Author's Accepted Manuscript

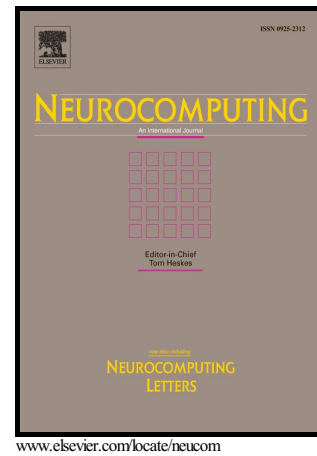## Time series forecasting with the WARIMAX-GARCH method

J.M. Corrêa, A.C. Neto, L.A. Teixeira Júnior, E.M.C. Franco, A.E. Faria Jr

Time series forecasting with the WARIMAX-GARCH method

J. M. Corrêa[a*], A. C. Neto[b], L. A. Teixeira Júnior[c], E. M. C. Franco[d], A. E. Faria Jr[e]

[a]Department of Matematics, Federal Technological University of Paraná, Medianeira, Paraná, Brazil.
[b]Department of Statistic, Federal University of Paraná, Curitiba, Paraná, Brazil.
[c]Latin American Institute of Technology, Infrastructure and Territory, Federal University of Latin America Integration, Foz do Iguaçu, Paraná, Brazil.
[d]Department of Electrical Engineering, State University of West Paraná, Foz do Iguaçu, Paraná, Brazil.
[e]Department of Mathematics and Statistics, The Open University, Milton Keynes, Buckinghamshire, England.

Abstract

It is well-known that causal forecasting methods that include appropriately chosen Exogenous Variables (EVs) very often present improved forecasting performances over univariate methods. However, in practice, EVs are usually difficult to obtain and in many cases are not available at all. In this paper, a new causal forecasting approach, called Wavelet Auto-Regressive Integrated Moving Average with eXogenous variables and Generalized Auto-Regressive Conditional Heteroscedasticity (WARIMAX-GARCH) method, is proposed to improve predictive performance and accuracy but also to address, at least in part, the problem of unavailable EVs. Basically, the WARIMAX-GARCH method obtains Wavelet "EVs" (WEVs) from Auto-Regressive Integrated Moving Average with eXogenous variables and Generalized Auto-Regressive Conditional Heteroscedasticity (ARIMAX-GARCH) models applied to Wavelet Components (WCs) that are initially determined from the underlying time series. The WEVs are, in fact, treated by the WARIMAX-GARCH method as if they were conventional EVs. Similarly to GARCH and ARIMA-GARCH models, the WARIMAX-GARCH method is suitable for time series exhibiting non-linear characteristics such as conditional variance that depends on past values of observed data. However, unlike those, it can explicitly model frequency domain patterns in the series to help improve predictive performance. An application to a daily time series of dam displacement in Brazil shows the WARIMAX-GARCH method to remarkably outperform the ARIMA-GARCH method, as well as the (multi-layer perceptron) Artificial Neural Network (ANN) and its wavelet version referred to as Wavelet Artificial Neural Network (WANN) as in [1], on statistical measures for both in-sample and out-of-sample forecasting.

Keywords

Wavelet decomposition, Exogenous variable, ARIMA-GARCH model, Forecasting.

List of adopted acronyms

- ARIMA: Auto-Regressive Integrated Moving Average.

- ARIMAX: Auto-Regressive Integrated Moving Average with eXogenous variables.

- GARCH: Generalized Auto-Regressive Conditional Heteroscedasticity.

- ARIMAX-GARCH: Auto-Regressive Integrated Moving Average with eXogenous variables and Generalized Auto-Regressive Conditional Heteroscedasticity.

- EVs: Exogenous Variables.

- WCs: Wavelet Components.

- WEVs: Wavelet Exogenous Variables.

- WARIMAX-GARCH: Wavelet Auto-Regressive Integrated Moving Average with eXogenous variables and Generalized Auto-Regressive Conditional Heteroscedasticity.

- ANN: Artificial Neural Network.

- WANN: Wavelet Artificial Neural Network.

1. Introduction

There is a vast body of literature on methods and techniques for the modeling and forecasting of time series (see e.g. [2]). One of the most well-known is the class of Auto-Regressive Integrated Moving Average (ARIMA) models proposed by [3] for stationary time series exhibiting linear auto-dependence characteristics. An ARIMA model makes use of available historical data from the underlying time series, denoted by y_t $(t = 1, ..., T)$, to quantify any auto-regressive and moving-average patterns and produce forecasts. Many time series are often affected or influenced by certain external factors such as special events (e.g. legislative activities, policy changes, environmental regulations), as well as by uncertain or random events (referred to as stochastic events), that generate data that may be available to be used as Exogenous Variables (EVs). Some models can accommodate one or more such variables to help improving the forecasting process. Box and Jenkins themselves proposed an extension to an ARIMA model, the transfer function model (see e.g. [2]), which can account for EVs. The general exogenous model employed by the ARIMA model has been discussed by [4], where it is referred to as an Auto-Regressive Integrated Moving Average with eXogenous variables (ARIMAX) model. [5] refers to the ARIMAX model as dynamic regression model. Both [4] and [5] show increased forecasting accuracy gains are achieved when EVs are properly used in the modeling process. ARIMAX models, however, require adequate EVs are available and that the underlying time series is stationary (or transformed into a stationary one).

In the conventional ARIMAX model, the conditional variance of its innovations is typically supposed to be constant (homoscedasticity). However, many time series often exhibit periods of unusual high volatility followed by periods of relative stability. In such situations, the constant conditional variance assumption may be considered inappropriate. In order to account for changes in conditional variance (heteroscedasticity), [6], [7], [8], amongst others, developed a class of conditional heteroscedastic models that allows the changing conditional variance (volatility) of a time series to be explicitly modelled. The Auto-Regressive Conditional Heteroscedastic (ARCH) model of [6] allows for the conditional variance to depend on past values of the conditional variance itself; while the Generalized ARCH (GARCH) model of [7] enable the volatility to depend on past values of both the squared innovation and conditional variance itself. The WARIMAX-GARCH method proposed here employs a GARCH model as one of its components. Other extensions of a GARCH model are, for example, the GARCH-in-mean (GARCH-M) of Engle, [9] that includes a heteroscedasticity term into the mean model (represented by either the ARIMA or ARIMAX components); and the Exponential GARCH (EGARCH) model of [10], which models the conditional variance in logarithmic form, that does not require non-negativity constraints and allows for asymmetric effect of information on the volatility. In effect, the GARCH term of the WARIMAX-GARCH method can be chosen to be any of such extensions of a GARCH model depending on the application.

In many practical cases, appropriate EVs are just not available to be employed. For instance, [11] claim that, contrary to price-based management, regression methods are somewhat less common in quantity-based revenue forecasting applications (such as airline and hotel revenue management) because it is often difficult to obtain data on the explanatory EVs as an automated data feed. In such cases, the only explanatory variables that can be used must come from the historical underlying data. In this context, the approach adopted in this paper is to generate WEVs from WCs of the time series to be used as EVs by the WARIMAX-GARCH method as we shall see. This way, it is possible to obtain improved forecasting performance gains similarly to the ones obtained by [5] when using the conventional ARIMAX-GARCH model albeit in another application as we shall see.

Note that, in practical terms, the Wavelet Exogenous Variables (WEVs) can be seen as representing the quantified frequency patterns present in the time series that the usual ARIMAX model does not account for. Now, mathematically, the WEVs consist of the wavelet components (the WCs), where a WC is defined by orthogonal projections of an original time series on orthogonal complete subspaces, called the "wavelet subspaces", of the l^2 space.

Thus, the WARIMAX-GARCH method consists of a new causal forecasting method that, based on ARIMAX-GARCH models, generates and uses the WCs as exogenous variables (the WEVs). Figure 1 shows a flow diagram with seven levels depicting the steps, described below, of the WARIMAX-GARCH method. To apply the WARIMAX-GARCH method, the underlying time series is first split into a training (in-sample) and a testing (out-of-sample) sample. The training sample, denoted by y_t ($t = 1, ..., T$), where T is a conveniently chosen time period, is decomposed via wavelet decomposition of level r as shown by the top two levels of the diagram in Figure 1. This decomposition produces (for each time $t = 1, ..., T$) one WC of approximation at level m_0, denoted by $\tilde{y}_{A_{m_0},t}$, and r WCs of detail at levels m_0, $m_0 + 1$, ..., $m_0 + (r - 1)$, denoted by $\tilde{y}_{D_{m_0},t}$, $\tilde{y}_{D_{m_0+1},t}$, ..., $\tilde{y}_{D_{m_0+(r-1)},t}$, respectively (as represented by the third level at the top of Figure 1). Each WC is then individually modeled by using an ARIMA-GARCH model, generating h-steps-ahead out-of-sample forecasts to its level. The $r + 1$ WCs of y_t ($t = 1, ..., T$) with their out-of-sample forecasts are the completed WCs and are denoted by $x_{1,t}$, $x_{2,t}$, $x_{3,t}$, ..., $x_{r+1,t}$ ($t = 1, ..., T, T + 1, ..., T + h$). Those are interpreted here as EVs ($WEV_1, ..., WEV_{r+1}$ in Figure 1) in an ARIMAX-GARCH model for forecasting the time series y_t ($t = 1, ..., T$), both in-sample and out-of-sample, to its level (conditional mean) and conditional variance. Finally, under Gaussian assumption, the respective predictive intervals are trivially calculated.
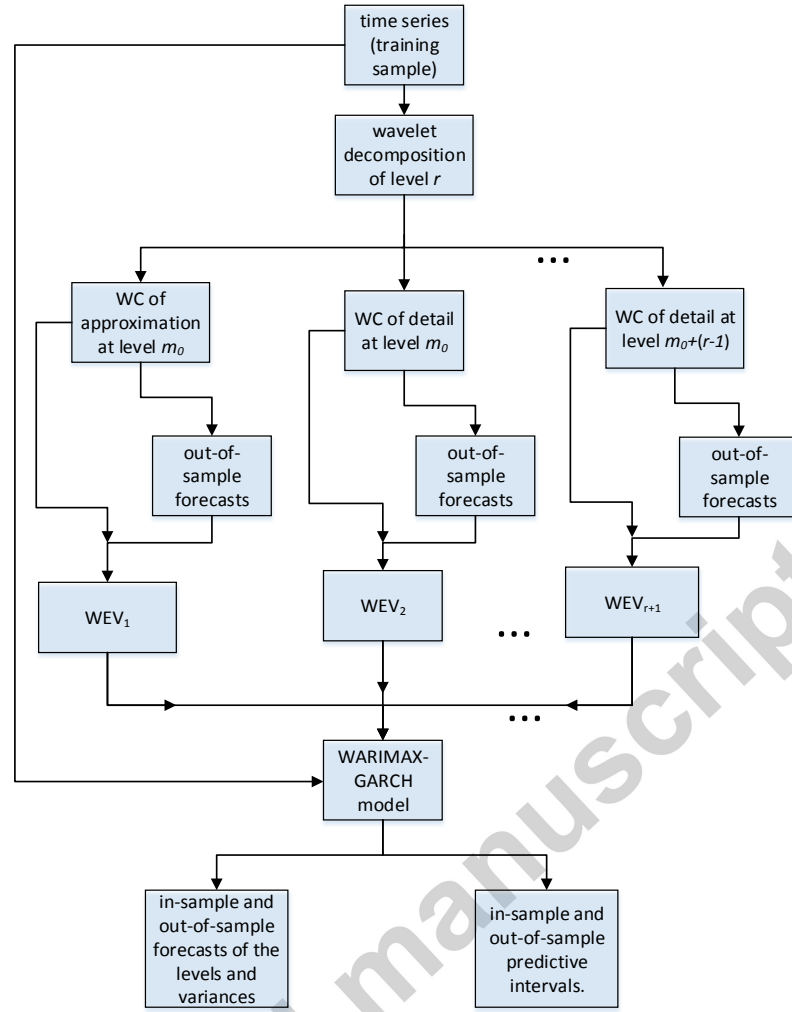
Figure. 1. Flowchart of the WARIMAX-GARCH method.

