

# Dam Uplift Pressure Forecast Using Dynamic Regression

**Júlio C. Royer<sup>1</sup>**

Instituto Federal do Paraná, [www.ifpr.edu.br](http://www.ifpr.edu.br)  
Campus Foz do Iguaçu, Av. Araucária, 780, Vila A  
CEP 85.860-000, Foz do Iguaçu, PR, Brazil

**Volmir E. Wilhelm**

Universidade Federal do Paraná, [www.ufpr.br](http://www.ufpr.br)  
Depto. Eng. Produção, Centro Politécnico, Av. Cel Fancisco H. dos Santos,  
210, Jardim das Américas, CEP 81531-970, Curitiba, PR, Brazil

**Josiele Patias**

Itaipu Binacional, [www.itaipu.gov.br](http://www.itaipu.gov.br)  
Depto. Eng. Civil, Av. Tancredo Neves, 6.731  
CEP 85856-970, Foz do Iguaçu, PR, Brazil

Copyright © 2015 Júlio C. Royer, Volmir E. Wilhelm, and Josiele Patias. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

In order to keep a large dam safe it is important to monitor the uplift pressure, and carry out some maintenance procedure if its value reaches a defined limit. The uplift pressure level oscillates according to temperature or reservoir level changes, for instance, which may be different in summer and winter. So, defining a fixed limit for the uplift alert doesn't help to discover an abnormal high uplift pressure when it should be close to the minimum for the year. This paper describes a Dynamic Regression Model using ARIMA to model the residuals to estimate the uplift pressure, based on air temperature, water temperature and reservoir level

---

<sup>1</sup> This research is partially funded by the Araucária Foundation, project number 44.721.

time series. The results obtained have a significant improvement when compared to the multiple regression procedure.

**Keywords:** Dynamic Regression, Time Series, ARIMA, Dam Uplift Pressure

## 1 Introduction

Concrete dams are usually built over rock foundations. And rock foundations are not continuous, but usually fractured due to its geological history. The water that seeps through these discontinuities generates the uplift pressure that relieves the structure weight and modifies the friction coefficient, bringing risks to the dam safety if it reaches certain values defined in the structural project [1, 10]. The hydraulic conditions should be monitored during all lifetime of the dam, because they may change over the time due to chemical or mechanical effects, like erosion or material deposition over the water path inside the foundations [7, 3, 12]. In large dams this monitoring process generates a huge amount of data, stored in databases, with graphics available for the analysis conducted by engineers. These systems usually offer alert levels for each instrument.

In Itaipu dam there are static alert levels, for instance, yellow alert if the uplift level is above the maximum historic level and red alert if the project safety limit for that instrument was reached. The alert limits are usually fixed and doesn't help to identify an increasing uplift pressure when it is close to the minimum of the yearly oscillation. By the other hand, there are few published works on methods for defining dynamic alert levels for uplift pressure. One example is [9], where the authors use multiple regression to specify dynamic limits for alert, when the *adjusted R<sup>2</sup>* coefficient is greater than 0.7, or, in the other cases, they use the mean and standard deviation, establishing the limits in a way that 95% of the data is between the limits. But this technique didn't work for Itaipu dam, where the multiple regression reaches *adjusted R<sup>2</sup>* values around 0.5, which means that the independent variables explains about 50% of the uplift pressure variation.

To improve the accuracy of the alert system, this work presents a methodology for forecasting the uplift pressure based on environmental variables in order to define the uplift pressure yellow alert for a given instrument, using dynamic regression with ARIMA modelling of the residuals time series.

The rest of this paper is organized as follows: the section 2 presents an overview of the ARIMA modelling for time series, section 3 deals with linear transfer function, used to predict a dependent variable as a consequence of other independent variables, including time lags and residuals time series modelling, and section 4 discusses the application of this technique to predict the uplift pressure level on Itaipu main dam. Finally section 5 presents the conclusions.

