XLIII CILAMCE
21-25, November, 2022
FOZ DO IGUAÇU - PR - BRAZIL

Forecast of time series earned by the Piezometer through Method Multiple Kernel Sarima Support Vector Regression Wavelet

Samuel B. Rodrigues[1], Jairo M. Corrêa[1], Tássia Hickmann[1], Lucas S. Ribeiro[1], Levi L. Teixeira[1], Etore F. Faria [2]

[1]*Federal University Of Technology – Paraná*
Avenida Brasil, 4232, 85884-000, Medianeira – Paraná, Brasil
samuelb@utfpr.edu.br, jairocorrea@utfpr.edu.br, hickmann@utfpr.edu.br, l.ribeiro@utfpr.edu.br, levilopes@utfpr.edu.br
[2] *Laboratory of Concrete Technology of Itaipu, Itaipu Binacional*
Av. Tancredo Neves, 6731, 85867-900, Foz do Iguaçu – Paraná, Brasil
etore@itaipu.gov.br

Abstract. In this study time series predictions were made from measurements of the instrument called piezometer (PS), located in the key block (I10) of stretch I of the Itaipu hydroelectric dam. The results show that the forecasting performance attained by the method called SARIMA Support Vector Regression Wavelet of Multiple Kernels (SSVRWMN) was notably superior to predictive methods SARIMA, SVR, and SARIMA-SVR combined. Comparing it to the second-best result (namely, the SVR method), the relative reduction was approximately 39.1% in the mean square error (MSE) accuracy measure.

Keywords: structural health monitoring, hybrid method, statistical techniques, machine learning techniques.

1 Introduction

Itaipu Binacional is one of the largest hydroelectric plants in the world and is located on the Paraná River upstream of the international bridge that links the city of Foz do Iguaçu, in Brazil, to Ciudad del Este, in Paraguay [1]. Itaipu's main dam is made of concrete, of the relieved gravity type, composed of 20 blocks, each one equipped with a generating unit. In total, there are 2,792 instruments installed in the concrete and in the foundation, providing temporal data that help in the analysis of the behavior of the blocks and, consequently, of the dam as a whole. Some blocks are designated as key blocks (as is the case of Block I10, an object of our study), which are equipped with a greater number of instruments.

The blocks are subject mainly to the action of the level of the lake due to the greater force that they receive from the water in the upstream-downstream direction, that is, in the direction of the river bed. The volume of water also exerts pressure on the lower parts of the blocks (called sub-pressure) which creates an effect contrary to the one its mass exerts on the foundation, according to the variations of temperature, as in the case of summer when dilatations of the concrete occur that provoke a tendency of deformation of the block toward the upstream direction, which by its turn may provoke increase the compression tensions at the upstream foot of the blocks. In winter the concrete contracts, provoking a tendency of deformation of the block at downstream. One may thus identify a cyclic behavior of the structure, closely conditioned by the environmental conditions of the region [2].

The piezometer is an instrument installed in the foundation, being responsible for the admeasurement of the sub-pressure acting where it is installed, that is, its measurements gauge the sub-pressures on the foundation. According to [2], the greatest gravity disasters in dams occurred by inadequate resistance to horizontal movement, that is, rupture by shear stress in the foundation. For this reason, the effects of sub-pressure must be acknowledged and carefully studied. In the foundation of a concrete dam sub-pressure acts in the upward direction, that is from the bottom to the top, reducing the effective weight of the structure and consequently its resistance to shearing of the potential slipping points that exist in the rock bed.

Statistical techniques, machine learning techniques, and technological progress are increasingly permitting the use of forecasts that help decision-makers in analyses and planning of future operational needs.

The method SARIMA Support Vector Regression Wavelet of Multiple Kernels (SSVRWMN) is ensemble learning method what combine forecasts by a linear model (SARIMA) and a non-linear one (Support Vector Regression Wavelet of Multiple Kernels [3]. Box and Jenkins' methodology aims to identify a plausible probabilistic system that generates a time series that exhibits second-order stationarity (ie, constant mean and covariance) and linear self-dependence structure (autocorrelation), using only the information contained therein. The hybrid model requires the forecast of a linear component and for this purpose are considered the forecasts obtained by the SARIMA model which aim at capturing the structures of linear self-dependence, taking into consideration the seasonal effect. In case there is no seasonality, the ARIMA model will be considered [4].

