La Houille Blanche, n° 3, 2016, p. 53-64 DOI 10.1051/lhb/2016031

Water level forecasting and navigability conditions of the Tapajós River - Amazon – Brazil

Nelio Moura DE FIGUEIREDO[1] & Claudio José CAVALCANTE BLANCO[2]

[1.] *Faculty of Naval Engineering (FENAV/ITEC/UFPA), Technological Institute, Federal University of Pará, Belém, PA, Brazil, nelio@ufpa.br, neliomfigueiredo@outlook.com*
[2.] *Faculty of Sanitary and Environmental Engineering (FAESA/ITEC/UFPA), Technological Institute, Federal University of Pará, Belém, PA, Brazil, blanco@ufpa.br, blancocjc@outlook.com*

ABSTRACT. – In the Amazon, hydrological modelling is of great importance to the safety and operation of inland waterway transport. In this paper, a stochastic model named ARIMA (Autoregressive Integrated Moving Average) was used to forecast water levels and seaworthiness 24 months in advance in the Tapajós River. Additionally, a time series analysis aimed to model and predicting water levels in the Tapajós river basin was completed. During validation, the model had an average coefficient of Nash and Sutcliffe R^2 of 0.95 and an average RMSE of 0.06. Drought forecast for the model predicted depth gains of 30 cm and a load capacity gain of 2,800 tons per trip. In the analysis of the two seasonal periods, the model predicted annual savings in freight cost of U.S. $350,000.00. The performance of the model allows us to conclude that ARIMA models should be used to forecast water levels and navigation conditions in the Amazon.

Key-words: Hydrologic modelling, water levels forecasting, navigability forecasting, ARIMA model, transportation planning, Amazon.

Prévisions du niveau d'eau et les conditions de navigabilité de la rivière Tapajós - Amazonie – Brésil

RÉSUMÉ. – En Amazonie, la modélisation hydrologique a une grande importance pour la sécurité et pour l'exploitation du transport par voie navigable. Dans cet article, le modèle stochastique nommé ARIMA (autorégressif à moyennes mobiles intégré) a été utilisé pour prédire les niveaux d'eau et la navigabilité 24 mois à l'avance pour la rivière Tapajós. En outre, une analyse des séries chronologiques visant à modéliser et prédire les niveaux d'eau pour le bassin de la rivière Tapajós a été achevée. Lors de la validation, le modèle a un coefficient moyen de Nash et Sutcliffe de l'ordre de 0,95 et un RMSE moyenne de l'ordre de 0,06. La sécheresse simulée par le modèle a prédit des gains de profondeur de 30 cm et d'une capacité de 2 800 tonnes par voyage de charge. Dans l'analyse des deux saisons, le modèle a prédit des économies annuelles de coût de fret de 350 000 $ US. La performance du modèle nous permet de conclure que les modèles ARIMA peuvent être utilisés pour prévoir les niveaux d'eau et les conditions de navigation dans l'Amazonie.

Mots-clés : Modélisation hydrologique, prévision des niveaux d'eau, prévision de navigabilité, Modèle ARIMA, la planification des transports, Amazonie.

I. INTRODUCTION

Fluvial navigation is the most important means of passenger and freight transport in the Amazon, linking communities and poles of production, as commercialisation and consumption are established along its vast waterways. Antagonistically, its economic dynamics, their operational peculiarities and the quantitative and qualitative information of activity are not well known and rarely systematised. By virtue of regional conditions, the vast Amazon River Basin is predominantly used as a means of access. Knowing the behaviour of the water levels and consequently knowing the depths of the Amazonian rivers is essential for satisfactory and safe navigation.

Waterway transport efficiency is a function of the navigability conditions and the appropriate sizing of the vessels that use the waterway. These vessels must fit the conditions and hydrological variables that directly influence waterway transport safety and the amount of load. In this context, hydrological stochastic modelling assumes increasingly strategic importance in predicting water levels and predicting navigability conditions. According to Babcock and Lu [2002], several studies on demand and transport simulation have been conducted. However, only a very small number of these studies refer to waterway transport. In this area, Jack Faucett and associates [1997, 2000] performed a study of river transport forecasts for the upper Mississippi River basin, focusing on the assessment of the benefits and costs of improvements in the navigation system. Additionally, Tang [2001] developed a quarterly forecasting ARIMA model for the transportation of grain in the Arkansas River.

Works such as Ledolter [1976, 1977] introduced the time-series analysis in hydrology. Salas *et al.* [1980] and Salas [1992] introduced key concepts and theories related to the modelling of time series in hydrology, with applications focused on engineering. According to Nourani *et al.* [2013], auto regressive integrated moving average (ARIMA) and seasonal ARIMA are widely used to predict hydrological time series. As a modelling technique, ARIMA has been useful in predicting hydro meteorological parameters [Boochabun *et al.*, 2004, Chattopadhyay and Chattopadhyay, 2010 and Chattopadhyay *et al.*, 2011]. Pektas and Keren Cigizoglu [2013] used the ARIMA model to develop a more accurate and generalisable model to predict the monthly direct runoff coefficient time series. Narayanan *et al.* [2013] used the ARIMA to forecast rainfall in western India for the period of 2010–2030. Lohani *et al.* [2012] used auto regressive models and the fuzzy inference system to predict monthly flows. Wu and Chau [2010] used ARIMA models and neural networks for streamflow forecasting in different rivers basins in China. Birinci and Akay [2010] used ARIMA type models to predict rainfall, which in turn, are inputs into the artificial neural network models used to forecast daily streamflow. Koutroumanidis *et al.* [2009] developed a forecasting flow model for the basin of the Nestos River in Bulgaria using a hybrid proposal based on the ARIMA models.

In this context, the objective of this work was to forecasting of the average monthly water levels and navigability conditions of the Tapajós River (Amazon Basin), up to twenty-four months in advance. The developed analysis considered the seasonal stochastic ARIMA models (Auto Regressive Integrated Moving Average).

II. STUDY AREA

The ARIMA model was applied to the Tapajós River basin, which contributes to the Amazon River basin, the largest watershed on the globe, which covers a total area of 6.11 million km², from its headwaters in the Peruvian Andes to its mouth in the Atlantic Ocean. This continental basin extends over several countries in South America: Brazil (63%), Peru (17%), Bolivia (11%), Colombia (5.8%), Ecuador (2.2%), Venezuela (0.7%) and Guyana (0.2%).

The basin of the Tapajós River is situated in the states of Pará, Mato Grosso and a small portion of the Amazonas state between latitudes 2° and 15° south and 53° and 61° west; the basin has a drainage area of 493,200 km² with an elongated longitudinal configuration. Its feeder rivers are the Juruena and Teles Pires, with drainage areas of 191,100 km² and 139,000 km², respectively (Fig. 1). The altitude varies from 900 m at the headwaters of the basin to approximately 50 m above sea level, together with the mouth in the Amazon River. The location of the Tapajós River basin covers the western Pará State, near the Amazonas state border. The Tapajós River, with a length of 851 km, is the most important influencer of the right margin of the Amazon River. The Tapajós River is born at the confluence of the Teles Pires and Juruena rivers, near the border of the Pará, Amazonas and Mato Grosso states, corresponding to the middle portion of the Amazon basin. In terms of climate, the basin is dominated by the Intertropical Convergence Zone (ITCZ). It has low pressure, which is characteristic of regions close to the equator. The average rainfall of the basin is between 1,800 and 2,300 mm per year. Fig. 1 shows the location of the study area.

