An intelligent hybridization of ARIMA with machine learning models for time series forecasting

Article *in* Knowledge-Based Systems · March 2019

DOI: 10.1016/j.knosys.2019.03.011

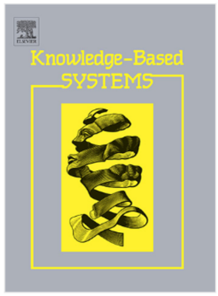
3 authors:

Domingos Santos
Universidade de Pernambuco

Joao Fausto Lorenzato de Oliveira
Universidade de Pernambuco

Paulo S. G. De Mattos Neto
Federal University of Pernambuco

Accepted Manuscript



An intelligent hybridization of ARIMA with machine learning models for time series forecasting

Domingos S. de O. Santos Júnior, João F.L. de Oliveira, Paulo S.G. de Mattos Neto

PII: S0950-7051(19)30132-7
DOI: https://doi.org/10.1016/j.knosys.2019.03.011
Reference: KNOSYS 4707

To appear in: *Knowledge-Based Systems*

Received date : 25 September 2018
Revised date : 14 March 2019
Accepted date : 15 March 2019

Please cite this article as: D.S.d.O. Santos Júnior, J.F.L. de Oliveira and P.S.G. de Mattos Neto, An intelligent hybridization of ARIMA with machine learning models for time series forecasting, *Knowledge-Based Systems* (2019), https://doi.org/10.1016/j.knosys.2019.03.011

An intelligent hybridization of ARIMA with machine learning models for time series forecasting

Domingos S. de O. Santos Júnior[a,b], João F. L. de Oliveira[b], Paulo S. G. de Mattos Neto[a,*]

[a] *Centro de Informática, Universidade Federal de Pernambuco, Recife-PE, Brazil*
[b] *Universidade de Pernambuco, Recife-PE, Brazil*

Abstract

The development of accurate forecasting systems can be challenging in real-world applications. The modeling of real-world time series is a particularly difficult task because they are generally composed of linear and nonlinear patterns that are combined in some form. Several hybrid systems that combine linear and nonlinear techniques have obtained relevant results in terms of accuracy in comparison with single models. However, the best combination function of the forecasting of the linear and nonlinear patterns is unknown, which makes this modeling an open question. This work proposes a hybrid system that searches for a suitable function to combine the forecasts of linear and nonlinear models. Thus, the proposed system performs: (i) linear modeling of the time series; (ii) nonlinear modeling of the error series; and (iii) a data-driven combination that searches for: (iii.a) the most suitable function, between linear and nonlinear formalisms, and (iii.b) the number of forecasts of models (i) and (ii) that maximizes the performance of the combination. Two versions of the hybrid system are evaluated. In both versions, the ARIMA model is used in step (i) and two nonlinear intelligent models—Multi-Layer Perceptron (MLP) and Support Vector Regression (SVR)—are used in steps (ii) and (iii), alternately. Experimental simulations with six real-world complex time series that are well-known in the literature are evaluated using a set of popular performance metrics. Our results show that the proposed hybrid system attains superior performance when compared with single and hybrid models in the literature.

*Corresponding author
Email address: psgmn@cin.ufpe.br (Paulo S. G. de Mattos Neto)

Keywords: Machine learning, Autoregressive integrated moving average (ARIMA), Time series forecasting, Error series, Hybrid system, Artificial Neural Networks

1. Introduction

Time series forecasting is a relevant task in several areas of science [1], such as economics, finance, engineering, health sciences, and meteorology. Consequently, forecasting systems have been designed to model different social, natural, ecological, financial phenomenon, among others. However, the development of accurate forecasting systems that are able to satisfactorily model different temporal phenomena is still a considerable challenge [1–4].

In the literature, statistical methods and Artificial Neural Networks (ANNs) have been employed with success in several applications of time series forecasting [2, 5–9]. The Autoregressive Integrated Moving Average (ARIMA), Autoregressive (AR) and Moving Average (MA) models have been employed in several applications and can be designed by a Box & Jenkins methodology [1]. These statistical methods are simple, flexible and they can be calibrated to model several temporal phenomena [1, 10–12]. Meanwhile, AR, MA and ARIMA are purely linear models and, therefore, have a limited performance in the real-world time series modeling, which commonly present linear and nonlinear temporal patterns [10, 13–15].

ANNs, among nonlinear forecasting models, have been highlighted due to their relevant results in terms of accuracy [2, 16]. ANNs are flexible and data-driven models that are able to perform nonlinear modeling from a training procedure. In time series forecasting, the learning process of an ANN consists of modeling past observations with the objective of estimating the underlying temporal relationship of the phenomenon. However, in real-world time series forecasting, the adoption of a single ANN may not be sufficient for modeling both linear and nonlinear patterns equally well due to problems with misspecification, under-fitting or overfitting of the model [10, 14, 15, 17–19].

In this scenario, hybrid systems that combine classical statistical models and ANNs have reached relevant results in terms of accuracy in different fields of application [3, 5, 6, 10, 11, 13–15, 20–23]. These architectures employ the error series (residuals)[1] modeling to increase the performance of entire system [3, 5, 6, 10, 11, 13–15, 20–23]. Hybrid systems which combine different techniques through error series modeling, aim to deal with the problems of selection and specification of models with little extra effort. Thus, combining linear and nonlinear models can improve the accuracy of the system [7, 14, 15] since linear and nonlinear patterns in data can be modeled more effectively.

The combination of linear and nonlinear models usually assumes a linear relation between their patterns [5, 10, 24]. Hybrid systems based on [10] perform an unweighted sum to combine the forecasts of the series and residuals. Despite being widely used in the literature [3, 5, 6, 23, 25], the linear combination can underestimate or degenerate the performance of the system because there may be no additive relationship between linear and nonlinear forecasts [26]. The second assumption, which was initially proposed by Khashei and Bijari [14], supposes a nonlinear combination between the linear and nonlinear components. Works that follow this assumption [14] employ machine learning (ML) methods to perform the nonlinear modeling. Hybrid systems proposed by Kashei and Bijari [14, 15] and Zhu *et al.* [25] generate the final forecast through a joint nonlinear modeling of residuals and time series. In general, the nonlinear combination reaches better results than the linear approach [14, 15, 25] showing the potential of this modeling. However, there is no guarantee that the nonlinear combination is the most appropriate for the modeling of any temporal phenomena [27]. Thus, the search of an appropriate combination of forecasts can be very challenging in hybrid systems [7, 14, 15, 27].

This work proposes a new strategy for combination of the forecasts from linear and nonlinear models. The proposed system can be divided into three sequential steps: (a) a linear model is employed to perform forecasts, (b) the residuals are modeled by a nonlinear model, (c) forecasts from the time series and residuals are combined by a data-driven model using a new strategy. The data-driven model employed in step (c) consists of a computational intelligence method that aims to analyze the data with the objective of finding temporal patterns (between input and output variables) without *a priori* knowledge of the phenomenon behavior. Thus, forecasts from the time series and residuals are both used as input to an intelligent method to find these

[1]Difference between actual series and forecasting

patterns and produce better results. In this way, any supervised learning model can be employed to perform the combination of the forecasts (time series and residuals).

In this work, we evaluated two computational intelligence models: Multi-Layer Perceptron (MLP) and Support Vector Regression (SVR). These models were chosen because they are widely used in the literature and to clearly show the contribution of the proposed method in comparison with existing hybrid systems that employ the same intelligent models [5, 7, 10, 14, 15, 19, 24, 28].

The main contribution of this paper is a versatile hybrid system for time series forecasting, which aims to find the most suitable combination function for describing the relationship between the forecasts of linear and nonlinear models. The system employs an intelligent model that performs the combination of linear and nonlinear techniques based on their respective forecasts. Experimental results are presented for well known time series: Canadian Lynx, Sunspot, British Pound/US Dollar Exchange Rate, Colorado River flow, Airline passengers and Star Brightness. The simulation results using a set of three evaluation measures show that the proposed hybrid system: (i) overcomes state-of-the-art hybrid systems, which perform combination of forecasters using residual modeling; (ii) reaches better accuracy than some single models; and (iii) improves the performance of the initial linear statistical model.

The rest of this paper is organized as follows. In the next section, we review hybrid systems of the literature that are the basis for the proposal. Section 3 describes the proposed hybrid system. Section 4 shows the experimental setup and results of the proposed system. Section 5 presents some considerations about the results and relevant aspects of the proposed system. Finally, Section 6 contains the concluding remarks and lists the future works.