The Wavelet function is capable of decomposing or representing a time series originally described in the domain of time in such a way that the series can be analyzed in different scales of frequency and time [5].

The Support Vector Machines (SVMs) form a system based on the theory of statistical learning or VC theory (Vapnik – Chervonenkis) being extended for the case of regression, so-called Support Vector Regression (SVR), making thus possible for the estimation of functions of real values [6]. The basic idea of SVR consists of imagining a margin around the outline of the transformation function and finding a function having at most a previously fixed error in all examples, thus trying to obtain a margin as narrow as possible.

According to [7] combination is an attractive method to obtain forecasts, since instead of choosing the best technique, the problem becomes deciding which techniques may be of greater help for improved accuracy, since the performance of combination is, in general, assessed through accuracy.

The Bootstrap method is a non-parametric statistical procedure whose main idea is the re-sampling with repositioning of the original data in order to create new sets of data that allow the estimation of a measure of interest [8]. Bootstrap residual, in its turn, requires a regression model adjusted over the original set of data and the calculation of the residues. In this way, the residues are re-sampled and new series are obtained [9].

The application of regression models involves the assessment of uncertainties that surround those who estimate parameters. The trust interval is a valuable statistic and performs this important function. They are used to estimate the parameters of the regression model [10].

In this study, time series predictions were made from measurements of the instrument called piezometer, located in the key block (I10) located in stretch I of the Itaipu hydroelectric dam, and that contributes to the monitoring of the behavior of the block, through the hybrid predictive method called SARIMA Support Vector Regression Wavelet of Multiple Kernels (SSVRWMN), to produce predictions that are aggregators of distinct stochastic information captured by different methods. The confidence interval is estimated using the bootstrap-residual, technique with the point predictions.

2 Results

In this section are presented the results obtained from the modeling of the time serie from the piezometer measurements through the hybrid method SSVRWMN, the results were compared with the Box and Jenkins approaches, Support Vector Regression, and the SARIMA-SVR composite. It was noticed that the proposed method SSVRWMN showed a reduction in the adherence statistics MSE, MAPE, and MAE in the forecasts of the tested time serie.

Initially, the treatment of the time series information from the piezometer measurements was carried out, to be structured according to the forecast methodology. An option was made to work with the period 1994-2014, with monthly values. Of the 252 values that the time series possesses, the first 228 data were used for the adjustment of the base models (SARIMA and SVR, SSVRWMN); the next 12 values were reserved for validation, and the last 12 values, referring to the year 2014, for testing. The methodology used is multi-step ahead forecasting, that is, h steps ahead; in this application h = 12, therefore the limit of forecasting coincides with the total of steps.

In the evaluation of the method the MAPE (mean absolute percent error), MAE (mean absolute error), and MSE (mean square error) errors were used [11], represented in eq. (1):

$$\mathtt{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y(t)-\hat{y}(t)}{y(t)}\right|;\ \mathtt{MAE} = \frac{1}{T}\sum_{t=1}^{T}|y(t)-\hat{y}(t)|;\ \mathtt{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y(t)-\hat{y}(t))^{2}. \tag{1}$$

The values of the time series from the measurement of the piezometer instrument, denoted by $\left(y_{ps}\left(t \right) \right)_{t=1}^{252}$, are given in meters above sea level (MSNM) and are presented in Fig. 1.

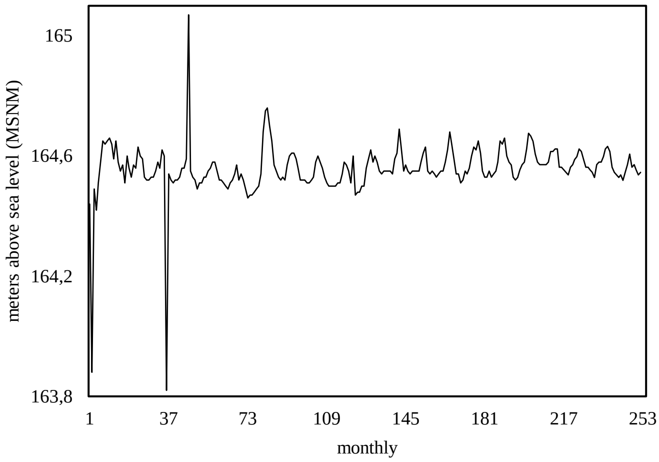

Figure 1. Original series of the piezometer

The list of forecasts, within and without the sample, (multi-step predictions) consists of the results of the method Multiple Kernel SARIMA Support Vector Regression Wavelet (SSVRWMN). The SSVRWMN method is given in six steps [3].