The major soil types occurring in the Tapajós River basin are lithic, quartz sands, oxisols, podzolic, lowland soils, concretionary soils and cambisols. Approximately 75% of the land is covered by forest, 10% by pasture, whereas 7% of the land has little coverage, 10% is used for agricultural cultivation and approximately 1% is covered by water courses [Collischonn, 2006].

III. UTILIZED DATA

The data used in the model was obtained from the database of the National Water Agency (ANA), which contains historical series of daily average water levels. These data are available at http://hidroweb.ana.gov.br/.

The modelling of the average water level was developed using data from 7 water level gauging stations located in the Tapajós river basin. The identifications and locations of the water level gauging stations used in the modelling are shown in Fig. 1. Historical series containing 33 years of observation data, from the period of 1976-2008, were used. Table 1 summarises the data and information from the water level gauging stations that were used in the

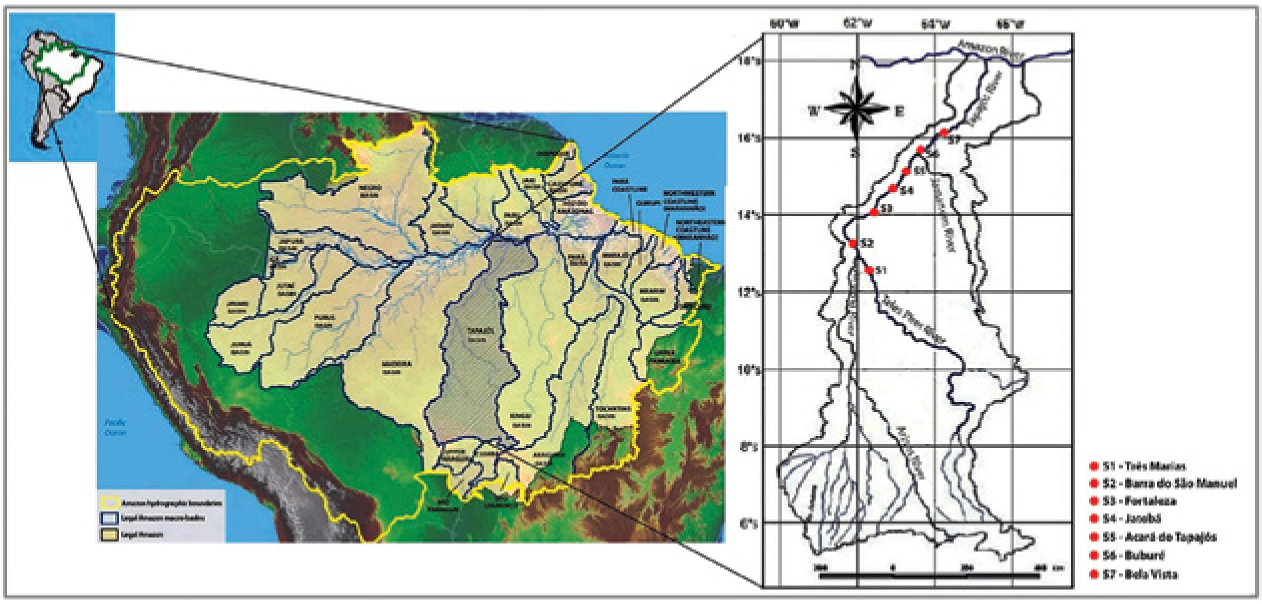

Figure 1: *Location of the Tapajós River basin and water level gauging stations.*

Table 1: Water level gauging stations.

Code ANA	Name	Area (km²)	River
17420000	Três Marias	138,000	Teles Pires
17430000	BSM	333,000	Tapajós
17500000	Fortaleza	363,000	Tapajós
17650000	Jatobá	387,000	Tapajós
17650002	Acará	390,000	Tapajós
17710000	Buburé	450,000	Tapajós
17720000	Bela Vista	453,000	Tapajós

modelling, namely: Três Marias, Barra do São Manuel (BSM), Fortaleza, Jatobá, Acará do Tapajós, Buburé and Bela Vista.

After filling in the gaps and data inconsistencies, the behaviour of the series of water levels in the period of 1976-2008 was analysed using boxplot and monthly variation graphs. As an example, Fig. 2 shows the behaviour of the series of water levels for the Barra do São Manuel water level gauging station. In all stations, the variation range of the seasonal component of the water level series is uniform. The flood period occurs between the months of January and May, the flow period occurs between June and December and the dry season occurs between August and October. The boxplot graph shows the occurrences of a few outliers and demonstrates that a considerable amount of gaps and inconsistencies in these data series were corrected.

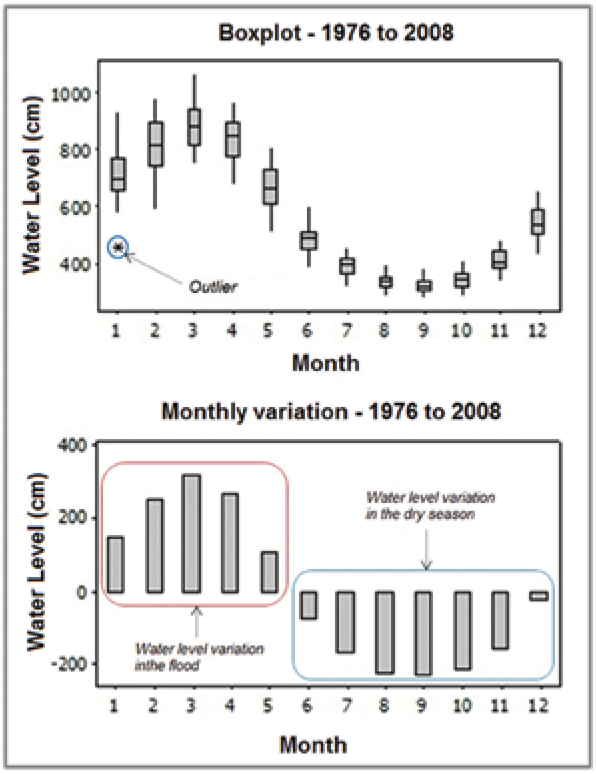

Figure 2: *Behaviour of the water levels series at the Barra do São Manuel station.*

IV. METHODOLOGY

Inland waterway transporting is recognised as one of the most attractive options to facilitate transport growth in Amazon. The safety, efficiency and reliability of inland waterway transport is directly or indirectly related to different meteorological variables. These variables change as climate changes - in the past and in the future. As result of climate change, inland waterway transport may experience problems related to higher volatilities in the water levels. Low water levels imply restrictions on the cargo capacity of barges convoy. These suggests that the capacity of the inland waterway transport fleet is severely reduced in periods with low water levels, which has economic consequences.

Hydrological processes' understanding is of great importance in the water resources management and hydraulic designs. The Hydrological modelling was use as a tool for observance of the behaviour of hydrological variables. River systems are dynamic and stochastic by nature. This paper presents an analysis of predictions of monthly average water levels and navigability conditions, in advance of 24 months, for the Tapajós river, in Brazil, using stochastic models such as ARIMA, for computing the difference between water levels observed and predicted, that will show gains or losses of depths that will be used in the navigability conditions analysis.

The methodology of stochastic processes is based on dynamic mechanisms that provide a structured analysis of a sequence of observations. Thus, a stochastic process should be viewed as a model that describes the probabilistic structure of an observation sequence.