This paper is structured as follows. In Section 2, a more in depth review of the wavelet decomposition method and the ARIMAX-GARCH model, which are component parts of the proposed WARIMAX-GARCH method, are presented, as well as the ANN and WANN models used here as benchmark methods; Section 3 formally defines the WARIMAX-GARCH method; while Section 4 shows the main numerical results of the application to a real time series of dam displacement at the Itaipu hydroelectric, in southern Brazil. Finally, Section 5 concludes the paper.

2. Oriented review of the literature

The purpose of this section is to present a brief review of the concepts and methods required to define the WARIMAX-GARCH method that is introduced in Section 3. It also reviews two neural network methods – a conventional Artificial Neural Network (ANN) and the Wavelet ANN (WANN) proposed by [1] – that are used as benchmark methods for comparing forecasting performances in the application described in Section 4.

It starts in Subsection 2.1 by describing the wavelet decomposition of level r, which is the algorithm adopted in initial step of the WARIMAX-GARCH method. This is followed in Subsection 2.2 by the introduction of the ARIMAX-GARCH models that are posteriorly employed to generate out-of-sample forecasts associated with the wavelet components (WCs)

that are used as exogenous variables. Finally, Subsections 2.3 and 2.4 describe the ANN and the WANN forecasting methods, respectively.

2.1 Wavelet decomposition of level r

Let l^2 be the collection of all scalar-valued complex infinite sequences $y_t (t \in \mathbb{Z})$ in $\mathbb{C}^{\mathbb{Z}}$ such that $\sum_{t \in \mathbb{Z}} |y_t|^2 < \infty$, where $\mathbb{C}$ and $\mathbb{Z}$ denote, respectively, the sets of all complex and integer numbers; and assume that the function $\langle\ ;\ \rangle$ of l^2 into $\mathbb{C}$ is the usual inner product (as in [12]). An element $\omega(.) \in l^2$ – with $\langle\ ;\ \rangle$ – is a l^2 – wavelet function, if the sequence $\omega_{m,n}(.) := 2^{\frac{m}{2}}\omega(2^m(.) - n)$, where $(n, m) \in \mathbb{Z}x\mathbb{Z}$, forms an orthonormal basis for l^2; and a member $\phi(.) \in l^2$ – with $\langle\ ;\ \rangle$ – consist of a l^2-scaling function, if the sequence $\phi_{m,n}(.) := 2^{\frac{m}{2}}\phi(2^m(.) - n)$, where $(m, n) \in \mathbb{Z}x\mathbb{Z}$, holds: $\langle\phi_{m',n'}(.); \phi_{j,k}(.)\rangle \neq 0$, whenever $m' = j$ and $n' \neq k$; and, if otherwise, $\langle\phi_{m',n'}(.); \phi_{j,k}(.)\rangle \neq 0$, where $m', n', j, k \in \mathbb{Z}$ (see e.g. [13] and [14]). Based on [15], it can be seen that the subset $\{\omega_{m,n}(.)\}_{\substack{(m,n\in\mathbb{Z}\\m\geq m_0}} \cup \{\phi_{m_0,n}(.)\}_{n\in\mathbb{Z}}$ of l^2-wavelet and l^2-scaling functions, where m_0 is a fixed integer value, is, in fact, an orthonormal basis for l^2. Accordingly, any sequence $y_t (t \in \mathbb{Z})$ in l^2 can be orthogonally decomposed, in terms of an orthonormal basis $\{\omega_{m,n}(.)\}_{\substack{(m,n\in\mathbb{Z}\\m\geq m_0}} \cup \{\phi_{m_0,n}(.)\}_{n\in\mathbb{Z}}$, as represented by

$$y_t = y_{A_{m_0},t} + \sum_{m=m_0}^{+\infty} y_{D_m,t}, \tag{1}$$

where $y_{A_{m_0},t} := \sum_{n\in\mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(.)$ is the WC of approximation at level m_0 of the state y_t, with $a_{m_0,n} := \langle y_t; \phi_{m_0,n}(.)\rangle$; and $y_{D_m,t} := \sum_{n\in\mathbb{Z}} d_{m,n} \omega_{m,n}(.)$ is the WC of detail at level m of y_t, with $d_{m,n} := \langle y_t; \omega_{m,n}(.)\rangle$. The orthogonal decomposition in (1) is usually called a wavelet decomposition.

Tautologically, any finite (scalar-valued complex) time series $y_t (t = 1, ..., T)$ can be interpreted as an infinite sequence $y_t (t \in \mathbb{Z})$ in l^2, defined as $y_t = y_t$, if $t \in \{1, 2, ..., T\}$; and $y_t = 0$, if $t \in \mathbb{Z} - \{1,2,...T\}$. Therefore, any finite time series $y_t (t = 1, ..., T)$ can be orthogonally decomposed by the wavelet decomposition in (1).

In practical terms, once it is impossible to model separately all WCs generated by the expansion (1), an adaptation is required to obtain a finite number of components. Thus, according to [16], a good alternative may occur thoroughly by means of the wavelet decomposition of level r of $y_t (t = 1, ..., T)$, where $r > 0$ and $r \in \mathbb{Z}$, is given by

$$y_t = \tilde{y}_{A_{m_0},t} + \sum_{m=m_0}^{m_0+(r-1)} \tilde{y}_{D_m,t} + \varepsilon_t, \tag{2}$$

where m_0 is the level parameter (which is often assumed to be equal to r); ε_t is the approximation error term, that is, the difference between the state y_t and its (wavelet) approximation $\tilde{y}_{A_{m_0},t} + \sum_{m=m_0}^{m_0+(r-1)} \tilde{y}_{D_m,t}$ (nevertheless, in practice, it is usually assumed that ε_t is equals zero); $\tilde{y}_{A_{m_0},t} := \sum_{n=0}^{2^{M-m_0}-1} a_{m_0,n} \phi_{m_0,n}(t)$ and $\tilde{y}_{D_m,t} := \sum_{n=0}^{2^{(M-m_0)}-1} d_{m,n} \omega_{m,n}(t)$, which are, respectively, WCs of approximation at level m_0 and of detail at level m and consist, respectively, of the approximations to $y_{A_{m_0},t}$ and $y_{D_m,t}$, in (1); and M is a parameter that takes an integer value such that $T = 2^M$. If T is not an integer power of 2, the sequence $y_t (t = 1, ..., T)$ is usually completed with zeros such that its length T is increased up to the

next integer power of 2. This procedure may be carried out because the zeros added up do not affect the calculation of the WCs $\tilde{y}_{A_{m_0},t}$ and $\tilde{y}_{D_m,t}$ generate in (2) (see e.g. [17]), preserving the auto-correlation y_t and its components, in (2), for all t, where $t = 1, \ldots, T$.

After obtaining the $r + 1$ WCs in (2), that is, $\tilde{y}_{A_{m_0},t}$ and $\tilde{y}_{D_m,t}(m = m_0, \ldots, m_0 + (r - 1))$, they are individually modeled by an adequate ARIMA-GARCH model in order to produce their out-of-sample forecasts. Finally, the mentioned forecasts complete the WCs in (2), providing the WEVs to be used as exogenous variables by ARIMAX-GARCH models as described in the following section.

Note that the WEVs were not modeled individually by ARIMA models as those models cannot map non-linear auto-dependence that is often present in the WCs. Further to that, the WCs of detail usually present conditional heteroscedasticity (as is the case of the time series modeled in this paper). Therefore, ARIMA-GARCH models have been chosen for mapping both linear and non-linear effects in the WCs and obtain more accurate in-sample and out-of-sample forecasts of the WCs, in particular of the detail WCs. Consequently, the in-sample and out-of-sample forecasts that can be interpreted as aggregators of information from different sources (namely, the different $r + 1$ ARIMA-GARCH models used to model the WCs) to the original time series will contain both linear and non-linear information.

2.2 The ARIMAX-GARCH model

Let y_t $(t = 1, \ldots, T)$ be a stationary time series (or a non-stationary time series that can be transformed into a stationary one) that exhibits linear auto-dependence. Also, assume $\left((x_{1,t})_{t=1}^T, \ldots, (x_{(r+1),t})_{t=1}^T \right)$ denotes a list of $r + 1$ vectors of realizations from $r + 1$ stationary exogenous variables of y_t $(t = 1, \ldots, T)$. Based on [4] and [5], each realization y_t can be represented by an ARIMAX (p, d, q) model, with the mathematical formulation:

$$\nabla^d y_t = \phi_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=i}^q \theta_i\, e_{t-i} + \sum_{i=1}^{r+1} \sum_{l_i=0}^{L_i} \alpha_{il_i}\, x_{i,t-l_i} + e_t, \qquad (3)$$

where B is the backward operator defined by $B^k y_t \coloneqq y_{t-k}$, with k belonging to $\mathbb{Z}$; $\nabla^d \coloneqq (1 - B)^d$ is the difference operator, with d representing its order; $(\phi_i)_{i=0}^p$ and $(\theta_j)_{j=1}^q$ are the ordered lists of model complex parameters, with $\phi_p \neq 0$ and $\theta_q \neq 0$, and $(\alpha_{il_i})_{l_i=0}^{L_i}$ is the ith ordered list of complex model parameters associated with the exogenous component (these three lists of parameters need to satisfy both the invertibility and the stationarity conditions - see e.g. [18] and [19]); e_t is an innovation consisting of a state of the random variable ε_t from an uncorrelated stochastic process with zero mean; p and q are, respectively, the orders of the Auto-Regressive (AR) part $\sum_{j=1}^p \phi_j y_{t-j}$ (AR(p)) and of the Moving Average (MA) part $\sum_{j=i}^q \theta_i e_{t-i}$ (MA(q)); and, L_i is the maximum lag order in the sum $\sum_{l_i=0}^{L_i} \alpha_{il} x_{i,t-l}$ of the exogenous variable $x_{i,t}$, with $i = 1, \ldots, (r + 1)$. Particularly, if $\alpha_{il} = 0$, for all $i = 0, \ldots, (r + 1)$ and all $l_i = 0 \ldots, L_i$, then the model in (3) above becomes a conventional ARIMA (p, d, q). Note that, a SARIMAX (p, d, q) x $(P, D, Q)_S$ model - also known as a multiplicative ARIMAX (p, d, q) model – can be used for modeling a seasonal time series (as in [18]). It generically consists of an ARIMAX model with seasonal components (please, see Section 3 for details).