In stage 1, for the piezometer time series, the modeling was carried out through a SARIMA model, generating predictions within the training sample and predictions in the test sample on a horizon of 12 steps ahead, these predictions are considered by the hybrid method as the linear component prediction, represented of $\left(\hat{L}_{ps}\left(t \right) \right)_{t=12}^{240+12}$. For the SARIMA modeling, the Eviews software was used. The graphic analysis of the time series was made and the appropriate model was identified. The definition of the orders of models was done through the analysis of the charts of the self-correlation functions FAC, partial self-correlation FACP, residues, and tests with several options of orders p, d, q.

The statistics found through the analysis of the piezometer time serie are represented in Tab. 1.

Table 1. SARIMA modelling for time serie.

Variable	Coefficient	Std. Error	t-Statistic	Prob
AR(1)	0.589922	0.044582	13.23221	0.0000
AR(10)	0.410084	0.044583	9.198260	0.0000
MA(1)	-0.262589	0.077803	-3.375077	0.0009
MA(10)	-0.959577	0.008765	-109.4830	0.0000
MA(11)	0.270075	0.076805	3.516387	0.0005
R-squared	0.401165	Mean dependent var		164.5615
Adjusted R-squared	0.390519	S.D. dependent var		0.076879
S.E. of regression	0.060019	Akaike info criterion		-2.766822
Sum squared resid	0.810505	Schwarz criterion		-2.692081
Log likelihood	323.1845	Hannan-Quinn criter.		-2.736673
Durbin-Watson stat	1.971314			

In stage 2 the orthogonal decomposition Wavelet of $r = 2$ level was made for the time series $y_{ps}\left(t \right)$, considering the daubechies bases (db 1,2,...,45), coiflets (coif 1,2,…….,5), symlets (sym 1,2,……30) and bi-orthogonal (1.1, 1.3,….. 6.8), a total of 95 decompositions, generating an approximation component that is the representation of the original series in low frequency, and two detail components which are high frequency series (more noisy series) in each decomposition carried out, thus obtaining 285 sub-series for $y_{ps}\left(t \right)$ time serie.

In stage 3 forecasts through Multiple Kernel Support Vector Regression, were carried out for each of the 285 sub-series obtain in stage 2. The RBF, polynomial, and sigmoidal kernels considered [12].

For example, for the time series of the piezometer $\left(y_{db1\_ps}(t)\right)_{t=1}^{240}$ the forecasts, were generated, represented in eq. (2):

$$\begin{aligned}
\left(y_{db1_ps_RBF}(t)\right)_{t=1}^{240+12} &= \left(\hat{y}_{db1_ps_RBFA_2}(t)\right)_{t=12}^{240+12} + \left(\hat{y}_{db1_ps_RBFD_2}(t)\right)_{t=12}^{240+12} + \left(\hat{y}_{db1_ps_RBFD_3}(t)\right)_{t=12}^{240+12} \\
\left(y_{db1_ps_Poly}(t)\right)_{t=1}^{240+12} &= \left(\hat{y}_{db1_ps_PolyA_2}(t)\right)_{t=12}^{240+12} + \left(\hat{y}_{db1_ps_PolyD_2}(t)\right)_{t=12}^{240+12} + \left(\hat{y}_{db1_ps_PolyD_3}(t)\right)_{t=12}^{240+12} \\
\left(y_{db1_ps_Sigm}(t)\right)_{t=1}^{240+12} &= \left(\hat{y}_{db1_ps_SigmA_2}(t)\right)_{t=12}^{240+12} + \left(\hat{y}_{db1_ps_SigmD_2}(t)\right)_{t=12}^{240+12} + \left(\hat{y}_{db1_ps_SigmD_3}(t)\right)_{t=12}^{240+12}
\end{aligned} \quad (2)$$

Being the forecast of the series $\left(y_{db1\_ps}(t)\right)_{t=1}^{240}$, the average of the forecasts, represented in eq. (3):

$$\left(\hat{y}_{db1\_ps}(t)\right)_{t=12}^{240+12} = \frac{1}{3}\left(\hat{y}_{db1\_ps\_RBF}(t)\right)_{t=12}^{240+12} + \frac{1}{3}\left(\hat{y}_{db1\_ps\_Poly}(t)\right)_{t=12}^{240+12} + \frac{1}{3}\left(\hat{y}_{db1\_ps\_Sigm}(t)\right)_{t=12}^{240+12}. \quad (3)$$

This process was repeated for all 95 Wavelet decompositions of each time series.