2. Related works

The combination of forecasts can be achieved through the employment of ensembles of forecasters [4, 29–31]. Another approach to combine forecasts is to deal separately with time series and error series, which can also improve the accuracy of the system [6, 16, 21]. The latter approach is based on the assumption that real world data is composed by linear and nonlinear patterns [10, 15] and a single forecast model may not deal equally well with the linear and nonlinear patterns. This limitation can occur due to the nature of the forecaster in case of linear models or because of the misspecification of the parameters, overfitting or under-fitting model in case of nonlinear intelligent models [7].

Zhang [10] introduced the use of the linear combination of linear statistical models with ANNs. Zhang [10] proposed that a given time series $\mathbf{Z}_t$ could be composed of linear ($\mathbf{L}_t$) and nonlinear ($\mathbf{N}_t$) patterns as shown in Equation 1. ARIMA and MLP models were employed to forecast linear and nonlinear components respectively, where the ARIMA produces linear forecasts ($\hat{\mathbf{L}}_t$).

$$\mathbf{Z}_t = \mathbf{L}_t + \mathbf{N}_t. \tag{1}$$

The error series ($\mathbf{E}_t$) is obtained through the difference between the original series ($\mathbf{Z}_t$) and the linear forecasts ($\hat{\mathbf{L}}_t$) as shown in Equation 2.

$$\mathbf{E}_t = \mathbf{Z}_t - \hat{\mathbf{L}}_t. \tag{2}$$

Box & Jenkins methodology for designing ARIMA models states that the residual series should not present linear correlations. If the linear model is well specified, then statistical tests such as correlation function or Box-Pierce [1] should not find a linear correlation on the residual series. Thus, because there is no linear pattern in the error series, it is reasonable to use a nonlinear model to handle some possible nonlinear pattern present in the residual series. In this way, a nonlinear model with n inputs of $\mathbf{E}_t$ can be used to perform a residual forecasting $\hat{\mathbf{N}}_t$ (Equation 3).

$$\hat{\mathbf{N}}_t = f(e_{t-1}, e_{t-2}, ..., e_{t-n}) + \epsilon_t, \tag{3}$$

where $f(.)$ represents the nonlinear modeling performed by ANN and ϵ_t is a random error, which cannot be predicted.

Finally, the final forecast $\hat{\mathbf{Z}}_t$ is performed by the sum (linear combination) of linear $\hat{\mathbf{L}}_t$ and nonlinear $\hat{\mathbf{N}}_t$ forecasts (Equation 4).

$$\hat{\mathbf{Z}}_t = \hat{\mathbf{L}}_t + \hat{\mathbf{N}}_t. \tag{4}$$

Linear combination of ARIMA and MLP (denoted in this work by ARIMA + MLP) has been employed in the literature in several applications: water quality [11], fuel wood prices [12], wind speed [23] and particulate matter [22]. Firmino *et al.* [3] proposed a recursive modeling of the residual series using ARIMA+MLP to stock exchanges forecasting.

Other works of the literature proposed hybrid systems using different models based on the supposition of Zhang [10] (Equation 1). Zhou *et al.* [20] employed the combination of the ARIMA model with Nonlinear Autoregressive Neural Network, ARIMA+NARNN, to schistosomiasis in humans forecasting. Yu *et al.* [32] combined a seasonal ARIMA model with NARNN (SARIMA+NARNN) to predict the incidence of cases of Hand, Foot and Mouth Disease (HFMD), and Pai and Lin [6] used the combination ARIMA+SVR to stock exchanges forecasting. Panigrahi and Behera [5] proposed the combination of the exponential smoothing method (ETS) with MLP (ETS+MLP) to forecast different time series.

Alternatively, Khashei and Bijari [15] considered a nonlinear combination of the linear ($\mathbf{L}_t$) and nonlinear ($\mathbf{N}_t$) patterns; that is,

$$\mathbf{Z}_t = f(\mathbf{L}_t, \mathbf{N}_t), \tag{5}$$

where $f(.)$ is a nonlinear function of the linear and nonlinear components. In [15], Khashei and Bijari used the combination of the ARIMA and MLP models. In the first step of this approach, the linear component ($\hat{\mathbf{L}}$) is estimated by ARIMA model and is used to calculate the residual series (Equation 2), similarly to the work of Zhang [10]. In the second step, the nonlinear modeling is performed with the objective to capture nonlinear patterns from the residuals $\mathbf{E}_t$. In this hybrid system [15], the MLP model is used with the objective to perform jointly the residuals modeling and the combination of the linear and nonlinear components. Therefore, the nonlinear model receives as input data the past residuals $(e_{t-1}, ..., e_{t-m_1})$, the linear forecasting $(\hat{\mathbf{L}}_t)$, and the past data of the time series $(z_{t-1}, ..., z_{t-m_2})$, generating the final forecasting $\hat{\mathbf{Z}}_t$ (Equation 6),

$$\hat{\mathbf{Z}}_t = f(e_{t-1}, ..., e_{t-m_1}, \hat{\mathbf{L}}_t, z_{t-1}, ..., z_{t-m_2}), \tag{6}$$

where the indexes m_1 and m_2 are integers that represent the size of the temporal window of $\mathbf{E}_t$ and $\mathbf{Z}_t$, respectively.

Some works of the literature follow the assumption proposed in [15]. Chen and Wang [33] combined the SARIMA and SVM models to forecast the production values of the machinery industry. Aguilar *et al.* [21] combined the SARIMA and MLP models to seaport inspection forecasting. Zhu and Wei [25] combined an ARIMA model with Least Squares Support Vector Machines (LSSVM) for carbon price forecasting.

From suppositions of these works [10, 15, 18], de Mattos Neto *et al.* [7] proposed a combination of models, named NoLiC method, that is composed of three steps: time series forecasting, residuals forecasting and combination of the forecasts. In the first step, a hybrid intelligent system composed of Genetic Algorithm and MLP was used to model time series instead a linear model. In the second step, linear and nonlinear models were evaluated to forecast the residuals. In the final step, an intelligent model was used to combine the time series and residuals forecasts. The hybrid system proposed in [7] was evaluated with particulate matter concentration time series.

3. Proposed hybrid system

The proposed hybrid system is based on previous works of the literature. Zhang [10] showed that to model separately linear and nonlinear patterns employing two (linear and nonlinear) predictors is an approach that is capable of improving the performance of single models. Khashei and Bijari [15] employed a nonlinear model to combine linear and nonlinear patterns. Their approach [15] used the lags of time series, linear forecast and past errors as ANN inputs. The ANN is employed with the objective of, in a single step, modeling and combining the previous patterns to generate the final forecast. Khashei and Bijari [15] showed that their hybrid system achieved a higher level of accuracy than the simple sum proposed by Zhang [10]. The NoLiC method [7] performs forecasts through the employment of three steps: time series forecast, error forecast and the combination of these forecasts. The combination is achieved by nonlinear ML models that take into consideration forecast from other methods. Results reached by NoLiC method [7] showed that an exclusive combination can lead to better performance. Thus, the design of the proposed hybrid system was inspired by these works [7, 10, 15].

In general, the proposed system performs time series modeling in three sequential steps: (I) the forecasting of the time series ($\mathbf{Z}_t$) using a linear model ($\mathrm{M_L}$), (II) the forecasting of the error series $\mathbf{E}_t$ using a nonlinear model ($\mathrm{M_{NL}}$), and (III) the combination of the forecasts of the series and of the residuals using an ML model ($\mathrm{M_C}$) to generate the final output. Each step (I, II and III) is composed of two phases: training and test, which are presented in Figures 1 and 2, respectively.

Figure 1 shows the Training Phase. In Step I of Training Phase, given the training set of a univariate time series ($\mathbf{Z}_t$), the objective is to train the forecasting models $\mathrm{M_L}$ and $\mathrm{M_{NL}}$. First, the training of the model $\mathrm{M_L}$ is performed using the time series $\mathbf{Z}_t$. Next, the error series, or residuals $\mathbf{E}_t$, is calculated from the difference between $\mathbf{Z}_t$ and the training output of $\mathrm{M_L}$, $\mathbf{E}_t = \mathbf{Z}_t$ - $\mathrm{M_L}(\mathbf{Z}_t)$. Then, the error series $\mathbf{E}_t$ is used to train the model $\mathrm{M_{NL}}$. The output of the Steps I and II of the training phase are the series $\mathbf{Z}_t$ and $\mathbf{E}_t$, and the models $\mathrm{M_L}$ and $\mathrm{M_{NL}}$, respectively.