Box and Jenkins [1976] conceptualised time series as a set of observations of a variable arranged sequentially in time. In this concept, it is assumed that there is a stochastic process generating the series and each possible random realisation of the variable is associated with a probability of the observation occurrence. The methodology of ARIMA models is built on the assumption that the series was generated by a stochastic process that can be described and characterised, based on the past behaviour of the variable.

The process of construction and adjustment proposed by the Box & Jenkins model is grounded in an iterative cycle. Fig. 3 summarises the flow of the Box-Jenkins methodology,

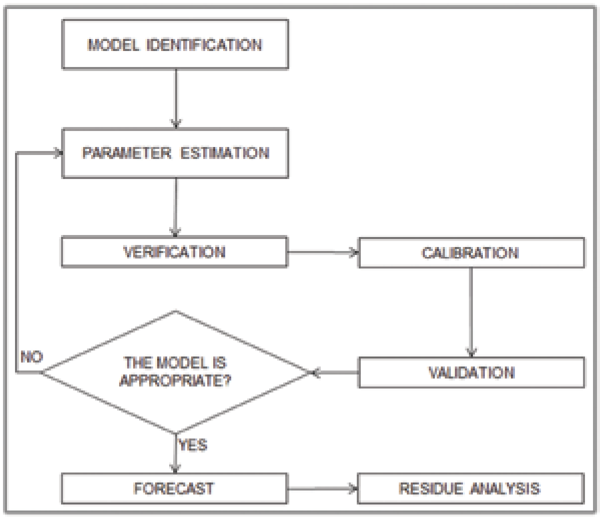

Figure 3: *Flowchart of the Box and Jenkins methodology.*

showing the inherent steps in the application and utilisation process of this model [Box and Jenkins, 1976].

Box and Jenkins [1976] suggested the application of a seasonal ARIMA model to describe the possessor series of serial correlation in seasonally lagged periods. This model is represented by ARIMA (p,d,q) $(P,D,Q)_S$. Where s define the number of months for the seasonal period that will repeat itself again for it season; p, d and q are the auto regressive, integration and moving average non-seasonal parameters and P, D and Q are the auto regressive, integration and moving average seasonal parameters, respectively, according to Equation 1.

$$\phi(B) . \Phi(B^s) . (1-B^s)^D . (1-B)^d . Y_t = \theta(B) . \Theta(B^s) . a_t \quad (1)$$

Where B is the translation operator, ϕ is the non-seasonal autoregressive coefficient, Φ is the seasonal auto regressive coefficient, θ is the non-seasonal average moving coefficient, Θ is the mobile average seasonal coefficient, $\phi(B)$ is the translation operator autoregressive non-seasonal of order "p", $\Phi(B^S)$ is the translation operator autoregressive seasonal of order "p", $\Phi(B^S)$ is the translational operator non-seasonal moving average of order "q" and $\Theta(B^S)$ is the operator of translation seasonal average of order "Q".

The operationalization of the Box and Jenkins methodology was performed using algorithms implemented in the R software (version 2.14.1) of the R Foundation for Statistical Computing, which is a language and environment for statistical computing and graph construction and a GNU project (General Public License of the Free Software Foundation).

The forecasting of observations using time series models are procedures that simulate future observations using models that have been calibrated, validated and adjusted with past and present values. The water level forecast was made using the general characteristic equation of an ARIMA (p,d,q) $(P,D,Q)_S$ seasonal (Equation 1).

Thus, for the ARIMA model $(1,0,0)(1,1,1)_{12}$, Equation 1, with the replacement and operation of the parameter values s=12 (s=12 is the number of months for the seasonal period); p=1, d=0, q=0 and P=1, D=1, Q=1, can be written in its linear form as Equation 2.

$$\begin{aligned}\widehat{Y}_t &= \Phi_1 Y_{t-12} + Y_{t-12} - \Phi_1 Y_{t-24} + \phi_1 Y_{t-1} - \phi_1 \Phi_1 Y_{t-13} \\ &\quad -\phi_1 Y_{t-13} + \phi_1 \Phi_1 Y_{t-25} - \Theta_1 a_{t-12}\end{aligned} \quad (2)$$

Where $\widehat{Y}_t$ is the predicted water level (Na) at time t, Y_t is the observation at time t, and a_t is the residue of the forecast at time t (difference between the predicted and observed water level in a at time t).

The identification stage of the model was performed with the observance of the function graphs of the ACF autocorrelation function and PACF partial autocorrelation function of the series of water levels. The model estimation was made using the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criteria). These criteria are based on the maximum likelihood methodology and are expressed by Equations 3 and 4, respectively.

$$AIC = -2\log\hat{L} + 2.(p+q) \quad (3)$$

$$BIC = -2\log\hat{L} + 2.(p+q)\log n \quad (4)$$

Where n is the number of observations; p and q are the auto regressive and moving average non-seasonal parameters the ARIMA model and $\bar{Y}$ is the maximised likelihood. The best model presents the lowest AIC and/or BIC.

The calibration and validation of the model were developed by utilising the objective function that adopted the coefficient of determination R² (Nash and Sutcliffe coefficient) and the RMSE (root mean square error) as performance coefficients, described by Equations 5 and 6, respectively.

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(Y_t - \widehat{Y}_t\right)^2}{\sum_{t=1}^{n}\left(Y_t - \bar{Y}\right)^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(\frac{Y_t - \widehat{Y}_t}{Y_t}\right)^2}{n}} \quad (6)$$

Where n is the number of observations, Y_t is the observation at time t, $\widehat{Y}_t$ is the simulated observation and $\bar{Y}$ is the average of n observations.

The analysis of seaworthiness is function of the water depth (P or Pr) and of the barge convoy draft (C or Cr). The water depth depends of the riverbed level (Cf) and of the predicted water level (Na or $\widehat{Y}_t$). The amount of transported cargo is related with the dimensions of the barges and with the barge convoy draft, thus, for a same barge, the highest cargo capacity can be reached with a biggest barge draft. The navigability condition, for a given stretch of the waterway, are checked for the most unfavourable situation. This unfavourable condition is a function of the bed level of the river and of a reference level or reduction (Nr), which is define by probability distributions. The navigability conditions and the load capacity of the barge convoy, always, are defined from a reference level (Nr), which varies for a given hydrological scenario.

The verification of variation of water level and of navigability condition were controlled using a horizontal datum, called reduction level (Nr). All the predicted water levels (Na or $\widehat{Y}_t$) and the predicted water depths (P) were referenced in relation to the reduction level. The reduction level is related to a low and critical streamflow, for a same number of days per year, anywhere along a river. E.g., for a reduction level of 95% of the year, means that, on an average, the predicted water level will be lower than the reduction level at only 18 days (5% of the year) per year, or that, the navigation can be paralyzed or will have their cargo capacity reduced in just 18 days a year.

In this paper, for the navigability conditions forecasting, the predicted water levels (Na or $\widehat{Y}_t$) were obtained by ARIMA models. For a navigation channel cross-section, the variations of the reference water depths (Pr = Nr - Cf) were obtained by of the difference between the reference level (Nr) and the riverbed level (Cf). Similarly, the variations of the predicted water levels (P = $\widehat{Y}_t$ – Cf) were obtained by of the difference between the predicted water levels (Na or $\widehat{Y}_t$) and the riverbed level (Cf). Thus, the Fig. 4 shows two situations: In the first situation (left), the predicted water level is above of the reduction level (high or intermediate waters). In the second situation (right), the predicted water level is below of the reduction level (low water or drought).