To obtain the optimum parameters of the SVR model, the data from the time serie reserved for adjustment were subdivided into training and validation. The model that presented the lowest Mean squared error (MSE) in the validation set was selected, being the cardinality of the validation set was equal to 12.

Table 2 shows the two best results obtained for each Wavelet family and the MSE statistics of the validation sample, for the time series of the y_{ps} piezometer.

Table 2. Two best results for each Wavelet family for series y_{ps}.

Families	Base	Validation MSE
Daubechies	**db10**	**0,00006624**
	db19	0,00006864
Symlets	**sym12**	**0,00006288**
	sym15	0,00006595
Coiflets	**coif3**	**0,00013086**
	coif5	0,00013105
Biorthogonal	bior3.5	0,00008236
	bior3.9	**0,00005207**

From Tab. 2 one can observe that the selected modeling is those that presented the lower MSE statistics. For the time serie of the piezometer, are the basis db10, sym12, coif3, and bior3.9.

In stage 4 were combined the forecasts of the best modelings were obtained when the MSE statistics were compared in the validation set regarding each Wavelet family.

For the time series of the piezometer y_{ps}, the combination is represented by: $\left(\hat{y}_{ps}(t)\right)_{t=12}^{240} = \left[\rho_1 \times \left(\hat{y}_{db10\_ps}(t)\right) + \rho_2 \times \left(\hat{y}_{sym12\_ps}(t)\right) + \rho_3 \times \left(\hat{y}_{coif3\_ps}(t)\right) + \rho_4 \times \left(\hat{y}_{bior3.5\_ps}(t)\right)\right]_{t=12}^{240}$, where the adaptive constants assume the following values: $\rho_1 = 0,866198$, $\rho_2 = 0,94482$, $\rho_3 = 0,746553$ and $\rho_4 = -1,547557$, after the minimization of the average quadratic error (MSE).

In this stage, the forecasts of the multiple kernels of the model SVR Wavelet are obtained, that is, of the non-linear component of the modeled time series and which is represented by $\left(\hat{N}_{ps}(t)\right)_{t=12}^{240}$.

In stage 5, the forecasts obtained through the linear model (SARIMA) and through the non-linear model (SVRWMN) are taken up and the mean is calculated to obtain the one-off forecasts for each t, for each of the modeled time series.

The hybrid SSVRWMN method presents in its structure of self-dependence linear and non-linear information and is considered as a filtered version of the time series $y(t)$ both through a linear filter (stage 1) and a non-linear filter (stages 3 and 4).

The list of forecasts for the time series of the piezometer is $\left(\hat{y}_{ps}(t)\right)_{t=12}^{240} = \frac{1}{2}\left(\hat{L}_{ps}(t) + \hat{N}_{ps}(t)\right)_{t=12}^{240+12}$.

At this stage, it is possible to compare the forecasts of the SSVRWMN method with those obtained through the SARIMA, SVR method, and the combination SARIMA-SVR. The compared adherence statistics are MSE, MAPE, and MAE.

It is observed from Tab. 3 that the forecasts of the piezometer time serie of via the method SSVRWMN showed a reduction in the three adherence statistics MSE, MAPE, and MAE among all other predictive methods listed. If compared with the second-best result (namely, SVR method) the relative reduction was approximately 39.1% in the accuracy measure of MSE.

Table 3. Comparison MSE, MAPE and MAE for the series y_{ps} multi step h=12

Methods	MSE	MAPE	MAE
SARIMA	0,00217	0,0233%	0,03841
SVR	0,00140	0,0188%	0,03107
SARIMA-SVR	0,00171	0,0208%	0,03435
SSVRWMN	**0,00085**	**0,0148%**	**0,02446**

Such results demonstrate the predictive efficiency of the method SSVRWMN against renowned approaches in the time series literature, such as SARIMA, SVR, and the hybrid SARIMA-SVR, in the present case.