In equation (3) above, it is assumed that the innovation term e_t is a realization of an uncorrelated random variable ε_t, with zero mean and constant conditional variance, that is, $\sigma_t^2 = \sigma^2$, for all t. However, many time series do not satisfy this stationarity condition (called

homoscedasticity). In some cases, the changing conditional variance (volatility) may depend on past squared innovations of the time series or past values of the variance itself such that an ARMA structure, as well as an extension of it, can be adopted for temporally projecting the volatility. This way, unconditionally, the variance is constant, but conditional on past values it is allowed to change in time. According to [20], the general GARCH (R, S) model for the conditional variance σ_t^2 of the innovation e_t is given by

$$\sigma_t^2 = \gamma_0 + \sum_{j=1}^{R} \gamma_j \, \sigma_{t-j}^2 + \sum_{i=1}^{V} \delta_i \, e_t^2, \tag{4}$$

where the following constraints must hold: $\sum_{j=1}^{R} \gamma_j + \sum_{i=1}^{V} \delta_i < 1$; $\gamma_0 > 0$; $\gamma_j \geq 0$ $(j = 1, \dots, R)$ and $\delta_i \geq 0$ $(i = 1, \dots, V)$. Equation (4) is used by the WARIMAX-GARCH method to generate in-sample and out-of-sample forecasts of the conditional variance of the time series y_t $(t = 1, \dots, T)$. In addition, σ_t^2 multiplied by a parameter φ can be used in (3) to account for a non-linear effect to construct the forecasting of its level (conditional mean). Notice that the generating mechanism for a GARCH innovation process e_t $(t = 1, \dots, T)$ is defined by $e_t := \sigma_t \, z_t$, where: z_t is a realization of a standardized, independent and identically distributed random variable Z_t; and σ_t is the conditional standard deviation.

A model compounded by (3) and (4), which accounts for the above mentioned constraints, is called an ARIMAX-GARCH model. In order to obtain the best possible ARIMAX-GARCH model, three basic steps should be carried out: (i) test the plausible values for the parameters

p, d, q and L_i, in (3), as well as the parameters R and V in (4) (which can be obtained through the profile analysis of the plots of simple and partial auto-correlation functions of the ordinary standard innovations z_t $(t = 1, \dots, T)$ and the squared standard innovations z_t^2 $(t = 1, \dots, T)$, as described by [20] and [6]; (ii) define the method to be used to estimate the ARIMAX-GARCH parameters - the most common is the Maximum Likelihood Estimation (MLE) method (as in [18]); and, (iii) make a diagnostic check to choose the most parsimonious and adequate model to be used for generating both the in-sample and the out-of-sample forecasts of y_t $(t = 1, \dots, T)$ and their volatilities.

It is worth pointing out that similarly to the way that the h-steps ahead out-of-sample forecasts of conditional means are produced (for instance, through an estimated ARIMAX model), the h-steps ahead forecasts of conditional variances (volatilities) are generated by means of an estimated GARCH model (or an extension of it). For more details, please see [18], wherein a GARCH-in-mean model produces out-of-sample forecasts of volatilities and employs them in the construction of the forecasts for the corresponding conditional means. In fact this approach aims to aggregate non-linear information (coming from the squared auto-dependence exhibited by a given time series) from the forecasts of the conditional means in order to increase its predictive accuracy. Note that the WARIMAX-GARCH method proposed here can adopt any GARCH approach to forecast volatilities.

2.3 Artificial Neural Networks

Artificial Neural Networks (ANNs) are well known to be flexible computing frameworks for modeling and forecasting a broad range of stochastic time series exhibiting either linear or non-linear auto-dependence structures. Contrary to many linear statistical forecasting models, stationarity is not required by ANN methods (see e.g. [1]). Another important aspect of ANNs is that they are universal approximates of compact (i.e., closed and bounded) support functions, as showed in [21]. In effect, since observations from a time series $\mathbf{y_t}$ $(\mathbf{t = 1, \dots, T})$ that exhibit dependency on past values may be seen as points of the domain of an unknown

compact support function, it follows that the ANNs are capable of approximating them (for modeling or forecasting) with a high degree of accuracy. According to [22], the predictive power of ANNs comes from the parallel processing of the information exhibited by the data. In addition, AAN models are largely determined by the stochastic characteristics inherent in the time series.

In this context, the feed-forward multi-layer perceptron ANNs (see e.g. [23]) are the most widely used neural prediction models for time series forecasting. Particularly, a single hidden layer ANN (henceforth, for simplicity, referred to as ANN) is characterized by an artificial network composed by three layers (namely, input, hidden an output layers) of simple processing units numerically connected by acyclic links. The relationship between the output y_t and the L-lagged inputs, y_{t-k} ($k = 1, \dots, L$), has the following mathematical representation

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j\, g\big(\beta_{0j} + \sum_{i=1}^{L} \beta_{ij}\, y_{t-i}\big) + \varepsilon_t, \qquad (5)$$

where α_j ($j = 0, 1, \dots, q$) and β_{ij} ($i = 0, 1,, \dots, L;\ j = 0, 1, \dots, q$) are the ANN parameters, called connection weights; L is the number of input nodes; q is the number of hidden nodes; ε_t is the approximation error at time t; and $g(.)$ is the transfer function, here, a logistic function - although it is possible to adopt other functions (see e.g [23]). The logistic function is widely employed as the hidden layer transfer function in neural network forecasting and is mathematically defined by

$$g(x_t) = \frac{1}{1+exp(-x_t)}, \qquad (6)$$

where $x_t := \beta_{0j} + \sum_{i=1}^{L} \beta_{ij}\, y_{t-i}$ and $exp(.)$ is the exponential function with Euler's basis (as in [23]). Due to $g(.)$ being a non-linear transfer function, the ANN model in (5), in fact performs a non-linear mapping of the past observations y_{t-k} ($k = 1, \dots, L$) to produce a forecast for y_t. In general, the model in (5) can be rewritten, as

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-L}, w) + \varepsilon_t, \qquad (7)$$

where w denotes a vector of all ANN parameters and $f(y_{t-1}, y_{t-2}, \dots, y_{t-L}, w)$ is the model determined by the network structure and the connection weights in (5). Note that the neural network as defined above is equivalent to a non-linear auto-regressive model.

In practice, w is unknown and hence needs to be numerically determined. So, in order to find its estimated value, $\hat{w}$, that accounts for some criteria (an objective function), an optimization algorithm is applied to training data. Although there are several methodologies available in the literature, we adopt, in line with [23], the Levenberg-Marquardt's algorithm. The minimization of the in-sample sum of squared errors (i.e., $min_w \sum_{t=1}^{T} \varepsilon_t^2$) is the numerical criteria of this ANN. Once $\hat{w}$ is determined we have that

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-L}, \hat{w}) + \hat{\varepsilon}_t \qquad (8)$$

with $f(y_{t-1}, y_{t-2}, \dots, y_{t-L}, \hat{w}) = \hat{y}_t$ consisting of the optimized ANN outcome at time t, which is the forecast of the state y_t, and $\hat{\varepsilon}_t$ being the forecasting error of $\hat{y}_t$.

2.4 Wavelet Artificial Neural Networks

There is currently a number of distinct wavelet decomposition methods combined with ANNs (referred to as wavelet ANNs) that are used for time series forecasting and achieve

remarkable accuracy gains. In fact, there are several studies showing the predictive accuracy gains achieved by wavelet ANN methods such as in [24], [25], [26] and [1] amongst others. Nevertheless, due to the remarkable results achieved in their experiments involving the forecast of the non-stationary solar radiation time series, the [1]'s wavelet ANN method (henceforth referred to as wavelet ANN method) was chosen to be used as a benchmark method in this paper.

For a time series y_t ($t = 1, \ldots, T$), the wavelet ANN method is carried out in two steps:

Step 1: a wavelet decomposition of level r (as in Section 2.1) of y_t ($t = 1, \ldots, T$) is performed, producing 1 WC of approximation at level m', denoted by $y_{A_{m'},t}$ ($t = 1, \ldots, T$), where $m' \in \mathbb{Z}$, and r WCs of detail at levels m', $m' + 1$, …, $m' + (r - 1)$, denoted by $y_{D_m,t}$ ($t = 1, \ldots, T$), respectively, where $r \in \mathbb{Z}$; and

Step 2: The WCs from Step 1 are simultaneously modelled through the ANN described in Section 2.3 producing forecasts similarly to (8), as follows:

$$y_t = f\left(Y_{A_{m'},t,L}; Y_{D_{m'},t,L}; \ldots; Y_{D_{m'+m'+(r-1)},t,L}; \widehat{w}\right) + \widehat{\varepsilon}_t \qquad (9)$$

where $Y_{A_{m'},t,L} := \left(y_{A_{m'},t}, \ldots, y_{A_{m'},t-K}\right)$ and $Y_{D_m,t,L} := \left(y_{D_m,t}, \ldots, y_{D_m,t-L}\right)$, notice that $m = m'$, $m' + 1$, …, $m' + (r - 1)$ are the input data. Similarly to the ANN in (8), the optimal solution $\widehat{w}$ in (9) is obtained via Levenberg-Marquardt's algorithm such that $min_w \sum_{t=1}^{T} \varepsilon_t^2$ is determined. The forecast of y_t is $\left(Y_{A_{m'},t,L}; Y_{D_{m'},t,L}; \ldots; Y_{D_{m'+m'+(r-1)},t,L}; \widehat{w}\right) = \widehat{y}_t$, and $\widehat{\varepsilon}_t$ is the forecasting error of $\widehat{y}_t$ (please, see [1], for more details).

3. The WARIMAX-GARCH method

The ARIMAX-GARCH model requires that EVs are available to be used. In cases, where EVs are not available, an ARIMA-GARCH model can be used but at the expense of a likely lower predictive accuracy. The WARIMAX-GARCH method defined below aims to, at least in part, address this problem. The general steps, mentioned at the end of Section 1, are described in more detail as the following five steps.