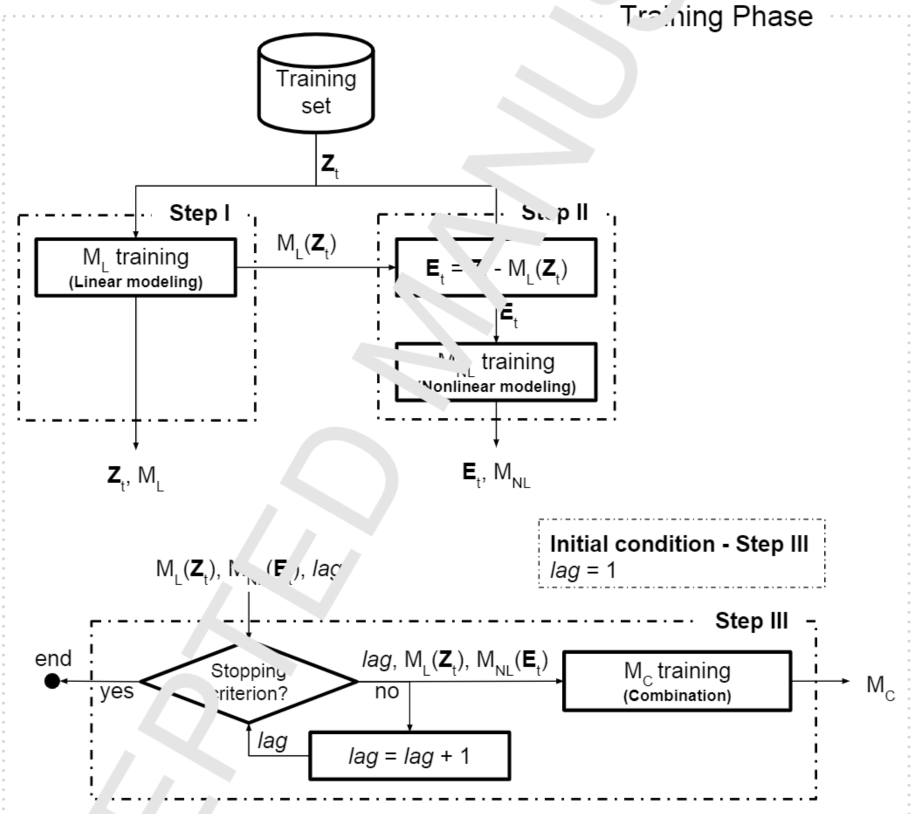


Figure 1: Training phase of the proposed hybrid system.

Several works [6, 7, 10, 13, 15, 21] have reported that the accuracy of the forecasting system can be improved from combining the prediction of the time series with its error series. In this way, the third step of the training phase of the proposed method (Step III in Figure 1) aims to combine the outputs of the models $\mathrm{M_L}$ and $\mathrm{M_{NL}}$, which were previously trained in Step I and II, respectively. Thus, the output of the Step III is the model M_C trained from the estimates of the M_L and M_{NL} models.

Figure 1 shows that Step III receives three inputs data: $M_L(\mathbf{Z}_t)$, $M_{NL}(\mathbf{E})$ and a variable called *lag*. Because the proposed method aims to find the combination function $f(.)$ most suitable to describe the relationship between the forecasts of the time series and the forecasts of the residual series, the model M_C must be versatile. From the input data the model M_C has to be able to search for the function combination $f(.)$ that best describes the relationship between the forecasts of the models M_L and M_{NL}, maximizing the predictive performance of the combination. Considering the level of adaptability of M_C, the proposed method employs an ML model because it can be flexible and can be adapted to temporal phenomena. Given the importance of the input data for training of the intelligent model M_C, the variable *lag* is a key point for the proposed method. The integer variable *lag* can assume values in the interval $[1, Lmax]$ and corresponds to the number of forecasts that are used of the M_L and M_{NL} models. The variable $Lmax$ is determined from cross-correlation [1] between time series and error series obtained from forecast of the M_L model. Thus, $Lmax$ can vary according to the time series under analysis. It is important to mention that other statistical measures can be used to assess the relationship between the series [1]. When the *lag* is equal to 1 (one), the M_C model is trained using only the forecasting for time $t+1$ of models M_L and M_{NL}, totalizing two inputs. For $lag > 1$, the M_C model receives, in addition to the forecast for $t+1$, $lag-1$ past predictions. For example, when $lag = 5$, M_C receives as input data the forecasts of the M_L and M_{NL} models in the temporal window $[t+1,\ t,\ t-1,\ t-2,\ t-3,\ t-4]$, totalizing 10 inputs. Thus, the model M_C, described in Equation 7, searches a suitable function of combination $f(.)$ between the forecasts of the time series (S) and the forecasts of the residuals ($\mathbf{R}$) obtained by M_L and M_{NL} using a given number of lags, respectively:

$$M_C = f(\mathbf{S}, \mathbf{R}), \tag{7}$$

where

$$\mathbf{S} = [M_L(\mathbf{Z})_{t+1}, M_L(\mathbf{Z})_t, \ldots, M_L(\mathbf{Z})_{t+1-(lag-1)}], \tag{8}$$

and

$$\mathbf{R} = [M_{NL}(\mathbf{E})_{t+1}, M_{NL}(\mathbf{E})_t, \ldots, M_{NL}(\mathbf{E})_{t+1-(lag-1)}]. \tag{9}$$

Step III of the training phase has one stopping criterion, which is the number maximum of time lags ($Lmax$) reached. After the training phase, the $\mathrm{M_L}$, $\mathrm{M_{NL}}$ and $\mathrm{M_C}$ models are used to forecast unseen patterns of the test set. The test phase, as shown in Figure 2, is also divided into three steps: forecasting of the time series by $\mathrm{M_L}$ model, forecasting of the residuals given by $\mathrm{M_{NL}}$ model, and the combination performed by $\mathrm{M_C}$ model. So, given an unseen test pattern $\mathbf{Z}_q$, the objective is to forecast $\mathbf{Z}_{t+1}$. The forecasting $\mathbf{Z}_{t+1}$ is given by Equation,

$$\mathbf{Z}_{t+1} = \mathrm{M_C}(\mathbf{S}, \mathbf{R}), \tag{10}$$

where $\mathbf{S}$ is a vector with *lag* forecasts of $\mathrm{M_L}$ and $\mathbf{R}$ is a vector with *lag* forecasts of $\mathrm{M_{NL}}$ given as input data to $\mathrm{M_C}$.

The proposed method in this work is flexible because it is able to combine the forecasts of the $\mathrm{M_L}$ and $\mathrm{M_{NL}}$ models with a linear or a nonlinear function. It is also able to select the number of forecasts required to improve this combination. Thus, the combination of predictors using residual analysis arises as a promising approach to model different temporal patterns. In this work we employed an ARIMA model to perform linear forecasting, MLP and SVR models to perform nonlinear forecasts and combination, alternately. Then, two versions of the proposed hybrid system were developed: MLP(ARIMA,SVR) and SVR(ARIMA,MLP), taking into consideration the adopted nomenclature $\mathrm{M_C}(\mathrm{M_L}, \mathrm{M_{NL}})$.

3.1. Linear modeling

The ARIMA is a traditional forecasting method commonly used to linear modeling. This method has a well established modeling methodology, proposed by Box and Jenkins [1], which aims to find the best ARIMA model order. An ARIMA(p,d,q) model consists of an AR model of order p (AR(p)), a MA model of order q (MA(q)) and a differentiation step d to produce a stationary series. The mathematical formulation of the ARIMA model is presented in Equation 11.

$$\mathbf{Z}_{t+1} = \mu+\phi_1 z_t+\phi_2 z_{t-1}+...+\phi_p z_{t-p+1}+\epsilon_{t+1}-\theta_1\epsilon_t-\theta_2\epsilon_{t-1}-...-\theta_q\epsilon_{t-q+1}, \tag{11}$$

where z_t and ϵ_t are the time series value and random error at time t, respectively; μ, $\phi_1, \phi_2, ..., \phi_p$ and $\theta_1, \theta_2, ..., \theta_q$ are the model parameters. The p and q are integers and refer to the model orders.