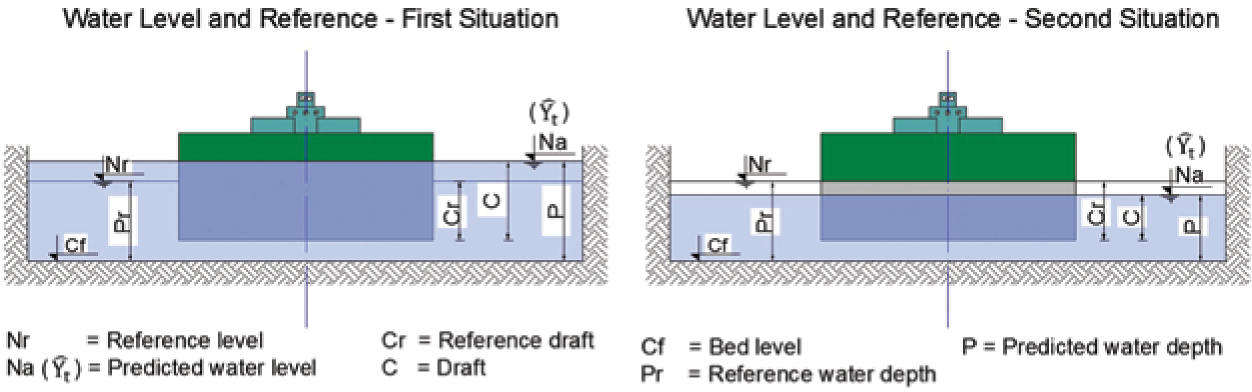

Figure 4: *Cross section of a hypothetical canal with water levels above and below the reference level.*

Gains or losses analyses at the cargo capacity of barge convoys is function of convoy's type and draft and of the water depth. The position of the predicted water level (obtained by ARIMA model) and the convoy draft, for a same convoy, directly influence on the price of freight. Smallest freight rates can be obtained with biggest barge draft. In the economic analyses and of maximum allowable cargo capacity was adopted a standard barge convoy used in Amazonian rivers, composed of six barges with dimensions of 61.00 m by 10.67 m and maximum draft of 4.50 m, as shown in Fig. 5.

V. RESULTS AND DISCUSSION

V.1. Model calibration

In the calibration phase, the model estimations understood the standardisation criteria of Equations 3 and 4. Among the various combinations of parameters assigned to the model ARIMA $(p,d,q)(P,D,Q)_S$, the parameters that showed the lowest values for these criteria were: ARIMA $(0,0,0)(1,1,1)_{12}$ and ARIMA $(1,0,0)(1,1,1)_{12}$. Other parameter settings showed similar values, but the principle of parsimony prevailed. Table 2 shows the values of AIC and BIC (Equations 3 and 4) measured in the water levels series. The ARIMA model $(1,0,0)(1,1,1)_{12}$ showed the best results.

Table 2: Criteria of the AIC and BIC for the water level series in the calibration phase.

Gauged Station	ARIMA Models	
	(0,0,0) (1,1,1)	(1,0,0) (1,1,1)
	AIC	AIC
Três Marias	3,821.04	3,647.32
Jatobá	3,700.79	3,640.47
Fortaleza	3,532.10	3,354.59
Buburé	3,466.26	3,269.62
BSM	3,767.57	3,588.15
Bela Vista	3,954.53	3,751.36
Acará	3,697.60	3,512.64
	BIC	BIC
Três Marias	3,832.49	3,666.39
Jatobá	3,712.24	3,659.54
Fortaleza	3,543.56	3,373.66
Buburé	3,477.71	3,286.34
BSM	3,779.02	3,607.22
Bela Vista	3,965.98	3,769.03
Acará	3,709.05	3,530.52

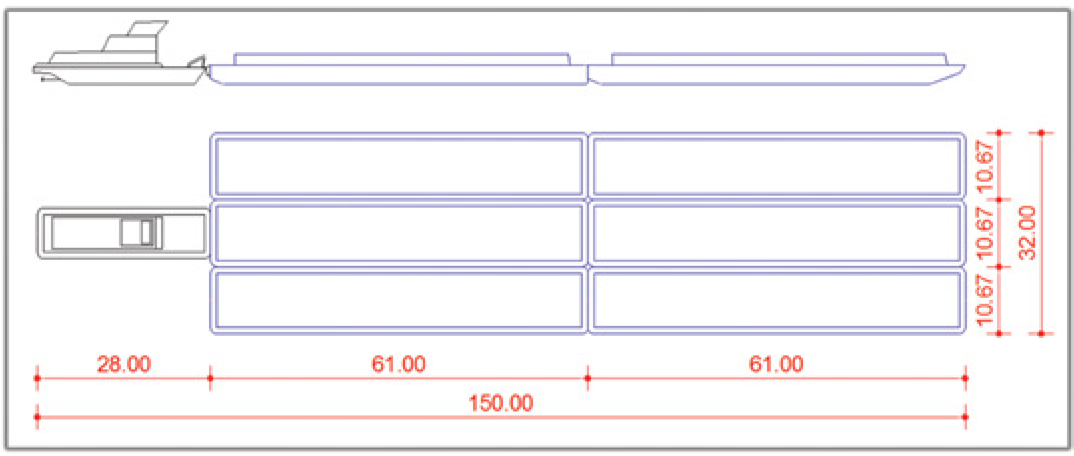

Figure 5: *Barge convoy default for the Amazonian Rivers.*

The performance of the calibration models was measured by the coefficient of Nash and Sutcliffe R^2 (Equation 5) and the relative root mean square error (RMSE) (Equation 6). Table 3 shows the values of R^2 and RMSE for the water level series for each model analysed.

Based on the values of R^2 and RMSE in the model calibration phase, it was found that the ARIMA $(1,0,0)(1,1,1)_{12}$ model performed better. This finding occurred by comparing the water level graphs of the simulated and observed levels. The data series for the period of 1976-2004 formed the model memory. Calibration was performed for the period of 2005-2006, and validation was performed for the period of 2007-2008. For example, Fig. 6 and Fig. 7 show the water level graphs with high and low precision, respectively, for the calibration phase, using the minimum water level simulations as the main criterion of performance measurement.

Table 3: R^2 and RMSE for the series of water levels in the calibration phase.

Gauged Station	ARIMA Models	
	(0,0,0) (1,1,1)	(1,0,0) (1,1,1)
	R^2	**R^2**
Três Marias	0.941	0.946
Jatobá	0.929	0.932
Fortaleza	0.922	0.923
Buburé	0.917	0.923
BSM	0.944	0.951
Bela Vista	0.909	0.917
Acará	0.921	0.928
	RMSE	**RMSE**
Três Marias	0.076	0.070
Jatobá	0.065	0.063
Fortaleza	0.078	0.075
Buburé	0.062	0.059
BSM	0.092	0.084
Bela Vista	0.155	0.137
Acará	0.090	0.084

In dry seasons, the model showed very satisfactory performance in the water level simulation and did not show significant behaviour differences between the seasons. In the floods of the first seasonal period, a water level delay was observed in relation to the simulated water level, and, in the second flood, a water level advance was observed. In the simulation of the maximum water levels, the model did not perform as well as during the dry season because the adjustment of the model parameters was directed to droughts, as the simulation of the minimum water level is the predominant factor when forecasting seaworthiness.