## 2 ARIMA Time Series Modelling

A time series is a sequence of data observed in a regular time interval basis and that depend on the adjacent previous observations. The time series analysis is

based on techniques for the analysis of the data dependency [2, 14]. In the ARIMA model, a time series of the Y variable, represented by the values  $\{Y_1, Y_2, \dots, Y_t, \dots, Y_{n-1}, Y_n\}$  or as the series  $\{Y_t, t=1, 2, \dots, n\}$  may contain autoregressive (AR) components, moving average (MA) components and/or integration (I). It may also have seasonal autoregressive, moving average and/or integration components. To express a time series in a compact way, the backward operator B, defined as  $B^m Y_t = Y_{t-m}$ , and the differencing operator  $\nabla$ , defined as  $\nabla^d = (1-B)^d Y_t$  are often used [2].

In an autoregressive model with degree 2, AR(2), each value of  $Y_t$  have a significant influence of the two previous ones,  $Y_{t-1}$  and  $Y_{t-2}$ , and the series may be written as  $Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t$ , or, in the backward operator form  $\phi(B)Y_t = C + a_t$ , where C is a constant,  $a_t$  is a random shock,  $\varphi_1$  and  $\varphi_2$  are impact coefficient of  $Y_{t-1}$  and  $Y_{t-2}$  on  $Y_t$ , and  $\phi(B)$  is the characteristic polynomial, which in this case (order 2) is  $\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2$ . An AR (p) can be expressed as AR (p) as  $Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + a_t$ , or  $\phi(B)Y_t = C + a_t$  where  $\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ .

In a similar way, an MA (q) model can be expressed as  $Y_t = C + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$ , or  $Y_t = C + \theta(B)a_t$ , where  $\theta(B)$  is the characteristic polynomial,  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ .

There are also time series with p autoregressive and q moving average components combined, making a ARMA (p,q) model, which can be expressed as:

$$Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Or, in the backward operator format:  $\phi(B)Y_t = C + \theta(B)a_t$ .

The Box&Jenkins methodology requires that the time series is stationary in the mean and variance. If the series is not stationary, it is possible to apply data transformations. One transformation often applied is do differentiate the series, represented by the ARIMA (p,d,q) model, where p is the  $\phi(B)$  order, q is the  $\theta(B)$  order and d is the differentiation order  $\nabla^d$ . Thus, the series can be expressed as  $\phi(B)\nabla^d Y_t = \theta(B)a_t$  [2]. The series also may have seasonal behavior, and  $Y_t$  value may also depend on the  $Y_{t-S}$  or  $a_{t-S}$  values, and may be necessary to seasonally differentiate the series. The representation of an ARIMA (p,d,q) (P,D,Q)<sub>S</sub> model in the backward operator format is:

$$\Phi(B^S)\phi(B)\nabla_S^D \nabla^d Y_t = \Theta(B^S)\theta(B)a_t$$

where S is the seasonality (number of samples in a cycle),

$$\begin{aligned} \phi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \\ \Phi(B) &= 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_q B^{pS}, \\ \Theta(B) &= 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_q B^{qS}, \\ \nabla^d &= (1 - B)^d \quad \text{and} \\ \nabla_S^D &= (1 - B^S)^D. \end{aligned}$$

### 3 Linear Transfer Function

In a set of observations of two variables  $X$  and  $Y$  of a population, it is possible to trace a linear regression between the variables to see if (and how) the variation of  $X$  influences the variation of  $Y$ . For instance, if  $X$  is the father's height and  $Y$  is his adult son's height, one might want to identify if (and how) the father's height influences the son's height. This can be done by a simple linear regression between  $X$  and  $Y$  [11].

$$Y_i = C + bX_i + a_i \quad (1)$$

This is a linear transfer function forming a straight line equation, where  $C$  is a constant,  $b$  is a coefficient that indicates how a variation of one unit on  $X_i$  impacts  $Y_i$ , and  $a_i$  represents the effect of all unconsidered independent variables on  $Y$ .