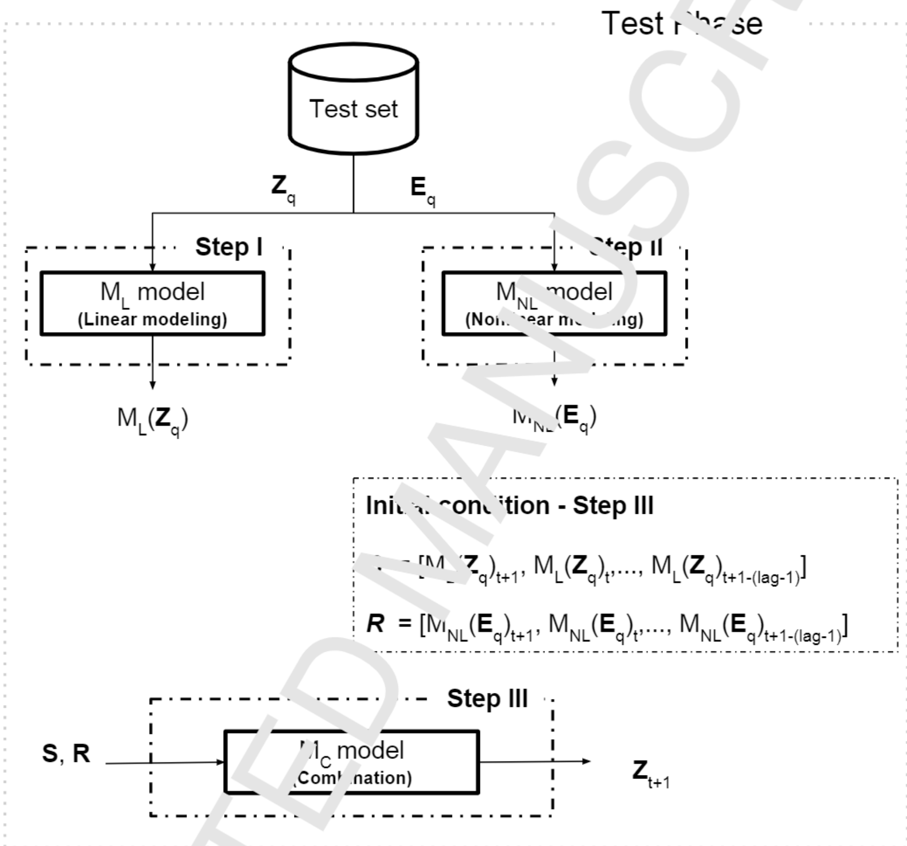


Figure 2: Test phase of the proposed hybrid system.

3.2. Nonlinear modeling

ANN are flexible, non-parametric methods which can perform nonlinear mappings from data. The mathematical formulation of the ANN for error modeling is presented in Equation 12.

$$\mathbf{E}_{t+1} = \alpha_0 + \sum_{j=1}^{s} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{r} \beta_{ij} e_{t-i+1}\right) + \vartheta_{t+1}, \tag{12}$$

where $\alpha_j (j = 0, 1, 2, ..., s)$ and $\beta_{ij} (i = 0, 1, 2, ..., r; j = 1, 2, ..., s)$ are connection weights. The number of input nodes and hidden nodes are represented by r and s, respectively, and ϑ_t corresponds to the forecasting error of the ANN. The sigmoid logistic is used as activation function (Equation 13).

$$g(e) = \frac{1}{1 + \exp(-e)}. \tag{13}$$

Likewise, the SVR also performs nonlinear mapping but using different training algorithm. It is based on a quadratic optimization procedure where the objective is to find a function in the form

$$\{f | f(\mathbf{e}) = \mathbf{w}^T\mathbf{e} + b, w \in \mathbb{R}^l, b \in \mathbb{R}\} \tag{14}$$

where w is a vector of weights estimated by optimizing the regularized risk function in Equation 15 and $\mathbf{b}$ is the bias. In Equation 15 $C > 0$ is a regularization factor, $\|.\|$ is a 2-norm and $L(.,.)$ is a loss function.

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} L(y_i, f(\mathbf{e}_i)) \tag{15}$$

By minimizing the regularized risk it is possible to find a suitable function for the problem. In order to introduce sparsity the $\varepsilon-$insensitive loss function is usually employed in the SVR presented in Equation 16. The loss is not computed if the estimated function values lies within the ε-tube.

$$L(y, f(\mathbf{x})) = \begin{cases} 0, & |f(\mathbf{x}) - y| < \varepsilon \\ |f(\mathbf{x}) - y| - \varepsilon, & \text{otherwise} \end{cases} \tag{16}$$

Thus, through the employment of the $\varepsilon-$insensitive function the SVR can be formulated as presented in Equation 17, where ξ and ξ^* are slack variables used to evaluate the errors of values that fall outside the limits determined by the ε-tube.

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{l}(\xi_i - \xi_i{}^*) \\ &\text{subject to} \quad \begin{cases} \mathbf{w}^T\mathbf{e_i} + b - y_i \leq \varepsilon + \xi_i \\ y_i - \mathbf{w}^T\mathbf{e_i} - b \leq \varepsilon + \xi_i{}^* \\ \xi_i, \xi_i{}^* \geq 0, i = 0...l \end{cases} \end{aligned} \tag{17}$$

In the SVR, nonlinear mappings can be achieved through the employment of kernels, thus the regression procedure is presented in Equation 18, where α and α^* are Lagrange multipliers and $k(\mathbf{e}_i, \mathbf{e})$ is a kernel function.

$$f(\mathbf{e}) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(\mathbf{e}_i, \mathbf{e}) + b \tag{18}$$

In this work, the radial basis function (RBF) kernel (also known as Gaussian kernel) is employed ($k(\mathbf{e}_i, \mathbf{e}_j) = \exp(\frac{-\|\mathbf{e}_i - \mathbf{e}_j\|^2}{\gamma^2})$) where γ is a parameter of the RBF kernel.

4. Simulations and experimental results

The database that is used in this work is composed of six times series: Canadian **Lynx**, Wolf's **Sunspot**, British Pound/US Dollar **Exchange Rate**, **Colorado River** flow, **Airline** Passengers and **Star** Brightness. These time series are from different application areas and have distinct statistical characteristics [5, 10, 14, 15]. Linear, non-linear and hybrid models have been employed to forecast these time series in several works [5, 10, 14, 15, 24, 28].

The Lynx time series contains the annual number of lynxes trapped in Northern Canada. This dataset has 114 observations between the years of 1821 and 1934. The Sunspot series is composed of the annual records of incidence of spots on the sun surface between the years of 1700 and 1987, totalizing 288 samples. The Exchange Rate time series consists of the weekly average of the British pound/US dollar exchange rate between the years of 1980 and 1993, totalizing 731 data points. The Colorado River time series is composed of the monthly record of Colorado River Lees Ferry flow between the years of 1911 and 1972, totalizing 744 samples. The Airline series corresponds to the monthly international airline passengers (in thousands). This dataset has 144 observations between January 1949 to December 1960. The Star series is composed of the daily star brightness observations (at the same place and hour), totalizing 600 points.

The Lynx, Sunspot, Colorado River and Airline datasets are available in [34], the Exchange Rate is available in the website of Federal Reserve Bank of St. Louis[2], and Star is available in the time series archive maintained by

[2]https://fred.stlouisfed.org/series/DEXUSUK

the University of York, England[3].

Table 1: Division of the time series used in the experimental evaluation.