Step 1: a wavelet decomposition of level r (described in Section 2.1) of the underlying time series y_t ($t = 1, \ldots, T$) is performed, generating $r + 1$ WCs. That is, one WC of approximation at level m_0, denoted by $\widetilde{y}_{A_{m_0},t}$($t = 1, \ldots, T$), and r WCs of detail at levels from m_0 to $m_0 + (r - 1)$, denoted by $\widetilde{y}_{D_m,t}$($t = 1, \ldots, T$) for m=m$_0$, …, m$_{0+(r-1)}$;

Step 2: each WC obtained in Step 1 is individually modeled by using a distinct ARIMA-GARCH in order to generate their out-of-sample forecasts;

Step 3: the WCs of the Step 1 are completed by their out-of-sample forecasts (of horizon h) of the Step 2, producing the Completed WCs (CWCs) consisting of the wavelet EVs (WEVs). Algebraically, they are given by the data sets below:

- $\tilde{y}_{A_{m_0},t,C}$ or $x_{1,t}$ $(t = 1, \ldots, T, T+1, \ldots, T+h)$ consist of the CWC of approximation at level m_0 of y_t $(t = 1, \ldots, T)$ and is such that: $\tilde{y}_{A_{m_0},t,C} = \tilde{y}_{A_{m_0},t}$, if $t = 1, \ldots, T$; and $\tilde{y}_{A_{m_0},t,C} = \hat{y}_{A_{m_0},t}$, if $t = T+1, \ldots, T+h$ - where $\hat{y}_{A_{m_0},t}$ represents the out-of-sample forecast at instant t generated by an ARIMA-GARCH in Step 2;

- $\tilde{y}_{D_m,t,C}$ or $x_{i,t}$ $(t = 1, \ldots, T, T+1, \ldots, T+h)$, where i=2,…, r+1, consists of the CWCs of the detail at level m, where $m_0 \leq m \leq m_0 + (r-1)$, of $y_t(t = 1, \ldots, T)$ and is such that: $\tilde{y}_{D_m,t,C} = \tilde{y}_{D_m,t}$, if $t = 1, \ldots, T$; and $\tilde{y}_{D_m,t,C} = \hat{y}_{D_m,t}$, if $t = T+1, \ldots, T+h$ - where $\hat{y}_{D_m,t}$ denotes the out-of-sample forecast at time t produced by an ARIMA-GARCH in Step 2.

Step 4: the $r+1$ CWCs generated in Step 3 are treated as $r+1$ wavelet exogenous variables by the following WARIMAX-GARCH (p, d, q) x (P, D, Q) model to generate in-sample and out-of-sample forecasts to the level and the conditional variance of y_t $(t = 1, \ldots, T)$, as well as their predictive intervals, under the assumption of Gaussian innovations out of sample.

A WARIMAX-GARCH (p, d, q) x (P, D, Q) model is mathematically defined by (10) and (11) below as

$$\phi(B)\, \Phi\big(B^S\big)\, \nabla^d\, \nabla_S^D\, y_t = \varphi_{A_{m_0}}\, \nabla^k\, g\left(\tilde{y}_{A_{m_0},t,C}\right) + \sum_{m=m_0}^{m_0+(r-1)} \nabla^{k_m}\, \tilde{y}_{D_m,t,C}\, \varphi_{D_m,C} +$$

$$\theta(B)\, \Theta(B^S)\, e_t + \varphi\, \sigma_t^2, \tag{10}$$

where the components $\phi(B):=\big(1 - \phi_1 B - \cdots - \phi_p B^p\big)$, $\theta(B) := \big(1 - \theta_1 B - \cdots - \theta_q B^q\big)$, $\nabla^d := (1-B)^d$, $\nabla_S^D := (1-B^S)^D$, $\Theta(B^S) := \big(1 - \theta_1 B^S - \cdots - \theta_Q B^{SQ}\big)$ and $\Phi(B^S):=\big(1 - \Phi_1 B^S - \cdots - \theta_Q B^{SP}\big)$, are the polynomials associated with a conventional SARIMA(p, d, q) x (P, D, Q) model; φ is the parameter linearly associated with the conditional variance σ_t^2; $\varphi_{A_{m_0}}$ and $\varphi_{D_m,C}$ are the parameters linearly associated with $\tilde{y}_{A_{m_0},t,C}$ and $\tilde{y}_{D_m,t,C}$, respectively. In turn, $g(.)$ is either the identity or a logarithmic function - which is very useful, in many cases, to turn the level of a time series constant (as in [18]) as well as to improve forecasting performance. Finally, ∇^k, for $k \geq 0$ and $k \in \mathbb{Z}$, is the difference operator associated with the WEV $\tilde{y}_{A_{m_0},t,C}$

($t = 1, \dots, T$) that (as in [18]) can be used to generate a new time series with constant mean. It is worth pointing out that ∇^d and ∇^k can assume different orders when one is searching for the optimal model.

Concerning the parameters of the WARIMAX part, they require assumptions similar to those required by the parameters of the conventional ARIMAX and are estimated here by Maximum Likelihood method. In order to obtain the model's conditional variance for y_t ($t = 1, \dots, T$), the WARIMAX-GARCH model uses its GARCH (R, V) component, which is, in the strict sense, given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{R} \alpha_i e_{t-i}^2 + \sum_{j=1}^{V} \beta_i \sigma_{t-j}^2, \tag{11}$$

where $e_t := \sigma_t z_t$, with z_t being the realization of a standardized, independent and identically distributed random variable Z_t and σ_t the conditional standard deviation.

Step 5: once the collection of in-sample and out-of-sample forecasts $\hat{y}_t$ ($t = 1, \dots, T, T + 1, \dots, T + h$) of the level of y_t ($t = 1, \dots, T$), as well as of their volatilities $\hat{\sigma}_t^2$ ($t = 1, \dots, T, T + 1, \dots, T + h$) are determined, the prediction intervals $[\hat{y}_{INF,t}; \hat{y}_{SUP,t}]$ ($t = 1, \dots, T, T + 1, \dots, T + h$), with any 1-$\alpha$ level of confidence, under the assumption of Gaussian innovations, can be determined straightforwardly. The inferior limit $\hat{y}_{INF,t}$ and the superior limit $\hat{y}_{SUP,t}$ of the prediction intervals at any time t are, respectively, defined by $\hat{y}_{INF,t} := \hat{y}_t - \hat{\sigma}_t^2 z_{\frac{\alpha}{2}}$ and $\hat{y}_{SUP,t} := \hat{y}_t + \hat{\sigma}_t^2 z_{\frac{\alpha}{2}}$, where $z_{\frac{\alpha}{2}}$ represents a state of a random variable Z_t associated with α.

Note that although wavelet methods are applicable to non-stationary and/or non-linear time series (as in [14]), all WEVs of a WARIMAX-GARCH model should be stationary to satisfy this requirement of the ARIMAX modelling approach. However, non-stationary WEVs formatted by the WC of approximation $\tilde{y}_{A_{m_0},t,C}$ ($t = 1, \dots, T$) and its lagged versions, can be made stationary via differencing with the use of the back shift operator ∇^k and/or by transforming with the mapping $g(.)$ associated with the component, as is usually done in ARIMA modelling when the original series is non-stationary.

Also note that, WEVs consisting of WCs of detail, as well as their stationary versions, are always stationary in level such that a difference operator ∇^{k_m} in (5) is applied just to help achieve improved forecasting performances and/or obtain a plausible model. In fact, according to [14], a wavelet function at level m, $\omega_{m,n}(.)$, consists of a short duration curve which image values have zero mean. Furthermore, a Wavelet Component (WC) of detail at level m, $\tilde{y}_{D_m,t}$ ($t = 1, \dots, T$), is mathematically defined by a linear combination of wavelet functions at level m such that $\tilde{y}_{D_m,t} := \sum_{n=0}^{2^{(M-m_0)}-1} d_{m,n} \omega_{m,n}(t)$, for each time t, where $t = 1, \dots, T$. Similarly to $\omega_{m,n}(.)$, a WC of detail $\tilde{y}_{D_m,t}$ takes values around zero; from the statistical point of view, it means that $\tilde{y}_{D_m,t}$ exhibits stationarity in the conditional mean (level). Now, based on [14], since m is a fixed parameter associated with the spectral frequency of $\tilde{y}_{D_m,t}$, it follows that its conditional variances (or volatilities) have stationary stochastic fluctuation. Therefore, a WC of detail at level m will always be stationary.

4. Empirical results

In this section, the main results of an application of the WARIMAX-GARCH method to a daily time series of the Itaipu dam displacement measures is described. For comparative purposes, ARIMA-GARCH models, as well as an ANN and a Wavelet ANN method (as in [1]) were also applied to the displacement series. In Section 4.1, the series is described and an initial statistical analysis conducted to justify the selected models. In Section 4.2, the ARIMA-GARCH models to be applied to the displacement series are identified. Section 4.3 shows the five basic steps of the applied WARIMA-GARCH method together with the main statistical tests used for its validation. Section 4.4 concludes with a comparative analysis showing the effectiveness of the proposed method in terms of forecasting performances relative to the ARIMA-GARCH, ANN and Wavelet ANN models.

4.1 The daily series of Itaipu dam displacement

A time series of physical displacement of the dam that supplies the Itaipu hydroelectric plant in southern Brazil is modeled due to its relevance and statistical properties. The Itaipu dam is the world's most powerful dam with a length of 170 Km and an average width of 7 Km, reaching an area of 1350 Km^2, and allowing electricity generation in excess of 90 billion KWh. It supplies 93% of the energy consumed by Paraguay and 20% of the Brazilian consumption. Please refer to http://www.aboutcivil.org/itaipu-dam-design-construction-facts.html for more information about the Itaipu dam. Monitoring and forecasting the dam's physical displacements allow engineers to take corrective actions to prevent structural damage and accidents that can lead not only to interruptions of electricity generation but also to failures of more serious consequences. The data in this application come from automatic measurements of displacement taken at daily intervals in the period from 28th October 2005 to 24th October 2012. Figure 2 shows the time plot of the corresponding 2554 daily observations. The initial 2506 observations were used as in-sample training and the last 48 observations were used as out-of-sample for model testing. Note that there were no exogenous variables, such as dam levels and water pressure measurements, available for use in the application. In fact, there is no other data available (of enough quality) than the dam displacement time series that could be used in the modeling process.
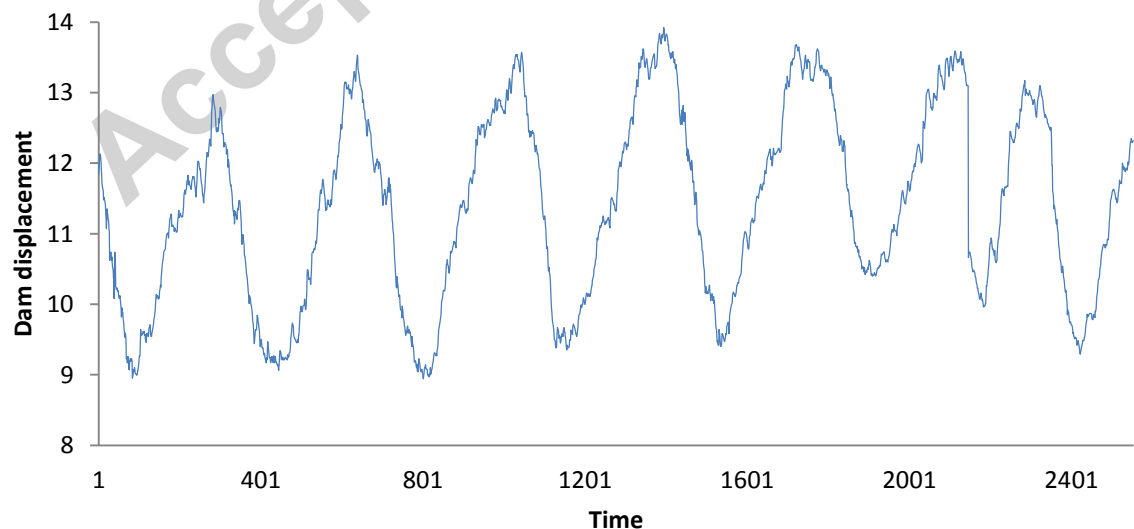

Figure. 2. The daily time series of dam displacement at the Itaipu electricity plant.