If besides father's height ( $X_1$ ) it is available also the mother's height ( $X_2$ ) it is possible to determine if (and how) the height of the son is influenced by his parents' height, tracing a multiple regression.

$$Y_i = C + b_1X_{1,i} + b_2X_{2,i} + a_i$$

Of course, if there are  $n$  independent variables ( $X_1, X_2, \dots, X_n$ ), the multiple regression is written as

$$Y_i = C + b_1X_{1,i} + b_2X_{2,i} + \dots + b_nX_{n,i} + a_i \quad (2)$$

Considering all data as time series, the minimum squared estimator for the  $b_i$  coefficients requires the independency of the residuals  $a_i$ . If there is autocorrelations between the  $a_i$  values, there is information to be extracted from residuals, the estimator is invalid and the  $b_i$  values are not as accurate as they could be [8]. Thus, if there is residual autocorrelations, (1) can be written as

$$Y_t = C + bX_t + N_t \quad (3)$$

where  $N_t$  is the autocorrelated residual series. Suppose  $N_t$  has the behavior of a ARIMA(1,0,1) model. Then it can be expressed as  $(1 - \varphi_1 B)N_t = (1 - \theta_1 B)a_t$ , or  $N_t = \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)} a_t$ .

Now using the  $N_t$  expression, equation (3) can be rewritten as:

$$Y_t = C + bX_t + \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)} a_t$$

Then, it is possible to use the minimum squared procedure to estimate  $C$ ,  $b$ ,  $\theta_1$  and  $\varphi_1$ , minimizing the  $a_t$  values. But the  $N_t$  ARIMA model is needed to estimate  $C$  and  $b$ . By the other hand,  $C$  and  $b$  are needed to calculate the  $N_t$  values, and then discover the better  $N_t$  ARIMA model. So, it is necessary to start with a low order ARIMA guess model, like ARIMA(1,0,1), to estimate the regression parameters and to calculate the residuals. If residuals are white noise, the result is found, and the process ends. Otherwise, it is necessary to check the residuals model and repeat the process until the residuals are white noise.

It is common to find time series in real situations where the independent variable  $Y_t$  receives influences not only from the  $X_{1,t}, X_{2,t}, \dots, X_{n,t}$ , but also from the  $k_j$  previous values, for example  $X_{1,t}, X_{1,t-1}, \dots, X_{1,t-k_1}, X_{2,t}, X_{2,t-1}, \dots, X_{2,t-k_2}, \dots, X_{n,t}, X_{n,t-1}, \dots, X_{n,t-k_n}$ . This can be expressed as a dynamic regression, and written as [8], [18], [19]:

$$Y_t = C + v_1(B)X_{1,t} + v_2(B)X_{2,t} + \dots + v_n(B)X_{n,t} + N_t \quad (4)$$

where  $v_i(B)X_t = v_{i,0}X_{i,t} + v_{i,1}X_{i,t-1} + \dots + v_{i,k}X_{i,t-k}$ ,

$C$  is a constant and

$N_t$  is the residuals series, possibly with autocorrelations

#### 4 Forecasting Uplift Pressure Time Series

This session presents results of using this technique for obtaining the linear transfer function to forecast the uplift pressure on the foundations of Itaipu main dam which have seasonal variations [4]. The data presented here are from the concrete-rock contact piezometer *PS-F-73*, from the gravity relieved *F19/20* block. The statistical software used was *R* [13] with the *forecast* package [5]. The procedure used in this work to obtain the linear transfer functions has the following steps [14]:

1. To study the time series and the factors that may impact the uplift pressure, consulting the specialists;

This step included technical visits to the dam foundations, interviews with engineers, search and copy of the time series data from Itaipu database, research on technical documents of Itaipu dam foundations, and study of technical books on water flow in fractured rock masses applied to dam foundations [1, 8, 10]. Thus the main theoretical influences found are the air temperature, the temperature of water that passes on the piezometer, the reservoir level and punctual interventions, like drain cleaning procedures or earthquakes. There were no significant interventions detected on *PS-F-73*.

2. To graph the time series, identifying correlations between independent variables and uplift pressure, time lags, outliers and interventions;

The period selected is from the beginning of 2006 to the beginning of 2014, where all the data is available. The original time series of uplift pressure on *PS-F-73* (in m), air temperature, water temperature and reservoir level are showed in figures 1, 2, 3 and 4, respectively.

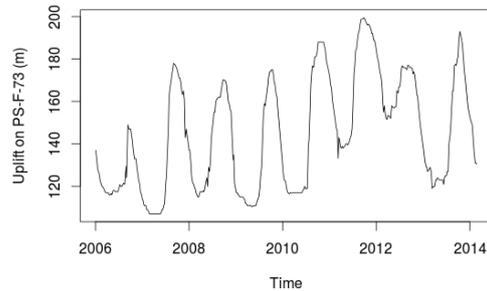


Figure 1: Uplift pressure on PS-F-73, in meters above sea level, from jan/2006 to feb/2014.