Time Series	Size			Percentage (%)	
	Sample	**Training set**	**Test set**	**Training set**	**Test set**
Lynx	114	100	14	88	12
Sunspot	288	221	67	77	23
Exchange Rate	731	679	52	93	7
Colorado River	744	595	149	80	20
Airline	144	115	29	80	20
Star	600	480	120	80	20

Table 1 shows the sample size and the percentage division for the training and test sets [5, 10, 15]. The test set is composed of the last values of the time series, and the remainder is used as training set. The logarithms (to the base 10) and the natural logarithmic are applied in the data of Canadian Lynx and Exchange Rate, respectively. The result analysis is performed using three well-used performance measures from the literature [7, 10, 15, 35]: Mean Square Error (MSE) (Equation 19), Mean Absolute Percentage Error (MAPE) (Equation 20) and Mean Absolute Error (MAE) (Equation 21), where N is the dataset size, y_t is the actual value at time t and $\hat{y}_t$ is the forecast value at time t. For these metrics, lower values represent better accuracy.

$$\mathrm{MSE} = \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2, \tag{19}$$

$$\mathrm{MAPE} = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right|, \tag{20}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t|. \tag{21}$$

All of the performance measures are based on one-step ahead forecasts from the models. The proposed model is compared against known single methods [10, 14, 15] (ARIMA and MLP) and hybrid methods from the literature [5, 10, 14, 15, 19, 24, 28]. In this work, a percentage comparison (PC)

[3] http://www.york.ac.uk/depts/maths/data/ts/welcome.htm (series number 26)

is also performed to measure the gain/loss reached by the proposed hybrid system in relation to models of the literature. The comparison is performed in terms of some evaluation measure following Equation 22

$$\mathrm{PC} = \frac{(\mathrm{Metric_{lm}} - \mathrm{Metric_{hs}})}{\mathrm{Metric_{lm}}}\%, \tag{22}$$

where Metric_{lm} and Metric_{hs} represent the values of a given performance metric reached by literature models and the proposed hybrid system, respectively. Thus, the greater the PC, the better is the performance of the proposed hybrid system.

For a better visualization and understanding of the results, we adopted a nomenclature to indicate the literature models and the proposed hybrid system. Table 2 shows the nomenclature of the proposed and literature approaches used in this work. In this work, for the sake of comparison, the NoLiC method is used with the same forecasters of the proposed hybrid system (ARIMA, MLP, and SVR) instead of the hybrid intelligent system used in the original article [7]. So, the best combination is selected between the two versions based on MSE value for each time series.

Table 2: Nomenclature of the proposed method and literature models used in this work.

Approach	Reference	Models
Single model	Zhang [10]	ARIMA
	Zhang [10]	MLP
Linear combination	Zhang [10]	ARIMA + MLP
	Panigrahi and Behera [17]	ETS + MLP
	de Oliveira and Ludermir [28]	ARIMA + SVR
Nonlinear combination	Khashei and Bijari [14]	ARIMA and MLP
	Khashei and Bijari [15]	ARIMA and MLP
	NoLiC [7]	ARIMA, SVR and MLP
Linear combination with moving-average filter	Babu and Reddy [19]	ARIMA + MLP
Linear combination with exponential smoothing filter	de Oliveira and Ludermir [24]	ARIMA + SVR
Proposed hybrid system	$\mathrm{MLP_{(A,S)}}$	MLP(ARIMA,SVR)
	$\mathrm{SVR_{(A,M)}}$	SVR(ARIMA,MLP)

Table 3 shows the parameter selection for the components of the proposed hybrid system. An automatic stepwise approach [36] is employed for the model selection for the ARIMA(p,d,q) in $\mathrm{M_L}$, which is a methodology that generates the same values used in [10, 14, 15]. Hyper-parameter selection for

SVR and MLP in M_{NL} and M_C is performed through a grid search approach. In case of the input selection for SVR and MLP as M_C model of proposed system, the integer values are chosen in the range $[1, Lmax]$. For each time series, the cross-correlation between the time series and its error series of the linear model (ARIMA) is analyzed to establish the maximum lag ($Lmax$) of the search. Considering the temporal correlation presented in each time series and the size of the data sets, in this work the cross-correlation is performed in the range $[1, 20]$, since values higher than 20 lags were evaluated but did not present relevant correlation to be considered. In this way, the maximum value of $Lmax$ is 20. Then, the significant lags of the past forecasts from time series and error series are used based on simulation in the range $[1, Lmax]$.

For example, if the value is equal to 1, then M_C model receives as input one forecast of each model (M_L and M_{NL}), totalizing 2 inputs. If the value is equal to 8, then M_C performs the final forecasting of the hybrid system based on 16 inputs data, 8 forecasts of M_L and 8 forecasts of M_{NL}. After parameter selection, the models are trained and tested 30 times, and the configuration with lower MSE is selected, for subsequent analysis [10, 14, 15].

Table 3: Values of the parameters for the models.

Model	Parameters	Values
ARIMA	p,d,q	Hyndman [36] Methodology
MLP	Input (Lags) - M_{NL}	[2, 24]
	Input (Lags) - M_C	[1, $Lmax$]
	Algorithm	Backpropagation
	Activation Function	Sigmoid
	Nodes in hidden layer	2, 5, 10, 15, 20
SVR	Input (Lags) - M_{NL}	[2, 24]
	Input (Lags) - M_C	[1, $Lmax$]
	Kernel	RBF
	γ	1,0.1,0.01,0.001
	C	0.1,1,100,1000,10000
	ε	0.1, 0.01, 0.001

Subsections 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 show the results obtained for Lynx, Sunspot, Exchange Rate, Colorado River, Airline and Star series, respectively.

4.1. Canadian Lynx time series

Figure 3 shows the MSE values achieved by different configurations of the proposed hybrid system for Lynx series. In this figure, the sensitivity analysis with MLP and SVR models shows the impact of the variation of time lags in the accuracy of the proposed hybrid system. Figure 3 shows that for the MLP model, the smallest MSE value was reached with 12 inputs, six forecasts of the time series performed by M_L model and six forecasts of the residuals generated by M_{NL} model. The best performance with SVR model was obtained with two time lags, resulting in 4 inputs (two forecasts of each model, M_L and M_{NL} models). For both combination models, the total of used time lags is composed of the forecasts of M_L and M_{NL}. From comparison between two best combination models, $MLP_{(A,S)}$ attained the smallest MSE value.

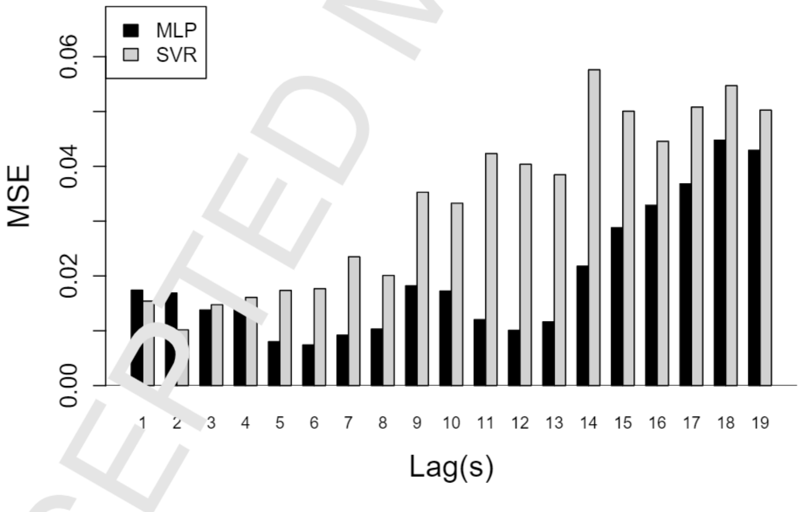


Figure 3: Sensitivity analysis of the number of input lags in terms of MSE values with the proposed hybrid system for test set of the Lynx time series.

Figure 4 sketches the forecasting for Lynx test set with initial model ARIMA, and the best configuration of the proposed hybrid system, $MLP_{(A,S)}$. It can be seen that the forecast of the ARIMA model was improved by the proposed combination method in almost every test point.

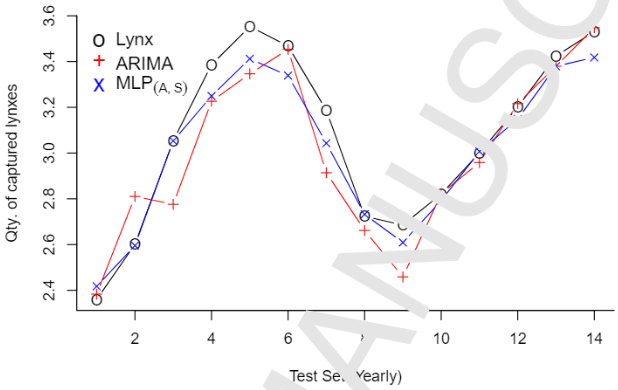


Figure 4: Forecasting for Lynx test set with ARIMA and MLP$_{(A,S)}$.

Table 4 shows the evaluation metrics MSE, MAE and MAPE reached by proposed hybrid system and models of the literature for Lynx test set. The performance results show that among analyzed models, the proposed hybrid system reached the best values for the three metrics.