It can be seen from Figure 2 that there are both high and low frequency oscillations in the displacement data. Although a seasonal ARIMA model can capture the low-frequency variations, it would not be able to deal with the high-frequency ones at all. A WARIMAX-GARCH model however can model the high-frequency oscillations in the underlying series time through the ARIMA-GARCH models integrated with the wavelet decomposition approach.

An Augmented Dikey-Fuller (ADF) unit root test was conducted and did not reject, at 1% significance level, the hypothesis of non-stationarity of the displacement data. Nevertheless, the ADF strongly rejected non-stationarity, at 1% level, of the purely log-transformed version, as well as tautologically of the second difference of the log-transformed series. The second difference here was not used to get stationarity, but also for reaching improved accuracy gains. The simple and partial auto-correlation functions relative to this last one, as well as its squared version, showed significant values, at 1% level − what supports the choice of an ARIMA and GARCH modeling. Therefore, since there were no conventional exogenous variables available to be used by the ARIMAX-GARCH model in this application, it follows that an ARIMA-GARCH model was consider plausible for modeling process.

The investigation conducted in this section generated multi-step in-sample and out-of-sample (point and interval) forecasts from both an ARIMA-GARCH model (used as a benchmark model) and the WARIMAX-GARCH method. Residual diagnostics for statistical validation were conducted using augmented Dickey-Fuller unit root tests, Ljung-Box and Durbin-Watson tests for first order auto-correlation, as well as ACF and PACF plots, BDS tests to detect non-linear serial auto-correlations and ARCH tests for unconditionally constant residual variance (see e.g. [18]). The EViews 8 software was used to perform the tests above.

4.2 The ARIMA-GARCH model

An ARIMA (2, 1, 1) method integrated with the GARCH (2, 1) model (that is, an ARIMA-GARCH (2, 1, 1) x (2, 1) model), with Generalized Error Distribution (GED) (see e.g. [27]), was identified to produce the best fit to the log-transformed double-differenced training sample of dam displacement, y'_t ($t = 1, ..., 2506$). The parameters of that model were estimated by MLE and were statistically significant at 1% level (as can be seen in Appendix A).

Among all plausible models ARIMA-GARCH obtained from the training sample, the ARIMA-GARCH (3, 2, 4) x (2, 1) produce the more accurate forecasts to the level of the underlying time series, in terms of the in-sample MAPE and MAE. Note that the results of the Ljung-Box (Q-Stat) test applied to the standard residuals of ARIMA-GARCH (3, 2, 4) x (2, 1), shown in Figures 3 and 4, suggest that there is no significant linear auto-dependence (at 1% level) in both the ordinary and the squared standard residuals up to lag 36 (corresponding to three years). In addition, an ARCH test was also conducted and confirmed that there is no significant auto-correlations in the residual variance (at 1% level) from lags 1 to 36. Furthermore, the calculated Durbin-Watson statistic of 2.092367 suggests there is no significant auto-correlation of lag 1 in the ordinary standard residuals. Please see Appendix A for more detail of tests performed and to [18] for more information on the statistical techniques behind those tests.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.021	-0.021	1.0781	
		2	0.013	0.013	1.5119	
		3	0.001	0.002	1.5164	
		4	0.025	0.025	3.1276	
		5	0.027	0.028	5.0089	0.025
		6	0.033	0.033	7.7102	0.021
		7	0.021	0.022	8.8475	0.031
		8	0.007	0.006	8.9698	0.062
		9	0.026	0.025	10.707	0.058
		10	0.020	0.019	11.760	0.068
		11	0.034	0.031	14.587	0.042
		12	0.021	0.020	15.748	0.046
		13	0.013	0.011	16.201	0.063
		14	0.025	0.022	17.754	0.059
		15	0.012	0.009	18.124	0.079
		16	0.012	0.007	18.483	0.102
		17	0.029	0.025	20.638	0.080
		18	0.037	0.033	24.051	0.045
		19	0.032	0.029	26.669	0.032
		20	0.013	0.010	27.129	0.040
		21	0.026	0.021	28.873	0.036
		22	0.026	0.021	30.529	0.033
		23	0.010	0.003	30.781	0.043
		24	0.029	0.021	32.842	0.035
		25	0.005	-0.002	32.895	0.047
		26	0.018	0.011	33.746	0.052
		27	0.031	0.024	36.139	0.040
		28	0.028	0.021	38.154	0.033
		29	0.020	0.014	39.200	0.035
		30	0.027	0.020	41.042	0.031
		31	0.033	0.026	43.726	0.022
		32	0.015	0.008	44.288	0.026
		33	0.036	0.027	47.617	0.016
		34	0.010	0.003	47.871	0.020
		35	0.001	-0.009	47.872	0.027
		36	0.007	-0.003	48.004	0.034

Figure. 3. ACF and PACF of ordinary standard residuals.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.001	-0.001	0.0019	0.965
		2	-0.001	-0.001	0.0044	0.998
		3	-0.001	-0.001	0.0074	1.000
		4	-0.001	-0.001	0.0093	1.000
		5	-0.001	-0.001	0.0118	1.000
		6	-0.001	-0.001	0.0125	1.000
		7	-0.001	-0.001	0.0143	1.000
		8	-0.001	-0.001	0.0172	1.000
		9	-0.001	-0.001	0.0199	1.000
		10	-0.001	-0.001	0.0230	1.000
		11	-0.001	-0.001	0.0252	1.000
		12	-0.000	-0.000	0.0255	1.000
		13	-0.001	-0.001	0.0283	1.000
		14	-0.001	-0.001	0.0314	1.000
		15	-0.000	-0.000	0.0317	1.000
		16	-0.001	-0.001	0.0345	1.000
		17	-0.001	-0.001	0.0357	1.000
		18	-0.000	-0.000	0.0361	1.000
		19	-0.001	-0.001	0.0381	1.000
		20	-0.001	-0.001	0.0402	1.000
		21	-0.001	-0.001	0.0434	1.000
		22	-0.001	-0.001	0.0463	1.000
		23	-0.001	-0.001	0.0491	1.000
		24	-0.001	-0.001	0.0504	1.000
		25	-0.001	-0.001	0.0527	1.000
		26	-0.001	-0.001	0.0552	1.000
		27	-0.001	-0.001	0.0577	1.000
		28	-0.001	-0.001	0.0602	1.000
		29	-0.001	-0.001	0.0627	1.000
		30	-0.001	-0.001	0.0646	1.000
		31	-0.001	-0.001	0.0657	1.000
		32	-0.001	-0.001	0.0685	1.000
		33	-0.001	-0.001	0.0709	1.000
		34	-0.001	-0.001	0.0737	1.000
		35	-0.001	-0.001	0.0762	1.000
		36	-0.000	-0.000	0.0767	1.000

Figure. 4. ACF and PACF of squared standard residuals.

Table. 1. BDS test of in-sample ordinary residuals of the ARIMA-GARCH model.

Dimensions	BDS statistics	p-values
2	-3.22E-07	0.9840
3	-9.65E-07	0.9785
4	-1.93E-06	0.9743
5	-3.22E-06	0.9707
6	-4.83E-06	0.9675

Table 1 shows the statistics and the corresponding p-values for dimensions 2 to 6 of a BDS test (which consists of a statistical test used to verify the existence of linear and non-linear

auto-dependence existing in a data set (see e.g. [28])) applied to the in-sample residuals of the ARIMA-GARCH (3, 2, 4) x (2, 1) model. According to those p-values it is possible to conclude that there is no evidence at 1% level of significance, in all five dimensions, of both linear and non-linear auto-dependences in the in-sample forecasting errors. In particular, the linear and squared auto-dependences previously present in the forecasting residuals have been properly mapped by the ARIMA-GARCH models specified above. Therefore, the in-sample forecasting residuals can be considered as a white noise process with zero mean.

It is worth mentioning that model selection among all identified plausible ARIMA-GARCH models was determined by comparing the forecasting performances of each candidate model as measured by their Absolute Percentage Error (APE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and R^2. The selected best model, the one with the smallest APE, MAPE and MAE and with the largest R^2, is described in the following section.

4.3 The WARIMAX-GARCH model

The first step of the WARIMAX-GARCH method was implemented in MATLAB (version 2013a). A wavelet decomposition of level 2 was obtained from the training sample of the dam displacement data. The plots of the WCs, with orthonormal basis db40 (as described in [29]), can be seen in Figure 5.

(a) WC of approximation at level 2, $\widetilde{y}_{A_2,t}$ ($t = 1, \ldots, 2506$).

(b) WC of detail at level 2, $\widetilde{y}_{D_2,t}$ ($t = 1, \ldots, 2506$).