The series analysis and topology of the forecasting equation of the model (Equation 2) show that the existing gaps in the maximum peak levels of the simulations are because the model adopts wastes related to the observed values, until the lag at month 12. For the lag at month 13, the model gauges their wastes using the simulated values as reference, not observed values. The lags of the second seasonal period, presented in a localised form during the calibration period, come from the linear fit of the model to the punctual inflection of the flood in draft along with the lag at month 14. On the other hand, ARIMA model types tend to respond late in the hydrograph rises because they depend only on information that occurs in a specific location. The error of the model is higher in the rising water level graphs because the forecasts in the initial lags are always slightly delayed compared to the observations.

V.2. Model validation

The model validation is based on the same calibration criteria. Table 4 shows the values of the AIC and BIC criteria (Equations 3 and 4) measured in the water level series. The ARIMA model $(1,0,0)(1,1,1)_{12}$ showed the best performance.

In the validation phase, the performance of the models was measured as the calibration performance, with the adoption of the coefficient of Nash and Sutcliffe R^2 and relative root mean square error (RMSE). Table 5 shows the values of the coefficients for each model analysed.

The values of R^2 and RMSE (objective functions - Table 5) show that the ARIMA model $(1,0,0)(1,1,1)_{12}$ performed better in this phase. In fact, the ARIMA model

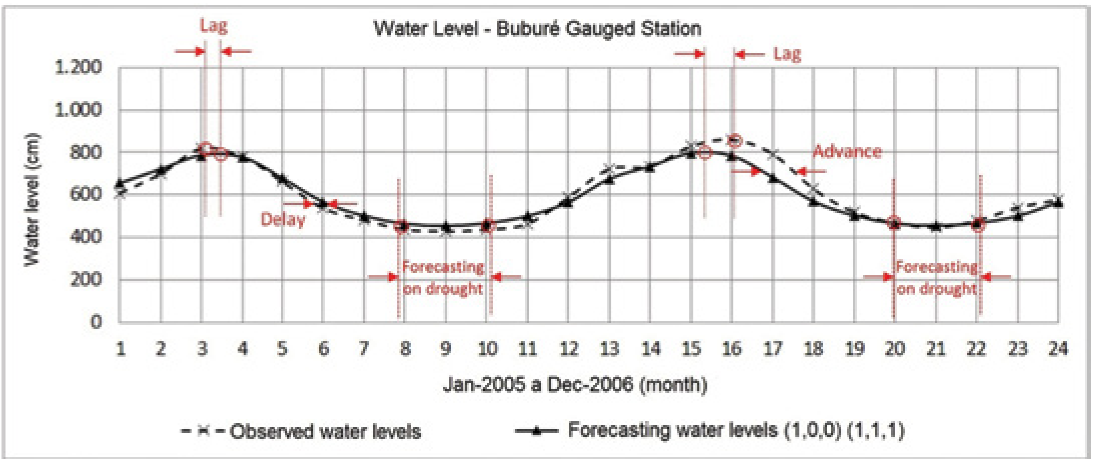

Figure 6: *Water levels observed and simulated at the Buburé station during the calibration phase.*

 DOI 10.1051/lhb/2016031

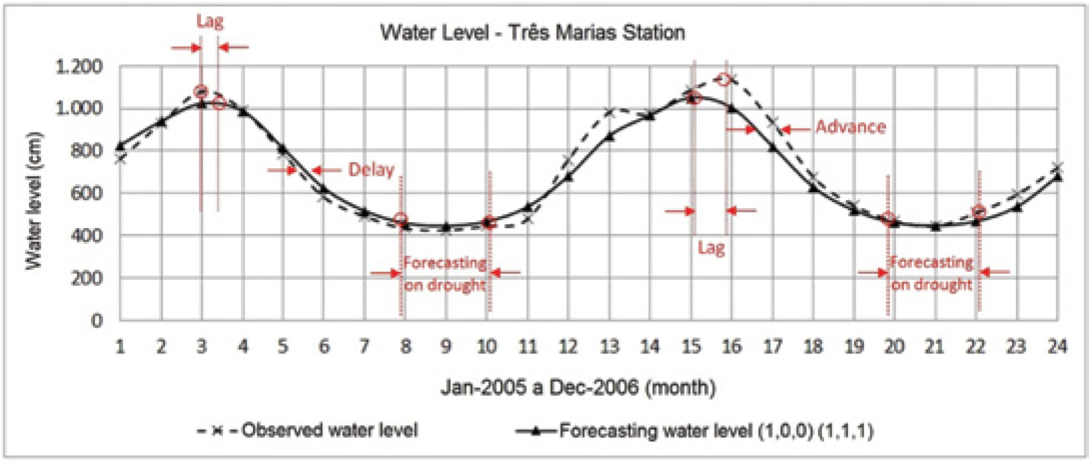

Figure 7: *Water levels observed and simulated at the Três Marias station during the calibration phase.*

Table 4: Criteria of the AIC and BIC for the water level series in the validation phase.

Gauged Station	ARIMA Model
	(1,0,0) (1,1,1)
	AIC
Três Marias	3,880.79
Jatobá	3,873.61
Fortaleza	3,567.18
Buburé	3,476.09
BSM	3,817.51
Bela Vista	3,992.43
Acará	3,736.48
	BIC
Três Marias	3,900.2
Jatobá	3,893.01
Fortaleza	3,586.61
Buburé	3,495.51
BSM	3,836.93
Bela Vista	4,011.85
Acará	3,755.91

Table 5: R² and RMSE for the series of water levels in the validation phase.

Gauged Station	ARIMA Model
	(1,0,0) (1,1,1)
	R²
Três Marias	0.945
Jatobá	0.952
Fortaleza	0.933
Buburé	0.957
BSM	0.941
Bela Vista	0.952
Acará	0.94
	RMSE
Três Marias	0.068
Jatobá	0.052
Fortaleza	0.065
Buburé	0.044
BSM	0.083
Bela Vista	0.099
Acará	0.078

used to forecast water levels presented good results, with R^2 values equal to 0.95 and 0.93 and RMSE values equal to 0.06 and 0.08 for the calibration and validation phases, respectively. The graphical standardisation of the model was developed by comparing the water level graphs of the simulated and observed levels. The data series for the period of 1976-2004 formed the model memory. To demonstrate the performance of the model in the validation phase, by way of example, Fig. 8 presents the water level graph that demonstrates higher accuracy in the validation phase.

In the validation phase, it was observed that the model showed satisfactory performance in simulating the minimum water level in the two seasonal periods. However, the model did not show significant behavioural differences between these phases. Concerning the floods in the first seasonal period, a delay of the observed water levels was noticed in relation to the simulated levels. However, advanced forecasting of the floods of the second seasonal period was observed. The water level graphs in the validation phase showed that the model performed better in the simulation of the minimum levels of water than that of the maximum levels, once the adjustments of the model parameters, such as the calibration, was directed to droughts. From the viewpoint of the maximum water levels, in the first seasonal period during the validation phase, a lag of the simulated flood was observed in relation to the observed flood. In the second seasonal period, a lag was not observed, demonstrating the better performance of the model during the second