In step 6, after obtaining the series of residues, the residual Bootstrap process was used to obtain the confidence interval (CI) at the 95% level.

Figure 2 shows the original piezometer series for the analyzed interval, the predictions obtained via the SSVRWMN method, and the respective confidence intervals, obtained through the hybrid method SSVRWMN – bootstrap.

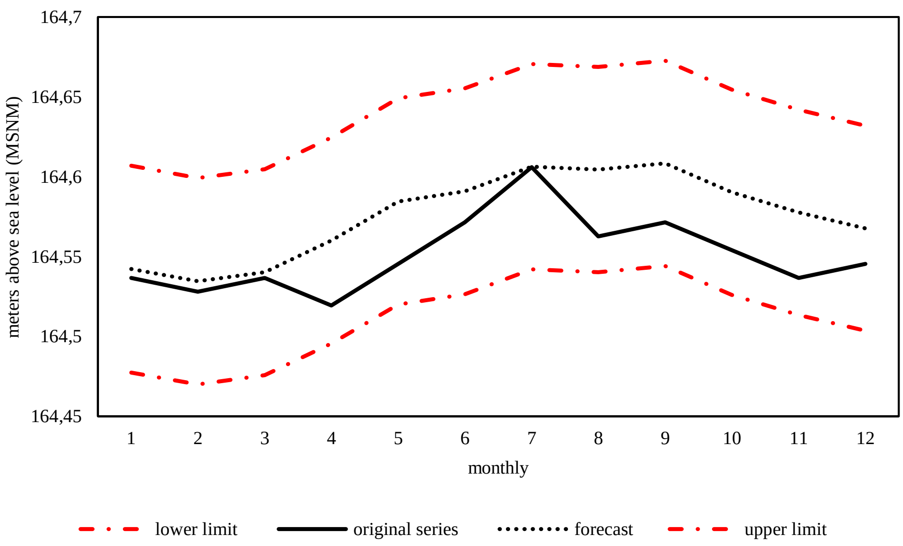

Figure 2. Original series, predictions via the SSVRWMN method, and confidence interval of the piezometer series

3 Conclusions Final

This article presented predictions for the piezometer instrument located in a key block of the dam of the Itaipu hydroelectric plant. The SSVRWMN method is an aggregator of linear and non-linear information, having demonstrated predictive efficiency in data from dam instrumentation.

It is noteworthy that the future forecasts and confidence intervals obtained by the SSVRWMN method are aggregators of valuable information for inspections, as well as decision making, for technicians and engineers responsible for the structural health monitoring of the dam.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] ITAIPU: Usina Hidrelétrica. Projeto: Aspectos de Engenharia. Foz do Iguaçu - Pr, 2009.
[2] C. I. Osako. A manutenção dos drenos nas Fundações de Barragens o caso da Usina Hidrelétrica de Itaipu, 2002. Dissertação Curso de Pós-Graduação em Construção Civil, Universidade Federal do Paraná.
[3] S. B. Rodrigues. Método Híbrido Iterativo Sarima Support Vector Regression Wavelet De Múltiplos Núcleos Na Previsão De Séries Temporais De Instrumentos De Barragens, 2015. Univerdidade Federal do Paraná.
[4] G. Box, G. Jenkins and G. Reinsel. Time Series Analysis: Forecasting and Control. 2008.
[5] I. Daubechies. Ten lectures on wavelets. Society for Industrial and Applied Mathematics, 1992
[6] V. N. Vapnik. The Nature of Statistical Learning Theory. 1995.
[7] R. T. Clemen. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 1989.
[8] B. Efron. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics, v. 7, n. 1, p. 1–26, 1979. Institute of Mathematical Statistics.
[9] R. A. Tibshirani. Comparison of Some Error Estimates For Neural Network Models. Neural Computation, V. 8, P. 152–163, 1996.
[10] L. Pan and D. N. Politis. Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. Journal of Statistical Planning and Inference, 2014.
[11] J. D. Hamilton. Time Series Analysis. Princeton University Press, 1994.
[12] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. Statistics and Computing, p. 199–222, 2004.