(c) WC of detail at level 3, $\widetilde{y}_{D_3,t}$ ($t = 1, \ldots, 2506$).
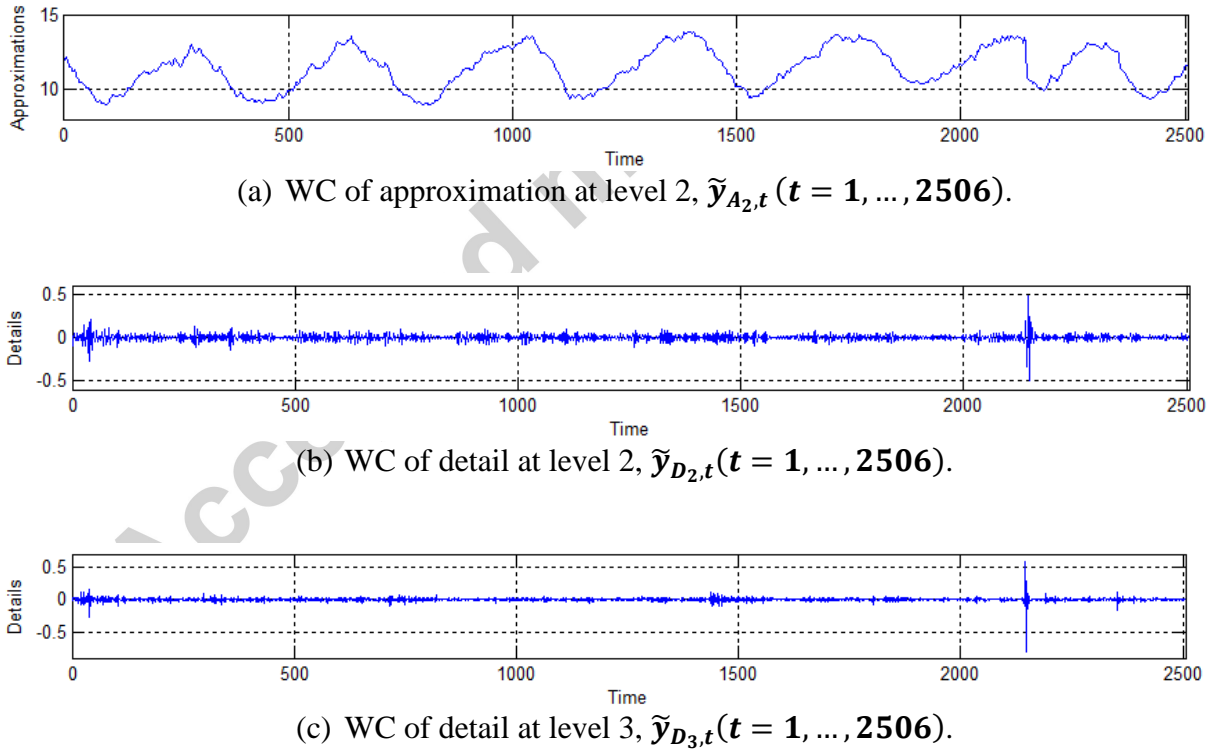
Figure. 5. WCs from the training sample of the dam displacement data set.

Recall from (2) that, for this application, $y_t = \widetilde{y}_{A_2,t} + \widetilde{y}_{D_2,t} + \widetilde{y}_{D_3,t} + \varepsilon$ ($t = 1, \ldots, 2506$), where ε is a null error (that is, $\varepsilon = 0$). In Step 2 of the WARIMAX-GARCH method, each one of the WCs $\widetilde{y}_{A_2,t}$, $\widetilde{y}_{D_2,t}$ and $\widetilde{y}_{D_3,t}$ ($t = 1, \ldots, 2506$) were individually modeled by three different ARIMA-GARCH models (which details are shown in Appendix B), providing the sequence of the out-of-sample forecasts to their respective levels, namely $\widehat{y}_{A_2,t}$, $\widehat{y}_{D_2,t}$ and

$\widehat{y}_{D_3,t}$ ($t = 2507, \dots, 2554$). The forecasting horizon was chosen as $h = 48$ (i.e., 48 days-ahead). All estimated parameters of the three models were statistically significant, at 1%, and the residual diagnostics confirmed their plausibility (as shown in Appendix B).

In Step 3, the three CWCs $\widetilde{y}_{A_2,t,C}$, $\widetilde{y}_{D_2,t,C}$ and $\widetilde{y}_{D_3,t,C}$ (or x_1, x_2 and x_3, respectively) consist of the WEVs generated from the dam displacement series. As mentioned before, they are easily obtained by filling in the WCs with their out-of-sample forecasts produced in the Step 2. Algebraically, it means that

I. $x_{1,t} = \widetilde{y}_{A_2,t,C} := \left(\left(\widetilde{y}_{A_2,t} \right)_{t=1}^{2506}; \left(\widehat{y}_{A_2,t} \right)_{t=2507}^{2554} \right)$;

II. $x_{2,t} = \widetilde{y}_{D_2,t,C} := \left(\left(\widetilde{y}_{D_2,t} \right)_{t=1}^{2506}; \left(\widehat{y}_{D_2,t} \right)_{t=2507}^{2554} \right)$; and

III. $x_{3,t} = \widetilde{y}_{D_3,t,C} := \left(\left(\widetilde{y}_{D_3,t} \right)_{t=1}^{2506}; \left(\widehat{y}_{D_3,t} \right)_{t=2507}^{2554} \right)$.

In Step 4, using the WEVs $x_{i,t}$ ($i = 1, 2, 3$) ($t = 1, \dots, 2554$), the WARIMAX-GARCH model was adjusted and then used to generate 48-steps-ahead forecasts to the level and the conditional variance of the underlying series. Algebraically, the forecasting formulation of the best WARIMAX model obtained is given in (12).

$$\left(1 - \phi_1 B - \cdots - \phi_5 B^5\right) y_t = \left(1 - \theta_1 B - \cdots - \theta_5 B^5\right) e_t + \sum_{i=1}^{2} \widetilde{y}_{A_2,t-i,C}\ \varphi_{A_2,t-i} + \widetilde{y}_{D_2,t-1,C}\ \varphi_{D_2,C,t-1} + \widetilde{y}_{D_3,t-2,C}\ \varphi_{D_3,C,t-2} + \varphi\ \sigma_t^2. \qquad (12)$$

In Appendix B, the estimated WARIMAX model above together with its main statistics can be seen in more detail. Note that all three WEVs, $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$, were required in the best WARIMAX model above. In addition, the volatility σ_t^2 provides a non-linear effect to represent the state y_t. Also, the best GARCH model was a GARCH (1, 1) with GED distribution, with the following algebraic formulation

$$\sigma_t^2 = \alpha_0 + \alpha_1\ e_{t-i}^2 + \beta_i\ \sigma_{t-1}^2.$$

Its estimates and main statistics can also be seen in Appendix B. The MLE method was used to obtain all estimates of the best WARIMAX-GARCH model parameters. Appendix C shows that all estimates are statistically significant, at 1%.

Figures 6 and 7 show, respectively, the ACF and the PACF (from lags 1 to 36) of both the ordinary and the squared standard residuals from the estimated WARIMAX-GARCH model. Note that all estimated ACF and PACF fall within the 99% confidence intervals suggesting those values were all non-significant. The Ljung-Box (Q-Stat) statistics on those figures also suggest there are no significant linear auto-dependences (at 1% level) in the in-sample ordinary and squared standard residuals of the WARIMAX-GARCH model. Also, note the ARCH test, (Prob) in Figure 7, confirmed there is no ARCH effect in the forecasting residuals. The Durbin-Watson statistic was 2.026328 confirming the lack of first-order auto-correlation of those in-sample standard residuals (please see Appendix C for more details).

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.008	-0.008	0.1531	
		2	-0.050	-0.050	6.3235	
		3	-0.016	-0.017	6.9561	
		4	-0.011	-0.014	7.2530	
		5	-0.001	-0.003	7.2569	
		6	0.006	0.005	7.3610	
		7	0.004	0.003	7.3970	
		8	-0.010	-0.010	7.6477	
		9	0.001	0.001	7.6512	
		10	0.022	0.022	8.9088	0.012
		11	0.007	0.007	9.0266	0.029
		12	-0.014	-0.012	9.5305	0.049
		13	-0.003	-0.002	9.5503	0.089
		14	-0.001	-0.001	9.5514	0.145
		15	-0.003	-0.003	9.5729	0.214
		16	0.017	0.016	10.304	0.244
		17	-0.002	-0.002	10.311	0.326
		18	0.034	0.036	13.227	0.211
		19	0.007	0.008	13.353	0.271
		20	0.032	0.035	15.929	0.195
		21	0.004	0.007	15.979	0.250
		22	0.005	0.010	16.036	0.311
		23	-0.004	-0.002	16.078	0.377
		24	0.003	0.004	16.097	0.446
		25	-0.002	-0.002	16.104	0.516
		26	0.028	0.028	18.115	0.448
		27	0.001	0.001	18.116	0.515
		28	0.017	0.019	18.842	0.532
		29	0.001	0.002	18.846	0.595
		30	0.046	0.049	24.202	0.337
		31	0.001	0.003	24.205	0.393
		32	0.029	0.035	26.329	0.337
		33	0.008	0.011	26.487	0.382
		34	-0.011	-0.006	26.773	0.421
		35	-0.007	-0.005	26.893	0.470
		36	-0.008	-0.011	27.042	0.516

Figure. 6. ACF and PACF of ordinary standard residuals.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.004	0.004	0.0343	0.853
		2	0.008	0.008	0.2130	0.899
		3	-0.002	-0.002	0.2233	0.974
		4	-0.000	-0.000	0.2235	0.994
		5	-0.003	-0.003	0.2396	0.999
		6	0.001	0.001	0.2433	1.000
		7	-0.003	-0.003	0.2594	1.000
		8	-0.000	-0.000	0.2599	1.000
		9	-0.003	-0.003	0.2763	1.000
		10	0.000	0.000	0.2768	1.000
		11	-0.002	-0.002	0.2922	1.000
		12	0.001	0.001	0.2931	1.000
		13	-0.003	-0.003	0.3097	1.000
		14	0.000	0.000	0.3103	1.000
		15	-0.003	-0.002	0.3261	1.000
		16	0.001	0.001	0.3311	1.000
		17	-0.003	-0.003	0.3476	1.000
		18	-0.000	-0.000	0.3481	1.000
		19	-0.002	-0.002	0.3637	1.000
		20	0.000	0.000	0.3640	1.000
		21	-0.003	-0.003	0.3810	1.000
		22	-0.000	-0.000	0.3811	1.000
		23	-0.003	-0.003	0.3986	1.000
		24	-0.001	-0.001	0.4002	1.000
		25	-0.003	-0.003	0.4171	1.000
		26	-0.000	-0.000	0.4171	1.000
		27	-0.003	-0.003	0.4341	1.000
		28	0.000	0.000	0.4342	1.000
		29	-0.003	-0.002	0.4503	1.000
		30	0.002	0.002	0.4642	1.000
		31	-0.003	-0.003	0.4808	1.000
		32	-0.001	-0.001	0.4816	1.000
		33	-0.003	-0.003	0.4983	1.000
		34	0.001	0.001	0.4991	1.000
		35	-0.003	-0.002	0.5155	1.000
		36	0.002	0.002	0.5232	1.000

Figure. 7. ACF and PACF of squared standard residuals.

Table. 2. BDS test of in-sample ordinary residuals of the WARIMAX-GARCH model.

Dimensions	BDS statistics	p-values
2	-3.22E-07	0.9840
3	-9.65E-07	0.9785
4	-1.93E-06	0.9743
5	-3.22E-06	0.9707
6	-4.83E-06	0.9675

Based on the p-values in Table 2, it can be concluded that there is no strong evidence (at 1% level) of both the linear or non-linear auto-dependence structure in the in-sample ordinary

standard residuals in all dimensions. Particularly, the linear and squared auto-dependence previously structures existing in them have been properly mapped by the adopted WARIMAX-GARCH model above. Therefore, the in-sample forecasting residuals can be considered as a white noise process with mean of zero, validating the WARIMAX-GARCH model.