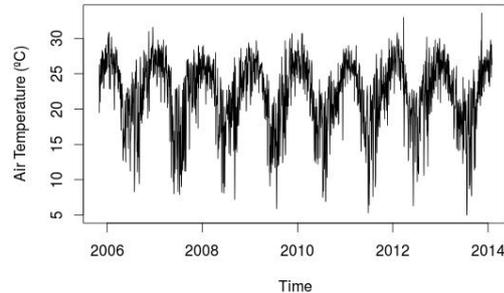


Figure 2: Daily average air temperature, in °C, from nov/2005 to jan/2014.

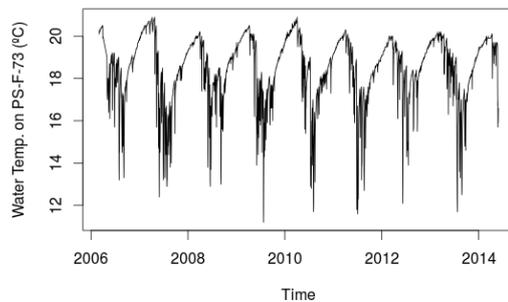


Figure 3: Daily average water temperature, in °C, on PS-F-73 from mar/2006 to may/2014.

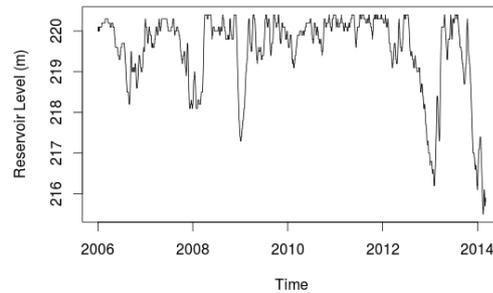


Figure 4: Daily average reservoir level, in meters above sea level from jan/2006 to feb/2014.

From these graphs it is possible to find:

- A visible correlation between air temperature and the uplift pressure, with about 3 month of time lag;
- Possible influence in the uplift pressure from the temperature of the water passing on the piezometer, without time lag;
- The water temperature has a fast cooling and a slow warming, due to water convection and heat conduction, respectively;
- No visible influence from the reservoir level. This variable remains in the analysis for statistical significance tests.

3. To establish regular interval bases for uplift pressure time series, by eliminating duplicated readings and extra readings (for maintenance procedures, for example);

The uplift pressure data are collected manually since the piezometer installation, during the dam construction. In critical moments, like when the reservoir was being filled, and on the first operation years, the instrument reading frequency was higher. After that, knowing the instrument behavior, the reading frequency became lower. For instruments with higher pressure variation the reading frequency remained higher. There are also extra readings during drain cleaning procedures. For *PS-F-73*, there are data periods with daily, twice a week

and weekly frequency, so the selected frequency was weekly data, available on the whole period, fixing the number of data per year in 52 due to seasonality tests.

4. To establish average of daily, weekly, monthly and quarterly readings for all independent variables and for uplift pressure, in order to identify what average better explains the variation of uplift pressure. Add columns with previous readings ( $t-1$ , ...,  $t-k$ ) for the variables which influence is suspected with time lags;

The best average found was the monthly average. The average of the last 30 days for each independent variable, and for the uplift pressure was calculated. Four columns were added with the air temperature average with 1, 2, 3 and 4 past months, since the graphs indicates a possible influence of the air temperature of 3 months before.

5. To perform significance tests with multiple regression with all independent variables, and for each average reading interval. Check if the *adjusted R<sup>2</sup>* result is acceptable and if the residuals are white noise;

The first multiple regression test included the monthly average of air temperature (with 3 month time lag), water temperature and reservoir level. The *adjusted R<sup>2</sup>* was 0.5631, indicating that this variables explain 56.31% of uplift pressure variation. The root mean squared error (RMSE) is 17.20986. Air temperature and water temperature are considered significant at 95% of confidence. Reservoir level is not considered significant.