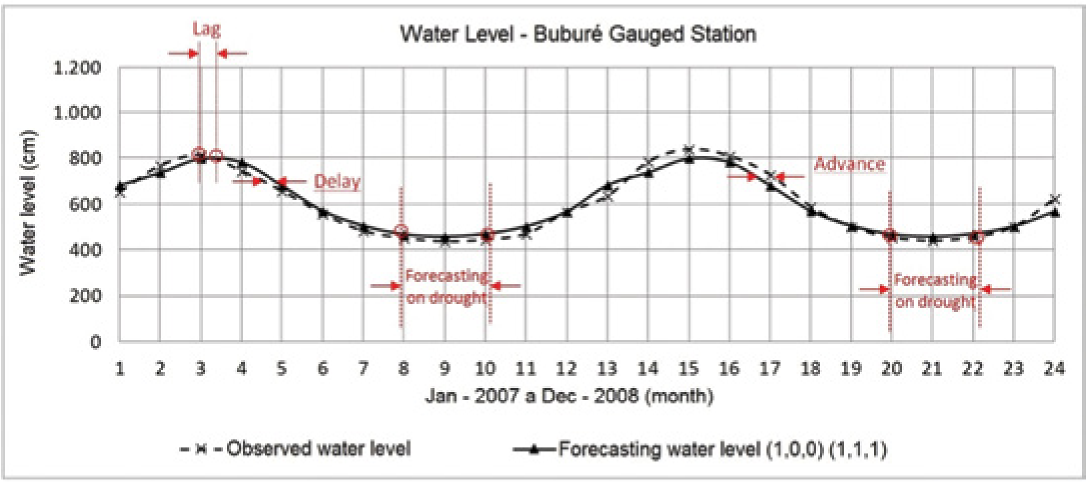

Figure 8: *Water level observed and simulated at the Buburé station during the validation phase.*

seasonal period of the validation phase in relation to that same period in the calibration phase. This slight superiority of the model performance in the validation phase, with regard to the calibration phase, can be confirmed by the values of R^2 and RMSE (Table 6). Similar to the calibration phase, in the validation phase, the water level simulation of drought in the second seasonal period showed better performance and greater linearity. This is because the model assumes the resulting wastes of the observed values until the lag at month 12; after the lag at month 13, the model assesses its waste using simulated water levels.

V.3. Forecasting and verification of the model

The forecasts of water levels were made based on Equation 2. Table 6 shows the seasonal and non-seasonal auto regressive coefficients and moving averages obtained from the application of the ARIMA $(1,0,0)(1,1,1)_{12}$ model to the water level series at each station. Thus, the equations for forecasting water levels $\widehat{Y_t}$ for each station were obtained by replacing the coefficients in Table 6 in Equation 2.

Verification of the models was carried out with significance limits of 95% got the residual graphs of ACF and PACF. It was found that the correlation coefficients were between the 95% significance limits, showing that the model satisfactorily incorporated the series characteristics of water level forecasting.

Table 6: Forecasting equation coefficients for the water levels.

Station	Coefficients ARIMA (0,0,0)(1,1,1) Water Levels		
	ϕ_1	Φ_1	Θ_1
Três Marias	0.6364	-0.0189	0.9438
Jatobá	0.4038	0.0407	0.9673
Fortaleza	0.6510	-0.0275	0.9458
Buburé	0.6750	0.0234	0.9531
BSM	0.6537	-0.0157	0.9477
Bela Vista	0.6868	0.6868	0.9538
Acará	0.6559	0.0230	0.9477

V.4. Navigability forecasting by the ARIMA model

For example, Fig. 9 shows the water level forecast for the Barra de São Manuel station for a period of 24 months and its behaviour in relation to a relative Nr with a return period of 10 years. It is observed that the predictions of Na in the dry season are below the Nr (same condition as Second Situation in Fig. 4). For the remaining months, the forecasts are above the Nr (same condition as Frist Situation in Fig. 4). Thus, the drought would have to interrupt navigation or adopt a draft C lower than the reference draft Cr because the water depth on the Na can be smaller than the water depth of the reference way. For periods outside of the drought, there would not be an interruption of navigation, and draft C could be greater than the draft reference Cr because the depth of Na would be greater than the reference depth to which the route would be designed.

The examination of navigability conditions of a waterway and freight transportation planning in periods of drought require a detailed analysis of the monthly behaviour of the Na simulations against the Nr, so that the permanence of the forecasts can be evaluated in relation to a projected Nr. Thus, the statistical analysis of the forecasts and their probability distributions are of great importance when planning river transport.

Fig. 10 to Fig. 12 shows the histograms of the monthly water level forecasts for Barra do São Manuel (in red) for the period of drought in the months of August, September and October; the curves of the normal probability distribution functions (in blue) and the cumulative upward curves of probability and confidence intervals are also shown.

In waterway studies, especially those aimed at scaling craft and seasonal capacity planning of freight transportation, it is necessary to implement studies on monthly levels. The monthly diagnosis of buoyancy and the permanence of expected water levels, in relation to a given reference level, are indispensable. Based on the reference levels of 306.0 cm, 292.7 cm and 307.3 cm (Fig. 10 to Fig. 12) for water levels that did not exceed 10% of the dry year for a period of recurrence of 10 years, it was observed that the model predictions of the water levels for the months of

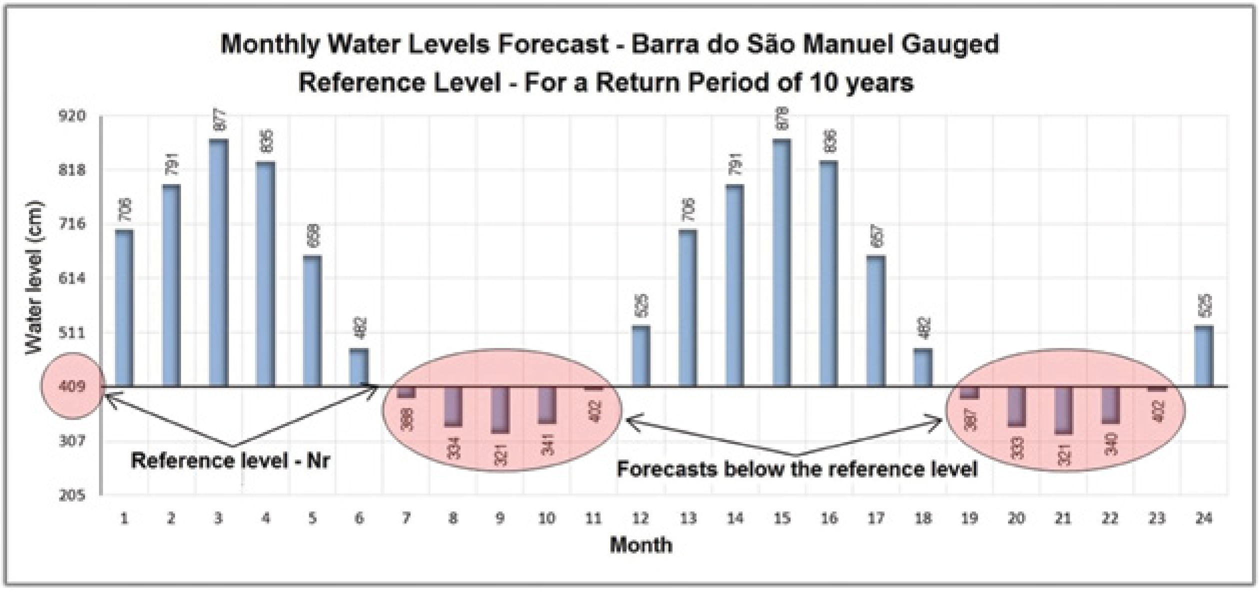

Figure 9: *Water level forecast and its variation in relation to the level of reduction.*