Among different approaches, ARIMA and MLP models reached the same MSE values between single models, de Oliveira and Ludermir [24] reached the best performance among hybrid systems with linear combination, and Khashei and Bijari [15] obtained the best accuracy between nonlinear combination hybrid systems.

From comparison of the proposed hybrid system with single and hybrid models of the literature, the version MLP$_{(A,S)}$ obtained a percentage gain in terms of MSE and MAE: of 64.21% and 39.78% over MLP, of 49.30% and 28.49% over de Oliveira and Ludemir [24], and over Khashei and Bijari model [15] of 20.26% and 20.58%, respectively.

4.2. Sunspot time series

Figure 5 shows the performance in terms of MSE of the 34 configurations of the proposed hybrid system for the test set of the Sunspot time series. It can be seen that the accuracy of the MLP and SVR models vary according to number of input lags used in the combination performed by hybrid system.

Table 4: Performance of the best configuration of the proposed hybrid system and other models found in the literature for Lynx series. The two best values for each evaluation metric are highlighted, in decreasing order, in bold and underlined.

Approach	Model	MSE	MAE	MAPE
Single Models	ARIMA [10, 14, 15]	0.0204	0.1122	-
	MLP [10, 14, 15]	0.0204	0.1121	-
Hybrid System (Linear combination)	Zhang [10]	0.0172	0.1039	-
	de Oliveira and Ludermir [28]	0.0204	0.1213	4.02
	Babu and Reddy [19]	0.0187	0.1250	4.20
	de Oliveira and Ludermir [24]	0.0144	0.0944	3.10
	Panigrahi and Behera [17]	0.0294	0.1381	4.77
Hybrid System (Nonlinear combination)	Khashei and Bijari [14]	0.0136	0.0896	-
	Khashei and Bijari [15]	<u>0.0099</u>	0.0850	-
	NoLiC [7]	0.0151	0.0988	3.24
Proposed Hybrid System	$\mathrm{MLP}_{(\mathrm{A,S})}$	**0.0073**	**0.0675**	**2.10**
	$\mathrm{SVR}_{(\mathrm{A,M})}$	0.0101	<u>0.0824</u>	<u>2.69</u>

Figure 5 shows that the MLP model reached the best performance using 24 inputs, 12 forecasts of the model $\mathrm{M_L}$ and 12 forecasts of the model $\mathrm{M_{NL}}$. The best performance obtained by SVR is reached with six inputs, three forecasts of the model $\mathrm{M_L}$ and three forecasts of the model $\mathrm{M_{NL}}$. From comparison between two best combination models, $\mathrm{SVR}_{(\mathrm{A,M})}$ obtained the best MSE value.

Figure 6 shows the forecasting for Sunspot test set using $\mathrm{SVR}_{(\mathrm{A,M})}$ and ARIMA model. It can be seen, that the proposed system was able to enhance the accuracy of the ARIMA model. The forecasting of the $\mathrm{SVR}_{(\mathrm{A,M})}$ is closer of the actual series than initial linear model.

Table 5 shows the forecasting results for Sunspot series in terms of MSE, MAE and MAPE for two periods of forecasting. It can be seen that the proposed hybrid system reached more accurate results than other models of the literature for all of the evaluated scenarios.

Among the models from literature: MLP obtained the best MSE and MAE for 35 points ahead, and ARIMA reached the best metric results for 64 points ahead between single models; Zhang [10] reached the best MSE in both periods among hybrid systems with linear combination; Khashei and Bijari model [14] obtained the best MSE for 35 points ahead, and Khashei and Bijari [15] found the best MSE and MAE for 64 points ahead among

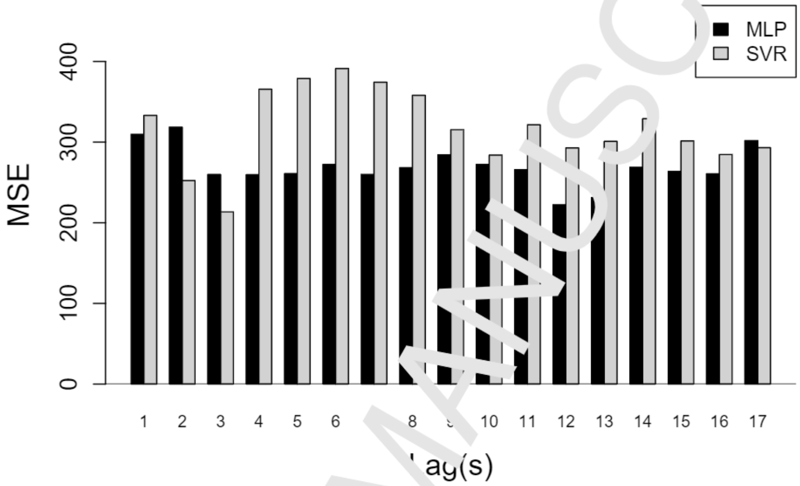


Figure 5: Sensitivity analysis of the number of input lags with the proposed hybrid system regarding MSE values for Sunspot series.

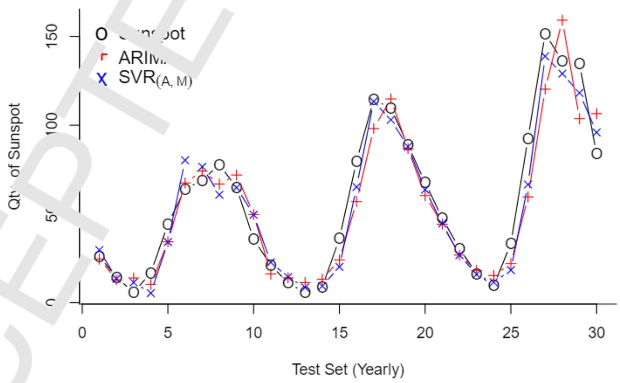


Figure 6: Forecasting for first thirty points of the test set for the Sunspot series with with ARIMA and $SVR_{(A,M)}$ models.

Table 5: Performance of the best configuration of the proposed hybrid system and other models found in the literature for Sunspot series. The two best values for each evaluation metric are highlighted, in decreasing order, in bold and underlined.

Approach	Model	35 points ahead			64 points ahead		
		MSE	MAE	MAPE	MSE	MAE	MAPE
Single Models	ARIMA [10, 14, 15]	216.9	11.[illegible]	-	306.0	13.0	-
	MLP [10, 14, 15]	205.3	10.[illegible]	-	351.1	13.5	-
Hybrid System (Linear combination)	Zhang [10]	186.8	[illegible].8		280.1	12.7	-
	de Oliveira and Ludermir [28]	189.7	11.4	39.1	306.8	13.2	40.3
	Babu and Reddy [19]	174.8	10.1	33.0	300.4	13.1	34.0
	de Oliveira and Ludermir [24]	195.4	10.8	34.9	300.4	13.1	37.2
	Panigrahi and Behera [17]	1[illegible].0	10.[illegible]	32.8	312.0	13.2	35.6
Hybrid System (Nonlinear combination)	Khashei and Bijari [14]	<u>125.8</u>	[illegible].9	-	234.2	12.1	-
	Khashei and Bijari [15]	[illegible]	[illegible]	-	<u>218.6</u>	11.4	-
	NoLiC [7]	[illegible]7.3	9.1	30.3	308.8	13.4	34.2
Proposed Hybrid System	$\text{MLP}_{(\text{A,S})}$	<u>105.2</u>	**7.4**	<u>26.5</u>	222.4	<u>11.1</u>	**27.4**
	$\text{SVR}_{(\text{A,M})}$	[illegible]1.1	<u>7.8</u>	**25.0**	**213.4**	**10.5**	<u>27.6</u>

hybrid systems with nonlinear combination.

From comparison of the proposed hybrid system with single and hybrid models of the literature, regarding 35 points ahead, the version $\text{SVR}_{(\text{A,M})}$ obtained a percentage gain in terms of MSE and MAE of: 50.75% and 23.52% over MLP, 45.87% and 2[illegible].77% over Zhang [10], and 19.63% and 12.35% over Khashei and Bijari model [14], respectively. For 64 points ahead, the version $\text{SVR}_{(\text{A,M})}$ obtained a percentage gain in terms of MSE and MAE of: 66.96% and 40.0% over ARIMA, 6[illegible].90% and 38.58% over Zhang [10], and 53.75% and 31.57% over Khashei and Bijari model [15], respectively.