The second test was done without reservoir level. Both air temperature and water temperature were significant, but the *adjusted R<sup>2</sup>* and RMSE remained almost the same (0.5677 and 17.21152).

In the third test the variable included were air temperature, with 2, 3 and 4 months of time lag, water temperature and reservoir level. The *adjusted R<sup>2</sup>* got 0.5896, and RMSE 16.49688. This is a bit better than the previous tests. But it is not yet acceptable and only air temperature with time lag of 4 month was considered significant at 95% of confidence and water temperature at 90%.

The fourth test included the air temperature with 2 and 4 months of time lags and water temperature. Both air temperature were significant at 95% and water temperature was considered significant at 90% and almost significant at 95% (*p-value*: 0.05431). The *adjusted R<sup>2</sup>* was 0.59065 and the RMSE was 16.65842. This is also an improvement, but not good enough. Figure 5 shows the estimate using the last model, the real time series values and the residuals. From the graph it is possible to see an approximation, but with relative large errors, from -36m to +37m. The residuals time series also doesn't seems to be white noise, which is confirmed by the acf and pacf graphs showed in figure 6.

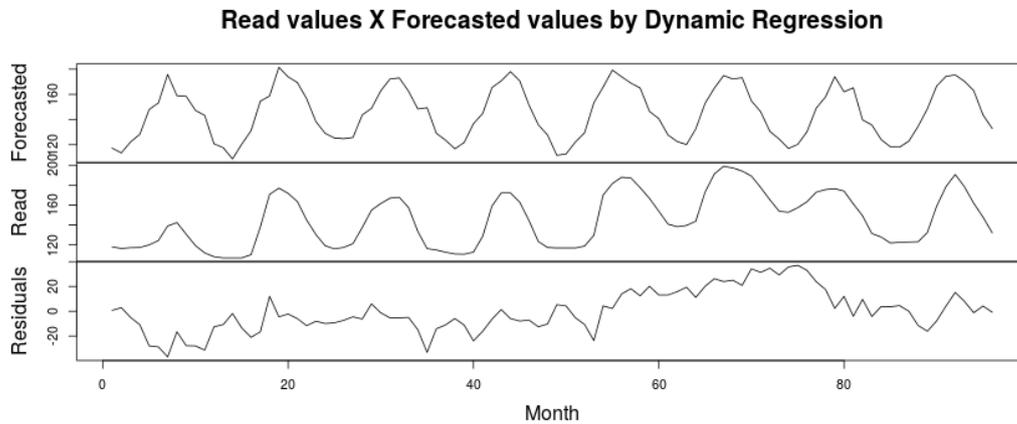


Figure 5: Graphs of *PS-F-73* forecasted uplift pressure, read uplift pressure and residuals time series, using dynamic regression from air temperature with 2 and 4 month lags and water temperature

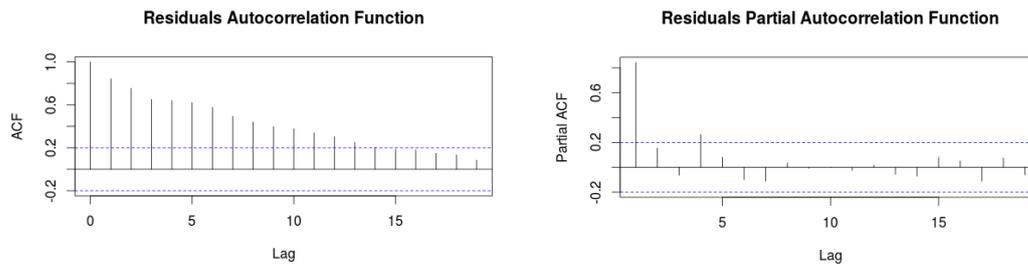


Figure 6: Residuals Autocorrelation Function and Residuals Partial Autocorrelation Function for *PS-F-73* Dynamic Regression

6. If the  $a_t$  residuals are not white noise or the *adjusted*  $R^2$  is not acceptable, the residuals series is called  $N_t$ , the ARIMA time series modelling is applied on  $N_t$ , starting with a low order AR model, new regression equation is calculated and the regression coefficients are estimated minimizing the new  $a_t$  residuals. If the  $a_t$  residuals are white noise, the model is accepted. Otherwise the residuals' *acf*, *pacf* and *eacf* are examined to try another ARIMA model, until the residuals series is white noise.