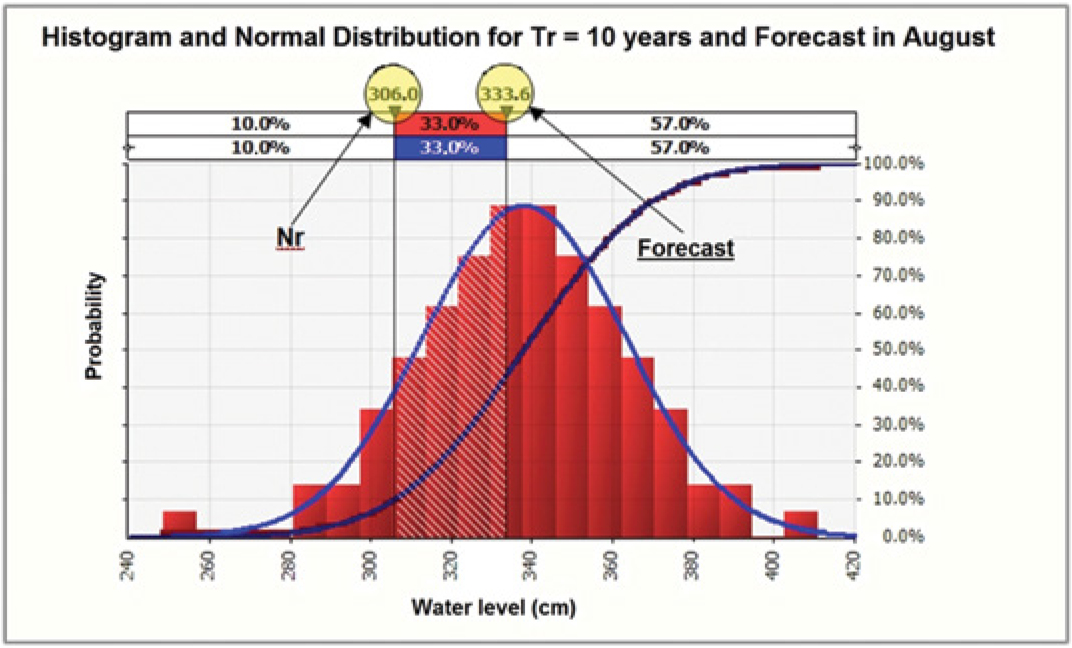

Figure 10: *Permanence of the water level forecast in relation to a reduction level for the month of August at the Barra do São Manuel station.*

August, September and October were 333.6 cm, 321.0 cm and 340.5 cm, respectively, which are above the reference level. Thus, it was observed that the model projected a gain in water depth of 27.6 cm, 28.3 cm and 33.2 cm compared to the reference levels of design. Percentage-wise, it appears that there is a probability of 33.0%, 32.0% and 32.0% that the water levels are above the levels of reduction.

The simulations showed the effects of drought on a barge convoy used in Amazon rivers (Fig. 5) composed of six (06) barges (61.00 m by 10.67 m); the gain of water depth generates an increase in the capacity of cargo per trip of 868 ton, of 890 ton and of 1,044 ton for the months of August, September and October, respectively. For example, considering a river cargo transport modelling with five trips per month, the gain in cargo capacity is 140,000 tons. Thus, due the transported surplus cargo, an economy was obtained at the freight rates of about U.S. $2.5 per ton and an annual freight savings of U.S. $350,000.00. Table 7 shows the gains of depth and of cargo capacity as allowed by the predictions of ARIMA model.

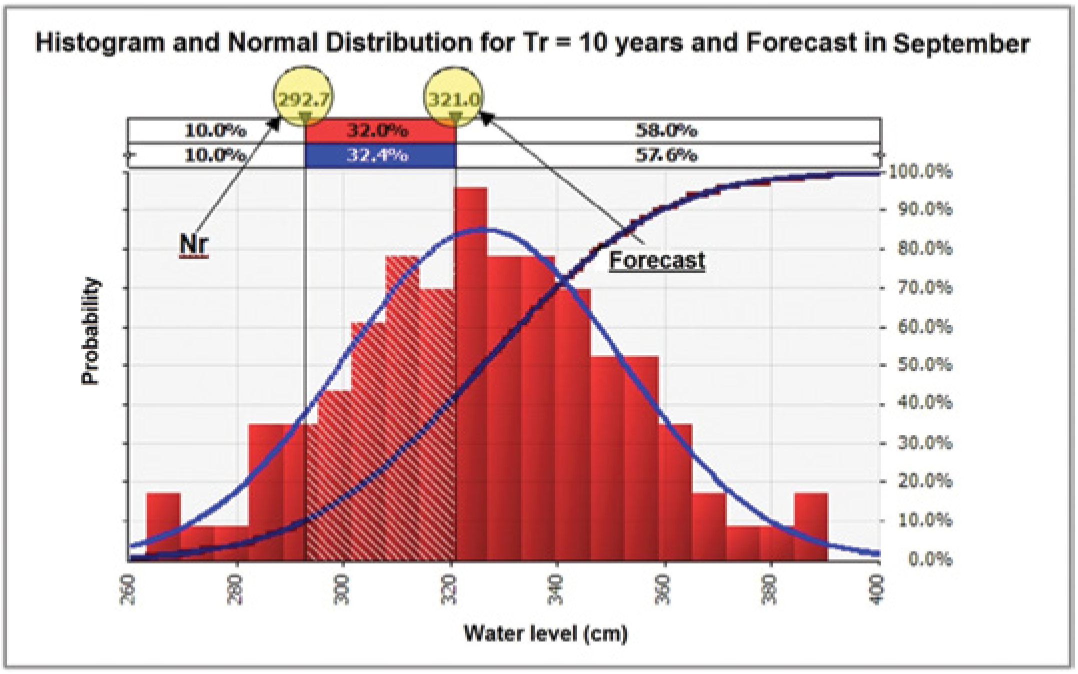

Figure 11: *Permanence of the water level forecast in relation to a reduction level for the month of September at the Barra do São Manuel station.*

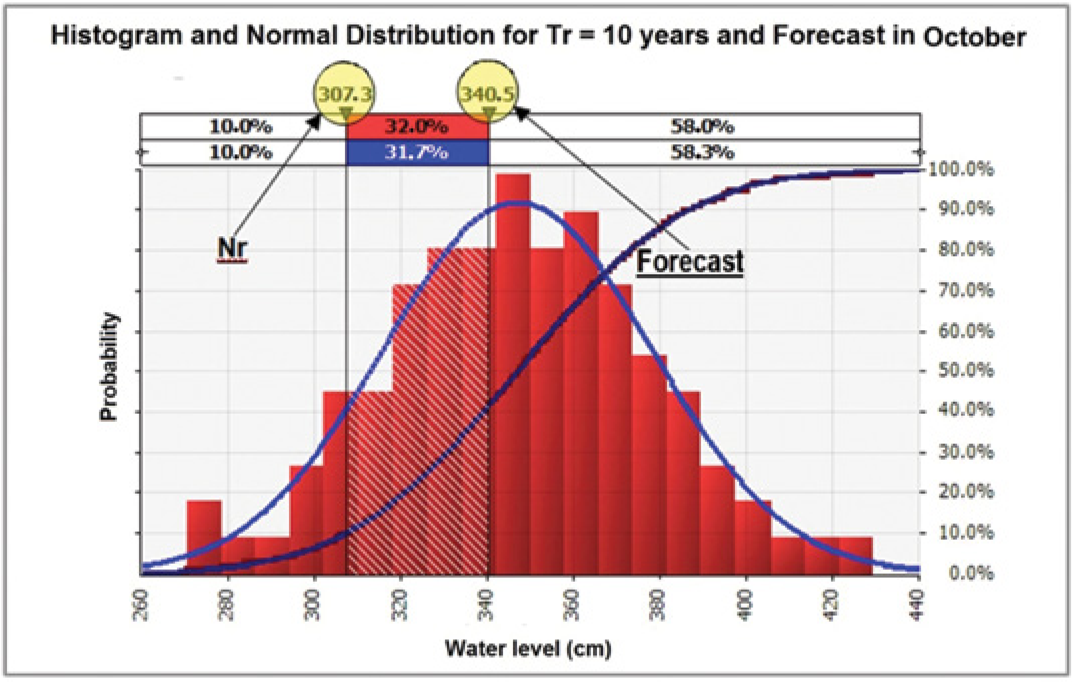

Figure 12: *Permanence of the water level forecast in relation to a reduction level for the month of October at the Barra do São Manuel station.*

