVISKI, E.; CUNHA, L. M. Estimativa de variáveis hidrológicas a partir de indicadores do fenômeno EL NIÑO. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 15., 2003, Curitiba. *Anais...* Curitiba: ABRH, 2003. p. 1-8.
- SILVA, D. F.; MOLION, L. C. B. Influência da variabilidade climática interanual na hidrologia da Bacia do Rio São Francisco. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 13., 2004, Fortaleza. *Anais...* Fortaleza: Sociedade Brasileira de Meteorologia, 2004. p. 1-5.
- SOARES, J. V.; CARRIELLO, F.; FERREIRA, N. J.; RENNÓ, C. D. Mapping the hydrologic response of the Brazilian hydrologic regions and their variability associated with El Niño and La Niña. *Revista Ambiente & Água*, v. 1, n. 1, p. 21-36, 2006. <http://dx.doi.org/10.4136/ambi-agua.3>.
- SOUZA FILHO, F. A.; LALL, U. Modelo de previsão de vazões sazonais e interanuais. *Revista Brasileira de Recursos Hídricos*, v. 9, n. 2, p. 61-74, 2004.
- SOUZA FILHO, F. A.; LALL, U. Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: applications of a multivariate, semiparametric algorithm. *Water Resources Research*, v. 39, n. 11, p. 1307, 2003. <http://dx.doi.org/10.1029/2002WR001373>.
- STUCK, J.; GÜNTNER, A.; MERZ, B. ENSO impact on simulated South American hydro-climatology. *Advances in Geosciences*, v. 6, p. 227-236, 2006. <http://dx.doi.org/10.5194/adgeo-6-227-2006>.
- UVO, C. B.; GRAHAM, N. E. Seasonal runoff forecast for northern South America: a statistical model. *Water Resources Research*, v. 34, n. 12, p. 3515-3524, 1998. <http://dx.doi.org/10.1029/98WR02854>.

Authors contributions

Cleiton da Silva Silveira: Analysis of the results and technical review of the manuscript.

Alan Michell Barros Alexandre: Manuscript structure, literature review, methods and analysis of the results.

Francisco de Assis de Souza Filho: Study orientation, literature review, methods, analysis of the results and technical review of the manuscript.

Francisco das Chagas Vasconcelos Junior: Manuscript structure, literature review, methods and analysis of the results.

Samuellson Lopes Cabral: Technical review of the manuscript.