Running an ARIMA automatic model selection for the residuals of a dynamic regression, considering all the tested combinations of independent variables tested, the best result was found including air temperature with time lag of 2, 3 and 4 months, water temperature and reservoir level. The *adjusted*  $R^2$  rose from 0.5906499 to 0.9379588 and RMSE felt from 16.65842 to 6.342698. From all independent variables, just the reservoir level is not significant at 95% of confidence and the water temperature is on the significance limit. Leaving out these two variables, and running the regression again, the results are almost the same (RMSE: 6.504334 and *adjusted*  $R^2$  : 0.9362063). Figure 7 shows the best forecast results, when compared with figure 5. In figure 8 it is possible to admit that the  $a_t$  residuals series is white noise, which is confirmed by Box-Pierce test.

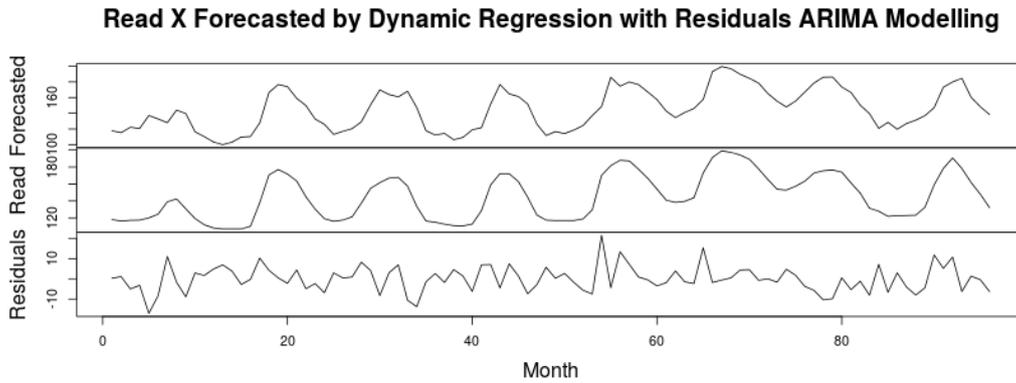


Figure 7: Graphs of *PS-F-73* forecasted uplift pressure, read uplift pressure and residuals time series, using dynamic regression with Residuals ARIMA modelling from air temperature with 2, 3 and 4 month lags, water temperature and reservoir level.

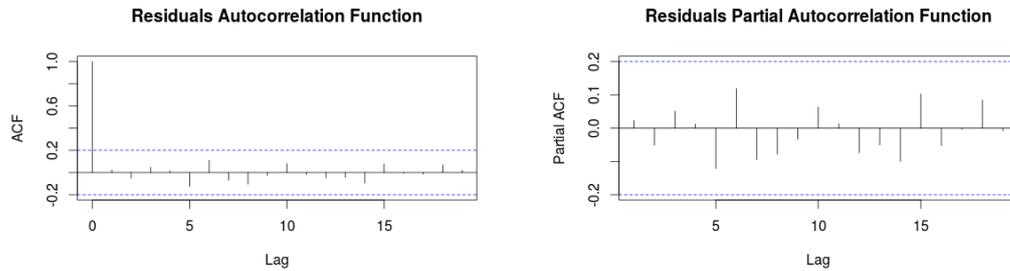


Figure 8: Residuals Autocorrelation Function and Residuals Partial Autocorrelation Function for *PS-F-73* Dynamic Regression with Residuals ARIMA Modelling

As the residuals seem to be white noise and the *adjusted R<sub>2</sub>* explains 93% of uplift pressure variation, which is good enough, the process ends. The residuals model selected was ARIMA (2,1,0), which means that all series must be differentiated once. For this model, the forecast equation is

$$\nabla Y_t = b_1 \nabla A_{t-2} + b_2 \nabla A_{t-3} + b_3 \nabla A_{t-4} + b_4 \nabla W_t + b_5 \nabla R_t + \frac{a_t}{(1 - \varphi_1 B - \varphi_2 B^2)}$$