4.4 ANN and WANN methods

As for the ANN and WANN methods, an iterative computational algorithm was used to test values for the ANN and WANN parameters and to choose their optimum values. The tested ANN parameters were the number of window lengths (L') (i.e., $L' = 1, ..., 10$), of neurons (N) in hidden layer (i.e., $N = 1, ..., 5$) and, in the case of Wavelet ANN method, of the null moments (N') of the Daubechies wavelet functions (as in [14]) for each WC, (i.e., $N' = 1, ..., 45$). In order to avoid excessive processing time, the following parameters were kept as fixed: premnmx normalization (as in [23]); one hidden layer; hyperbolic tangent and linear activation functions at the hidden and output (endowed with one neuron) layers, respectively; Levenberg-Marquardt training algorithm (for details about the mentioned ANN parameters, please, see [23]); and, in the case of Wavelet ANN method, $r = 2$ (i.e., wavelet decomposition of level 2), following the [1]'s approach. So, Table 3 exhibits the obtained optimal parameters of both the ANN and WANN methods.

Table. 3. Optimal ANN and WANN parameters.

METHOD	Parameters	Optimal values	Wavelet Component
ANN	L'	4	-
	N	2	-
Wavelet ANN	L'	6	Approximation at
	N	3	level 2
	N'	1	
	L'	2	Detail
	N	2	at level 2
	N'	1	
	L'	2	Detail
	N	3	at level 3
	N'	1	

4.5 Comparative of forecasting performances

Table 4 shows the MAPE and the MAE statistics for the in-sample and the out-of-sample forecasting performances of the three benchmark methods (namely, the ARIMA-GARCH, ANN and WANN approaches) and the WARIMAX-GARCH models. The optimal ARIMA-GARCH model obtained in Section 4.3 is an ARIMA-GARCH (3, 2, 4) x (2, 1).

It can be seen from the results in Table 4 that the WARIMAX-GARCH method had smaller MAPE and MAE values, and thus, better performances than the ARIMA-GARCH, ANN and WANN models both for the in-sample and out-of-sample periods. In particular, the WARIMAX-GARCH method produced greater improvements in out-of-sample forecasting performances, with 0.899% of MAPE and 0.1065 of MAE against 1.840% of MAPE and 0.2216 of MAE for the best benchmark method - i.e., WANN model (an improvement in excess of about 52% in both counts). It appears from those results that the use of WCs implied in significant forecasting performance improvements both in-sample and (in particular) out-

of-sample. In other words, the WARIMAX-GARCH has better modeled the dynamics of the referred time series and produced better forecasts than the WANN method.

Table. 4. The in-sample and out-of-sample forecasting performances.

METHODS	MAPE		MAE	
	In-sample	Out-of-sample	In-sample	Out-of-sample
WARIMAX-GARCH	0.300%	0.899%	0.0338	0.1065
ARIMA-GARCH (benchmark I)	0.398%	5.324%	0.0448	0.6368
ANN (benchmark II)	0.425%	2.247%	0.0474	0.2710
Wavelet ANN (benchmark III)	0.438%	1.840%	0.0495	0.2216

Figure 8 shows the plot of the Absolute Percentage Errors (APEs) calculated for the out-of-sample forecasts at each step-ahead by both models. Note that the WARIMA-GARCH model not only had lower APEs than the ARIMA-GARCH model for each step-ahead forecast from 1 to 48, but also produced comparatively better forecasts for larger steps-ahead. In fact, while the forecasting errors of the ARIMA-GARCH model showed a positive trend (growing from just under 1% to about 9%) the WARIMAX-GARCH showed fluctuations along the 1% line with increases in the forecasting horizon. Appendix B shows other plots of the temporal evolution of out-of-sample observations and the respective forecasts from both methods to support the conclusion that, in fact, the WARIMAX-GARCH model has a better power of generalization.
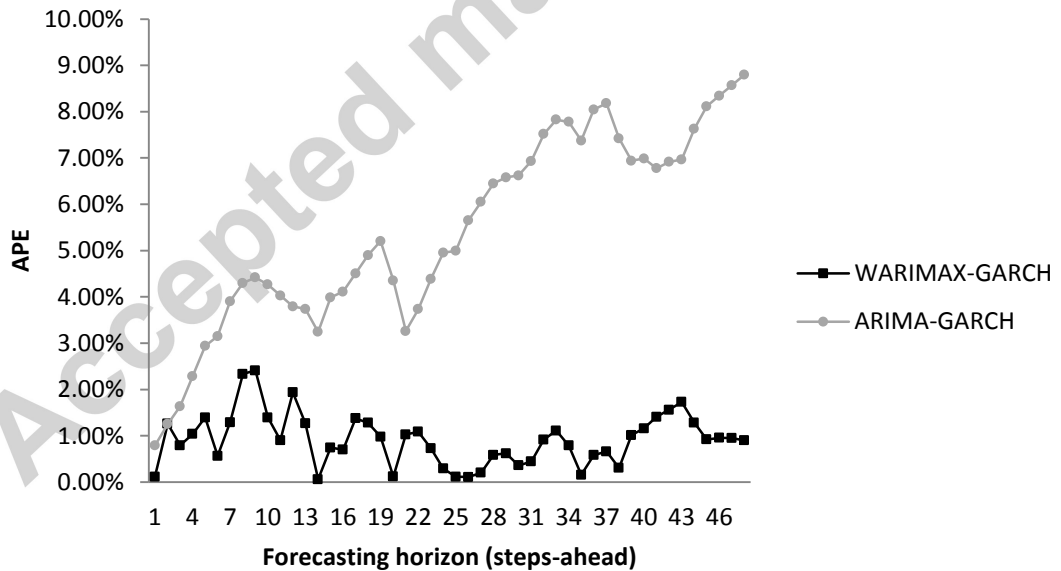

Figure. 8. Comparison of temporal evolution of APEs obtained from the WARIMAX-GARCH and ARIMA-GARCH methods.

Figure 9 also exhibits the plot of APEs calculated for the out-of-sample forecasts at each step-ahead by both methods. From step 20, it can be seen that the WARIMAX-GARCH method has obtained remarkably better forecasting accuracy than the Wavelet ANN (the

second best method), accounting for the best forecasting performance amongst all the benchmark methods in Table 4.

Note that, similarly to the ARIMA-GARCH model, the WANN method have lost accuracy for larger forecasting horizons, while the WARIMAX-GARCH has retained its forecasting performance.
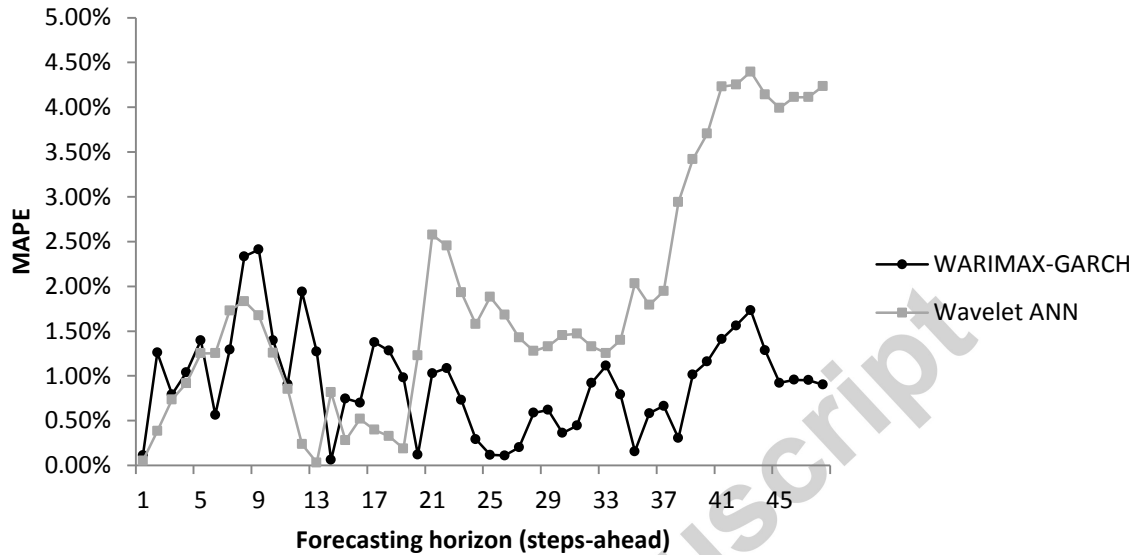

Figure. 9. Comparison of temporal evolution of APEs obtained from the WARIMAX-GARCH and WANN methods.

Figures 10 (a) and (b) show the plots of the actual observed displacements together the point forecasts and associated 99% upper and lower forecasting limits by (a) the ARIMA-GRACH model and (b) the WARIMAX-GARCH method. Notice that in Figure 10 (a) the forecasts by the ARIMA-GRACH model were all larger than the observed values at each time, however all within the 99% prediction intervals.