VI. CONCLUSIONS

The ARIMA model was implemented and verified based on the methodology proposed by Box & Jenkins (1976). Auto regressive models and moving averages were adjusted to the water level series of the Tapajós River. The calibration and validation of the model presented an average R^2 value of 0.90 and a RMSE value less than 0.16. These results showed that the model satisfactorily captured the behaviour of the series of water levels. The drought simulations showed

Table 7: Additional loads designed by the ARIMA model.

Month	Nr (cm)	Na (cm)	Gain Water depth (cm)	Additional Load (ton) /trip	Annual Load (ton)
Jan	587	706	119	3,740	18,588
Feb	687	791	104	3,269	16,247
Mar	782.8	877.1	94.3	2,964	14,731
Apr	738.6	834.5	95.9	3,014	14,,980
May	569.3	657.5	88.2	2,772	13,777
Jun	423.9	481.8	57.9	1,820	9,045
Jul	348.4	387.5	39.1	1,229	6,108
Aug	306	333.6	27.6	868	4,314
Sep	292.7	321	28.3	890	4,423
Oct	307.3	340.5	33.2	1,044	5,189
Nov	259	402.2	143.2	4,500	22,365
Dec	459.5	525	65.5	2,059	10,233

better results because of its decreased variability interquartile range. In the period of floods, the model must be used with caution because a large number of extreme events have been somewhat underestimated.

Forecasts of water levels, and consequently gains in depth and cargo capacity of the boats, might be obtained using the ARIMA model. For critical periods of drought, the model predicted a high probability of Na, 32% above the Nr, and designed a medium depth gain on the order of 30 cm, resulting in an additional charge per trip of approximately 2,800 tons and approximately 14,000 tons per year. When the analysis was conducted for the two seasonal periods (Jan. to Dec.), the model ensured average depth increments of approximately 75 cm and a freight economy of $350, 000.00 per year.

It should be noted that forecasts of water levels, with 24 months in advance, were fully reliable. Furthermore, this model can be used to forecast with intervals of times smaller monthly, or even weekly. In summary, the model allows a capacity planning of barge convoy load, depending on the depth of the river.

The adoption of the stochastic ARIMA model to predict water levels is of great importance and greatly helps in the achievement of terminals. Models that perform probabilistic analysis and bathymetric surveys that characterise relief funds are tools that allow for the exploration of navigability conditions and the scaling of waterway infrastructure works that are required to make the waterway. Overall, it was concluded that the chosen ARIMA model presented a fully satisfactory behaviour of the research objectives in question and allowed the fulfillment of the research objectives.

VII. REFERENCES

Babcock M. W. And Lu X. (2002) — Forecasting inland waterway grain traffic.

http://dx.doi.org/10.1016/S1366-5545 (01) 00017-5. *Transportation Research Part E: Logistics and Transportation Review.* **38(75)** 74

Birinci V.; Akay O. (2010) — A Study on Modeling Daily Mean Flow with MLR , ARIMA and RBFNN. *Proceedings of BALWOIS: Water observation and information system for decision support. Ohrid , Republic of Macedonia. 25-29 May 2010*

Boochabun K., Tych W., Chappell N.A., Carling P.A., Lorsirirat K., Pa-Obsaeng S. (2004) — Statistical modelling of rainfall and river flow in Thailand. *J. Geol. Soc. India.* **64** 503–515

Box G. E. P. And Jenkins G. M. (1976) — *Time series analysis: forecasting and control, 2nd ed. San Francisco: Holden – Day*

Chattopadhyay S., Chattopadhyay G. (2010) — Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. *CR Geoscience.* **342** 100–107

Chattopadhyay S., Jhajharia D., Chattopadhyay G. (2011) — Univariate modelling of monthly maximum temperature time series over northeast India: neural network versus Yule–Walker equation based approach. *Meteorol. Appl.* **18** 70–82

Collischonn B. (2006) — Use of precipitation estimated by the TRMM satellite on distributed hydrological model, Master's thesis. *IPHUFRGS* 193 p

Jack Faucett Associates (1997) — *Waterway trafc forecasts for the upper Mississippi River Basin, vol. I Summary and vol. II grain. Available online at http://www.mvr.usace.army.mil/pdw/nav_study/JFAreport.pdf*

Jack Faucett Associates (2000) — Review of historic and projected grain trafc on the upper Mississippi River and Illinois Waterway: an addendum. *Available online at http://www.mvr.usace.army.mil/pdw/nav_study/JFAreport.pdf*

Koutroumanidis T.; Sylaios G.; Zafeiriou E.; Tsihrintzis V. A. (2009) — Genetic modeling for the optimal forecasting of hydrologic time-series: Application in Nestos River, 10.1016/j.jhydrol.2009.01.041. *Journal of Hydrology.* **38(1-4)** 156-164

Ledolter J. (1976) — ARIMA Models and Their Use in Hydrologic Modelling Sequences. *IIASA Research Memorandum RM- 76-069.* 45 p

Ledolter J. (1977) — The Analysis of Multivariate Time Series with a View to Applications in Hydrology. *IIASA Research Memorandum RM- 77-011.* 33 p

Lohani A.K.; Kumar R.; Singh R.D. (2012) — Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques, 10.1016/j.jhydrol.2012.03.031. *Journal of Hydrology.* **442-443** 23-35

Narayanan P., Basistha A., Sarkar S., Kamna S. (2013) — Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. *Comptes Rendus Geoscience.* **345(1)** 22-27

Nourani V.; Baghanam A. H.; Adamowski J.; Gebremichael M. (2013) — Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. 10.1016/j.jhydrol.2012.10.054. *Journal of Hydrology.* **476** 228-243

Pektaş A.O., Kerem Cigizoglu H. (2013) — ANN hybrid model versus ARIMA and ARIMAX models of runoff coefficient. *Journal of Hydrology.* **500** 21-36

Salas, Jd; Delleur J.W; Yevjevich V.; Lane, Wl (1980) — Applied Modeling of hydrologic time series. *Water Resources Publications, Littleton, Colorado.* 484p

Salas, Jd (1992) — Analysis and modeling of hydrologic time series. *Handbook of Hydrology. Maidment, D. R. (editor). McGraw – Hill.* **19** 72p

Tang X. (2001) — Time of forecasting quarterly grain barge tonnage on the McClellan - Kerr Arkansas River navigation system series. *Journal of the Transportation Research Forum.* **40(3)** 91-108

Wu C.L.; Chau K.W. (2010) — Data-driven models for monthly streamflow time series prediction, 10.1016 / j.engappai.2010.04.003. *Engineering Applications of Artificial Intelligence.* **23(8)** 1350-1367