- Where  $A_{t-k}$  is the Air temperature average for month t-k
- $W_t$  is the Water temperature average for month t
- $R_t$  is the Reservoir level average for month t
- $b_i$  is the Estimated dynamic multiple regressive coefficient i
- $\varphi_i$  is the Residual autoregressive coefficient i

The final equation with coefficient found was:

$$Y_t = Y_{t-1} - 1.6483(A_{t-2} - A_{t-3}) - 2.5978(A_{t-3} - A_{t-4}) - 1.1556(A_{t-4} - A_{t-5}) - 1.5029(W_t - W_{t-1}) + 0.0020(R_t - R_{t-1}) + \frac{a_t}{(1 - 0.3060B + 0.4536B^2)}$$

## 5 Conclusions

One of the benefits of regression procedures to forecast a time series is the possibility of explanations of how each independent variable impacts the dependent variable. In the case of uplift pressure on dam foundations, if reservoir level is significant, it is possible to decide to operate the dam at a lower level in critical situations, for instance.

In Itaipu dam, the best *adjusted R<sup>2</sup>* got with dynamic multiple regression explains less than 60% of the variation of uplift pressure which is unacceptable for forecasting. But using the proposed technique, using dynamic multiple regression with residual ARIMA modelling got the *adjusted R<sup>2</sup>* explaining more than 93%, which is good enough.

The next steps of the study include the data treatment of punctual interventions, like drain cleaning or earthquakes, and defining the limits of the 95% confidence interval for the uplift pressure forecast, serving as yellow alerts.

**Acknowledgements.** The authors thank Itaipu engineers Dimilson Pinto Coelho and Silvia Frazão Matos, for contributing to technical discussions on dam behavior, the Araucaria Foundation for the partial financial support, the IFPR for granting a time for exclusively dedication to the research, and UFPR, Unioeste, Ceasb and PTI Foundation for promoting the PhD group in Numerical Methods on Engineering which motivated this research.

## References

- [1] R. M. Andrade, *Mechanics of flow in fractured rock masses applied to dams*, Engevix, Rio de Janeiro, 1988.
- [2] G. Box, G. Jenkins, G. Reinsel, *Time series analysis: forecasting and control*, John Willey & Sons, Hoboken, New Jersey, 2008.  
<http://dx.doi.org/10.1002/9781118619193>
- [3] CDBD, Comitê Brasileiro de Barragens, *Guia básico de segurança de barragens*, XXIII Seminário Nacional de Grandes Barragens, Belo Horizonte, 1999.
- [4] A. S. Fiorini et al, *Barragem de Itaipu: Comparação dos diferentes tipos de barragens de concreto*, XXIV Seminário Nacional de Grandes Barragens, Fortaleza, 2001.
- [5] R. Hyndmann, et al, *Forecasting functions for time series and linear models*, R Package Version 5.6, 2014. <http://CRAN.R-project.org/package=forecast>

- [6] R. Hyndmann, G. Athanasopoulos, *Forecasting: principles and practice*, 2014. <http://otexts.org/fpp>.
- [7] ICOLD, Internation Comission on Large Dams, *Dam Safety Guidelines*, Bulletin 59, Paris, 1987.
- [8] Itaipu Binacional, *Usina Hidrelétrica de Itaipu: aspectos técnicos das estruturas civis, Foz do Iguaçu, 2008*.
- [9] S. C. Kuperman, et al, *Critérios para a fixação de valores limite na instrumentação civil de barragens de concreto e terra*, XXV Seminário Nacional de Grandes Barragens, Salvador, 2003.
- [10] A. Mascarenhas, *Qualidade hidráulica e deteriorações em fundações de barragens de betão*, Laboratório Nacional de Engenharia Civil, Lisboa, 2005.
- [11] A. Pankratz, *Forecasting with dynamic regression models*, John Wiley and Sons, New York, 1991. <http://dx.doi.org/10.1002/9781118150528>
- [12] C. I. Osako, *A manutenção dos drenos das fundações de barragens: O caso da usina hidrelétrica de Itaipu*, Master dissertation on civil construction, UFPR, Curitiba, 2002.
- [13] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, 2014.
- [14] R. Yaffee, M. McGee, *Introduction to time series analysis and forecasting*, Academic Press, San Diego, 2000.

**Received: February 12, 2015; Published: March 5, 2015**