On the other hand, in Figure 10 (b), the forecasts by the WARIMAX-GARCH method tracked the trend in displacement more closely and also were all within the 99% prediction intervals which limits much smaller than those of the ARIMA-GARCH model. That is, the variance of the predictive density from WARIMAX-GARCH was lesser. This plot also shows the dynamics of the forecasts produced by the WARIMAX-GARCH method that tried to project the oscillations of the displacements into the 'future'.
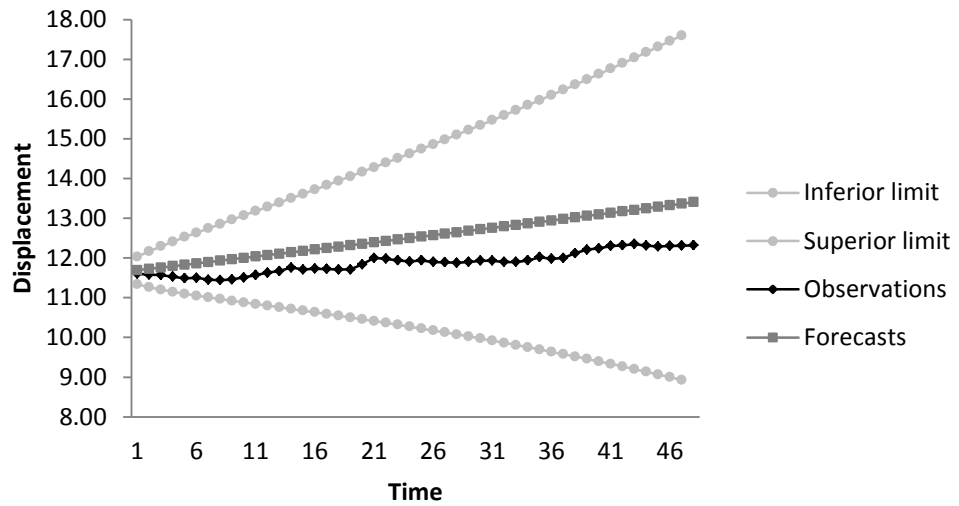
(a) ARIMAX-GARCH Method.
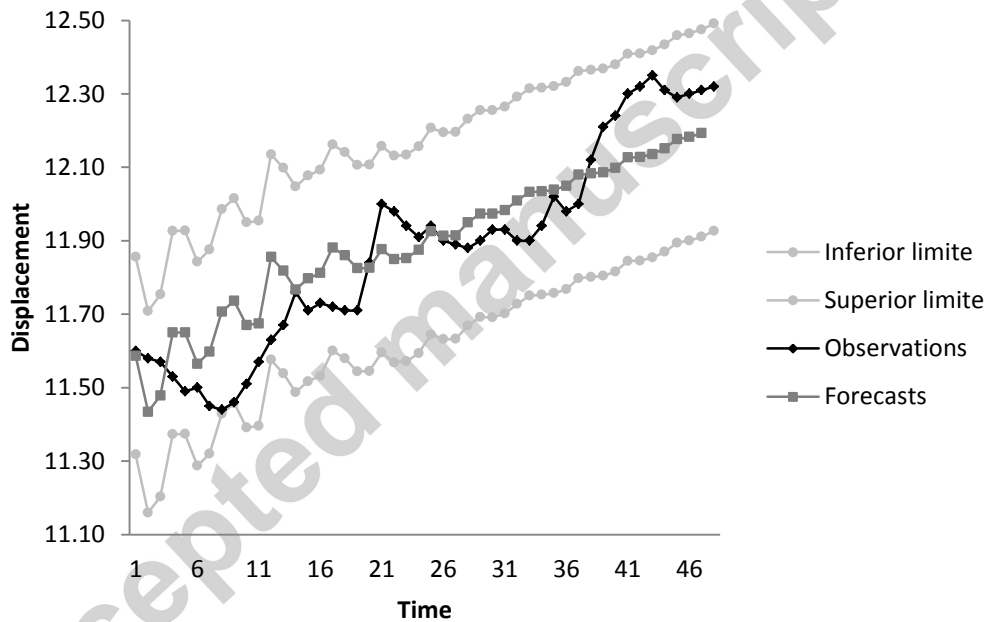
(b) WARIMAX-GARCH method.

Figure. 10. Forecasts (with prediction intervals) and observed displacements.

In terms of the R^2 coefficient, which is used to measure the amount of variation in the time series that is explained by the estimated method, the ARIMA-GARCH model had $R^2 = 0.361378$ and the WARIMAX-GARCH method had $R^2 = 0.997249$. Those results show the WARIMAX-GARCH explained approximately 99.72% of the variations in the time series of Itaipu dam displacements, while the ARIMA-GARCH only explained 36.14%.

5. Conclusions

In this paper, a new causal forecasting method called the WARIMAX-GARCH method is proposed that incorporates wavelet variables (obtained from wavelet decomposition of the underlying series) treated as exogenous variables incurring in substantial improvements in

forecasting performances over traditional ARIMA-GARCH, ANN and Wavelet ANN models. The incorporated wavelet components have good statistical properties to be used as exogenous variables by the WARIMAX-GARCH method. For instance, the detail components are always a second-order stationary process (usually required from exogenous variables that integrate a linear statistical regression model) and also always exhibit conditional variance (volatility) − similarly to a number of the financial time series (see e.g. [6]) − that enables nonlinear effects to be accounted for in the final model. Also, the approximation wavelet component can always be modeled by an ARIMA-GARCH model whenever the original underlying time series can also be. Furthermore, it can be easily seen that WCs always show strong correlations with the response variable they are obtained from.

The proposed method was applied to a daily time series of dam displacement in southern Brazil. Comparative results against the ARIMA-GARCH model showed the WARIMAX-GARCH method not only to produce significantly improved point forecasting performances but also improved predictive accuracy as measured by the prediction intervals that are straightforwardly operationally obtained. In addition, the WARIMAX-GARCH model has achieved considerably better forecasting performance than both the ANN and Wavelet ANN methods in the Itaipu dam displacement application.

This methodology has also been applied to other time series with similar results and will be subject of a future publication.

References

[1]   L.A. Teixeira Júnior, R.M. Souza, L.M. Menezes, K.M. Cassiano, J.F.M. Pessanha, R.C. Souza, Artificial Neural Network and Wavelet decomposition in the Forecast of Global Horizontal Solar Radiation, Sobrapo. 35 (2015) 1−16.

[2]   C. Chatfield, The Analysis of Time Series: an Introduction., Chapman & Hall/CRC, 6th ed., 2004.

[3]   G.E. Box, G.M. & Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, Oakland, Califórnia, 1970.

[4]   G.E.P. Box, G.C. Tiao, Intervention Analysis with Applications to Economic and Environmental problems, J. Am. Stat. Assoc. 70 (1975) 70−79.

[5]   A. Pankratz, Forecasting with Dynamic Regression Models, Wiley-Interscience., 1991.

[6]   R.F. Engle, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica. vol. 50 (1982) 987−1007.

[7]   T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econom. 31 (1986) 307–327.

[8]   T. Bollerslev, E. Ghysels, Periodic autoregressive conditional heteroscedasticity, J. Bus. Econ. Stat. 14 (1996) 139–151.

[9]   R. Engle, D. Lilien, R. Robins, Estimating time varying risk premia in the term structure: The ARCH-M model, Econometrica. 55 (1987) 391–407.

[10]   D.B. Nelson, Conditional Heteroscedasticity in Asset Returns: A New Approach, Econometrica. 59 (1991) 347–370.

[11]   T.T. Kalyan, J. van R. Garrett, The Theory and Practice of Revenue Management, Springer, 2006.

[12]   C.S. Kubrusly, The Elements of Operator Theory, 2nd edn, Birkhäuser/Springer, New York, 2011.

[13]   N. Levan, C.S. Kubrusly, A wavelet "time-shift-detail" decomposition, Math. Comput. Simul. 63 (2003) 73–78.

[14]   S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Third Edit, 2008.

[15]   C.S. Kubrusly, N. Levan, Abstract Wavelets Generated by Hilbert Space Shift Operators, Adv. Math. Sci. Appl. 14 (2006) 643–660.

[16]   D.L. Donoho, J.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika. 81 (1994) 425–455.

[17]   E. Haven, X. Liu, L. Shen, De-noising option prices with wavelet method, Eur. J. Oper. Res. 222 (2012) 104–112.

[18]   J. Hamilton, Time Series Analysis, Princeton University Press, 1994.

[19]   H. Lutkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2006.

[20]   T. Bollerslev, Generalized Autoregressive Conditional Heteroscedasticity, J. Econom. 31 (1990) 307–327.

[21] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control. Signals, Syst. 2 (1989) 303–314.

[22] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing. 50 (2003) 159–175.

[23] S.S. Haykin, Redes Neurais - 2ed., 2° ed., Bookman, Porto Alegre, 2001.

[24] H. Liu, H.Q. Tian, C. Chen, Y. Li, A Hybrid Statistical Method to Predict Wind Speed and Wind Power, Renew. Energy. 35 (2010) 1857–1861.

[25] B. Krishna, R.Y.R. Satyaji, P.C. Nayak, Time Series Modeling of River Flow Using Wavelet Neural Networks, J. Water Resour. Prot. 3 (2011) 50–59.

[26] K.K. Minu, M.C. Lineesh, J.C. Jessy, Wavelet Neural Networks for Nonlinear Time Series Analysis, Appl. Math. Sci. 4 (2010) 2485–2495.

[27] L.M. Liu, Time Series Analysis and Forecasting, Scientific Computing Associations, 2006.

[28] J.D. Hamilton, Causes and Consequences of the Oil Shock of 2007-08, (2009). http://www.nber.org/papers/w15002 (accessed May 2, 2015).

[29] I. Daubechies, Orthonormal bases of compactly supported wavelets, Commun. Pure Appl. Math. 41 (1988) 909–996.

Jairo Marlon Corrêa received the B.S. degree in Mathematics from State University of West Paraná, Foz do Iguaçu, Paraná, Brazil, in 2003; and M.S. degree in Numerical Methods in Engineering from Federal University of Paraná, Curitiba,

Brazil, in 2007. Currently, he is a Ph.D. student in Numerical Methods in Engineering from Federal University of Paraná, Curitiba, Brazil, as well as a professor in Federal Technological University of Paraná, Medianeira, Paraná, Brazil. His research interests include stochastic processes, time series forecasting, machine learning, data mining, applied statistics, ARIMA-GARCH models and their extensions, wavelet pre-processing methods, optimization, artificial neural networks, and risk analysis applied to Dam safety.

Anselmo Chaves Neto holds a B.S. Civil Engineering degree from Federal University of Curitiba, Paraná, Brazil, in 1974; M.S. degree in Statistics from University of Campinas, Campinas, São Paulo, Brazil, in 1985; and Ph.D. degree in Electrical Engineering from Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, in 1991. Presently, he is a full professor in Federal University of Curitiba, Paraná, Brazil. As researcher, his interests include multivariate statistical methods, time series forecasting, applied statistics, data mining, quality engineering, computationally intensive methods, pattern recognition, and Dam safety.

Luiz Albino Teixeira Junior holds a Ph.D. degree and a M.S. degree in Electrical Engineering from Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, in 2013. Part of his Ph.D. was done in in Department of Mathematics and Statistics in The Open University, Milton Keynes, Buckinghamshire, England. He is currently a professor in Federal University of Latin American Integration and his research interests include stochastic processes, analysis and forecast of stochastic time series, wavelet theory, singular spectral analysis, spectral analysis, multivariate statistics, data mining, artificial intelligence, machine learning, quality control of processes, pattern recognition, risk numerical analysis applied to Dam safety and Finances, optimization, and hybrid combination of forecasting methods.

Edgar Manoel Careño received the B.S. degree in Electrical Engineering from National University of Colombia, Bogotá, Colombia, in 1999; M.S. degree in Electrical Engineering from Technological University of Pereira, Pereira, Risaralda, Colombia, in 2003; and Ph.D. degree in Electrical Engineering in State University of São Paulo, São Paulo, Brazil, in 2008. Presently, he is a Professor in Department of Electrical Engineering in State University of West Paraná, Foz do Iguaçu, Paraná, Brazil. He has experience in Electrical Engineering with emphasis on planning of Transmission and distribution of electric power systems, acting on the following topics: forecast of demand, planning of transmission and distribution systems, and mathematical programming.

Álvaro Eduardo Faria holds a Ph.D. degree in Statistics from university of Warwick, in 1996; M.S. degree in Electrical Engineering; and B.S. degree in Systems Engineering from Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, in 1992 and 1984, respectively. He is presently a Professor in Department of Mathematics and Statistics in The Open University, Milton Keynes, Buckinghamshire, England. His research interests include multivariate Bayesian time series models, Bayesian modelling of space-time processes, Bayesian dynamic models for non-linear auto-regressive processes, multivariate statistics, Bayesian statistics, and hybrid forecasting models.