4.3. Exchange Rate time series

Figure 7 shows the MSE values obtained by different configurations of the proposed hybrid system for Exchange Rate series. The best performance with MLP model used 12 time lags, six forecasts of the time series performed by M_{L} model, and six forecasts of the residual generated by M_{NL} model. For SVR model, the best MSE value was achieved using two forecasts of each model (M_{L} and M_{NL}). From comparison between two best combination models, $\text{MLP}_{(\text{A,S})}$ obtained the best MSE value.

Figure 8 shows the forecasting of the $\text{MLP}_{(\text{A,S})}$ and ARIMA models for Exchange Rate test set. It is possible to observe that the $\text{MLP}_{(\text{A,S})}$ model improves the ARIMA forecasting. In this case the ARIMA forecast is similar to Random Walk model.

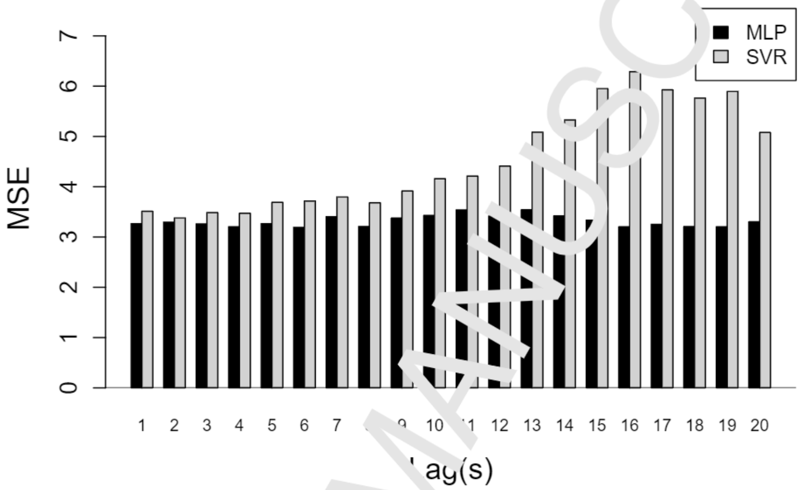


Figure 7: Sensitivity analysis of the number of input lags with the proposed hybrid system regarding MSE values for Exchange Rate series.

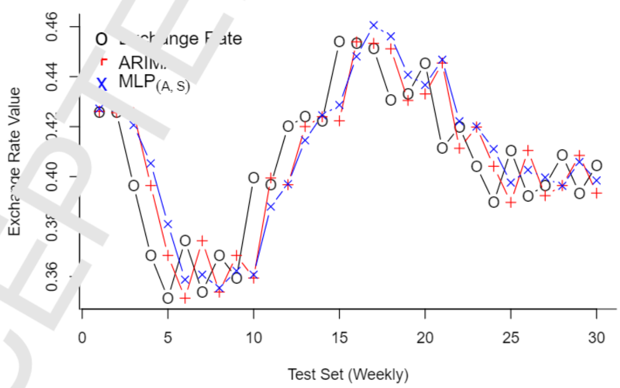


Figure 8: Forecasting for 30 first points in Exchange Rate test set with ARIMA and $\mathrm{MLP}_{(A,S)}$

Table 6 shows the metrics evaluation for Exchange Rate test set. It is possible to observe that among analyzed models, the proposed hybrid system reached the best result in terms of MSE, MAE and MAPE for the time horizon of 6 and 12 months.

Table 6: Performance of the best configuration of the proposed hybrid system and other models found in the literature for Exchange Rate series. The two best values for each evaluation metric are highlighted, in decreasing order, in bold and underlined.

Approach	Model	1 month			6 months			12 months		
		MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE
Single Models	ARIMA [10, 14, 15]	3.6849	0.0050	-	5.6574	[illegible]	-	4.5297	0.0053	-
	MLP [10, 14, 15]	2.7637	0.0042	-	[illegible]	0.0059	-	4.5265	0.0052	-
Hybrid System (Linear combination)	Zhang [10]	2.6725	0.0041	-	[illegible]	0.0058	-	4.3590	0.0051	-
	de Oliveira and Ludermir [28]	3.9244	0.0048	[illegible]	[illegible]	0.0057	3.28	3.5183	0.0047	<u>2.73</u>
	Babu and Reddy [19]	3.4900	[illegible]	2.46	4.5562	0.0057	3.28	3.7285	0.0049	2.84
	de Oliveira and Ludermir [24]	2.9133	0.0039	[illegible]	4.2952	0.0053	3.09	3.5944	0.0048	2.77
	Panigrahi and Behera [17]	**1.8008**	[illegible]	[illegible]	<u>3.9465</u>	<u>0.0051</u>	<u>2.95</u>	3.5313	0.0048	2.75
Hybrid System (Nonlinear combination)	Khashei and Bijari [14]	2.6093	0.0040	-	4.3164	0.0054	-	3.7639	0.0051	-
	Khashei and Bijari [15]	[illegible]	[illegible]	-	4.2782	0.0053	-	3.6477	0.0049	-
	NoLiC [7]	[illegible]	[illegible]	2.64	4.1081	0.0053	3.08	<u>3.2641</u>	0.0046	2.65
Proposed Hybrid System	$MLP_{(A,S)}$	[illegible]	[illegible]	4.13	3.9581	0.0051	2.97	**3.1904**	**0.0045**	**2.58**
	$SVR_{(A,M)}$	[illegible]	[illegible]	3.43	**3.8425**	**0.0050**	**2.93**	3.3783	<u>0.0047</u>	<u>2.73</u>

Considering the approaches in the literature, MLP reached the best results for 1 and 12 months between single models; Panigrahi and Behera [17] reached the best performance for 1 and 6 months ahead, and de Oliveira and Ludermir [28] were the best model for 12 months ahead among hybrid systems with linear combination; Khashei and Bijari [15] obtained the best result in all time horizon among nonlinear hybrid systems.

From comparison of the proposed hybrid system with single and hybrid models of the literature, the $MLP_{(A,S)}$ and $SVR_{(A,M)}$ versions were not able to improve the ARIMA forecasting for a time horizon of 1 month. For 6 months the $SVR_{(A,M)}$ reached a percentage gain in terms of MSE and MAE of: 32.[illegible]% and 16.29 over MLP, 2.63%, and 1.34% over Panigrahi and Behera [17], and 10.18% and 5.19% over Khashei and Bijari [15], respectively. For 12 months, the proposed model $MLP_{(A,S)}$ obtained a percentage gain in terms of MSE and MAE of: 29.51% and 13.99 over MLP, 9.31%, and 5.56%

over de Oliveira and Ludermir [28], and 12.53% and 9.11% over Khashei and Bijari [15], respectively.

4.4. *Colorado River time series*

Figure 9 shows the MSE values achieved by different configurations of the proposed hybrid model for Colorado River series. The best performance with MLP model used sixteen inputs, eight time lags of the forecasted values from the time series performed by M_L model and eight residual forecasts provided by M_{NL} model. For SVR, the best MSE value was obtained using 32 inputs, where 16 are time lags from M_L and 16 are time lags from M_{NL}. From the comparison between two best combination models, $SVR_{(A,M)}$ obtained the best results in all evaluation metrics.

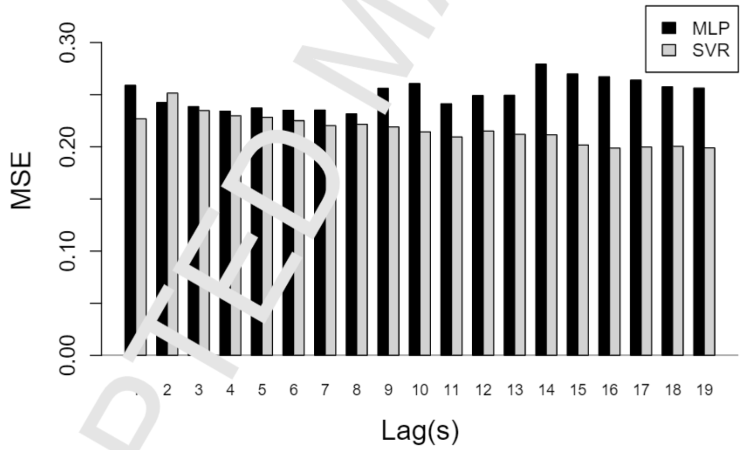


Figure 9: Sensitivity analysis of the number of input lags with the proposed hybrid system regarding MSE values for Colorado River series.

Figure 10 shows the forecasting for the Colorado River test set with the ARIMA model and the best configuration of the proposed hybrid system, $SVR_{(A,M)}$.

Table 7 shows the forecasting results for Colorado River test set in terms of MSE, MAE, and MAPE. It can be seen that the proposed hybrid system reached the best MSE and MAE values.

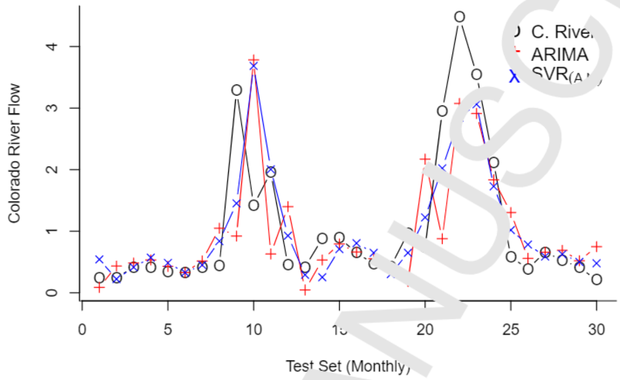


Figure 10: Forecasting for 30 first points in Colorado River test set with ARIMA and $SVR_{(A,M)}$.

Table 7: Performance of the best configuration of the proposed hybrid system and other models found in the literature for Colorado River series. The two best values for each evaluation metric are highlighted, in decreasing order, in bold and underlined.

Approach	Model	MSE	MAE	MAPE
Single Models	ARIMA [10, 14, 15]	0.2869	0.2879	75.44
	MLP [10, 14, 15]	0.2928	0.3114	116.8
Hybrid System (Linear combination)	Zhang [10]	0.2230	<u>0.2559</u>	65.16
	de Oliveira and Ludermir [28]	0.2867	0.2810	**32.56**
	Babu and Reddy [19]	0.2376	0.2917	49.34
	de Oliveira and Ludermir [24]	0.2612	0.3167	<u>33.95</u>
	Panigrahi and Behera [17]	0.2363	0.2615	49.52
Hybrid System (Nonlinear combination)	Khashei and Bijari [14]	0.2515	0.2862	101.9
	Khashei and Bijari [15]	<u>0.2214</u>	0.2633	75.53
	NoLiC [7]	0.2268	0.2657	62.63
Proposed Hybrid System	$MLP_{(A,S)}$	0.2316	0.2720	86.17
	$SVR_{(A,M)}$	**0.1987**	**0.2477**	72.62

Among the models from the literature: ARIMA model found the best results in terms of MSE, MAE, and MAPE between single models; Zhang [10] reached the best performance in terms of MSE and MAE among hybrid systems with linear combination; Khashei and Bijari model [15] found the best results in terms of MSE and MAE between hybrid systems with nonlinear combination.

From comparison of the proposed hybrid system with single and hybrid models of the literature, the version $SVR_{(A,M)}$ obtained a percentage gain in terms of MSE and MAE of: 30.75% and 13.90% over ARIMA, 10.91% and 3.22% over Zhang [10], and 10.26% and 5.94% over Khashei and Bijari model [15], respectively.

4.5. Airline time series

Figure 11 shows the MSE values achieved by different configurations of the proposed hybrid system for Airline series. This figure shows the impact of the variation of time lags in the accuracy of the proposed hybrid system. The best performance for MLP model was reached using 40 inputs, where 20 forecasts were provided by each model (M_L and M_{NL}). The best result with SVR model was obtained with six inputs, three forecasts of the M_L model and three forecasts of the M_{NL}. From comparison between two combination models, $MLP_{(A,S)}$ reached the best MSE value.

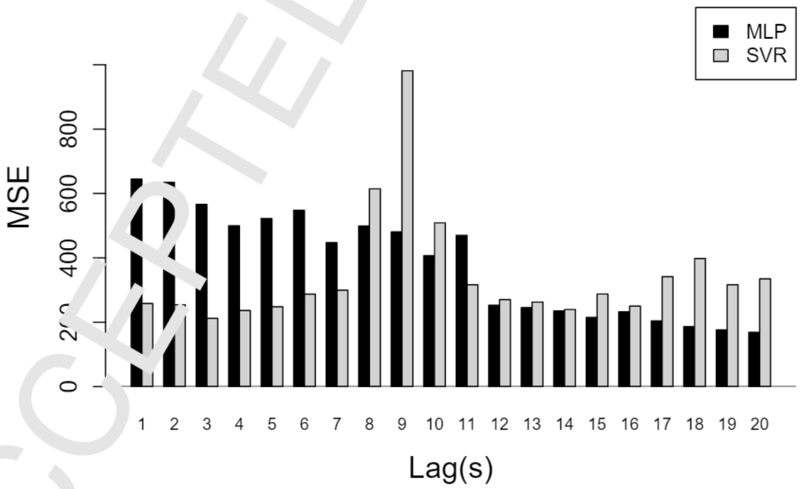


Figure 11: Sensitivity analysis of the number of input lags with the proposed hybrid system regarding MSE values for Airline series.

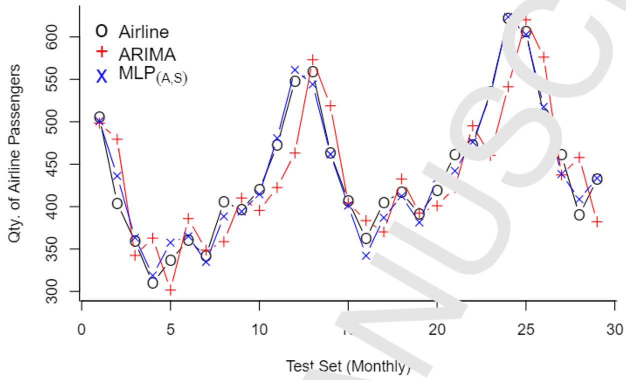


Figure 12: Forecasting for Airline test set with ARIMA and $\mathrm{MLP}_{(\mathrm{A,S})}$.

Figure 12 shows the forecasting for Airline test set using ARIMA model and $\mathrm{MLP}_{(\mathrm{A,S})}$. This figure shows that the proposed hybrid system was able to improve the performance of the initial model.

Table 8: Performance of the best configuration of the proposed hybrid system and other models found in the literature for Airline series. The two best values for each evaluation metric are highlighted, in decreasing order, in bold and underlined.

Approach	Model	MSE	MAE	MAPE
Single Models	ARIMA [10, 14, 15]	1918.6	36.02	8.15
	MLP [10, 14, 15]	507.7	18.12	4.25
Hybrid System (Linear combination)	Zhang [10]	485.7	16.59	3.93
	de Oliveira and Ludermir [28]	388.9	16.08	3.71
	Babu and Reddy [19]	793.3	22.05	4.90
	de Oliveira and Ludermir [24]	405.4	16.77	3.81
	Panigrahi and Behera [17]	400.3	15.95	3.85
Hybrid System (Nonlinear combination)	Khashei and Bijari [14]	253.3	13.08	3.01
	Khashei and Bijari [15]	258.0	13.57	3.14
	NoLiC [7]	257.9	12.76	2.97
Proposed Hybrid System	$\mathrm{MLP}_{(\mathrm{A,S})}$	**168.5**	**10.38**	**2.49**
	$\mathrm{SVR}_{(\mathrm{A,M})}$	<u>211.9</u>	<u>10.96</u>	<u>2.57</u>

Table 8 shows the performance for Airline series in terms of MSE, MAE and MAPE. It can be observed that the two versions of the proposed hybrid system reached more accurate results than other models of the literature. The combination $MLP_{(A,S)}$ obtained the best accuracy between proposed versions.

Among the models from literature: MLP found the best metric results between single models; de Oliveira and Ludermir [28] reached the best performance among hybrid systems with linear combination; NoLiC [7] reached the best MAE and MAPE values among hybrid systems with nonlinear combination.

From comparison of the proposed hybrid system with single and hybrid models of the literature, the version $MLP_{(A,S)}$ obtained a percentage gain in terms of MSE, MAE and MAPE of: 66.[illegible]%, 42.72% and 41.41% over MLP, 56.67%, 35.44% and 32.88% over de Oliveira and Ludermir [28], and 34.66%, 18.65% and 16.16% over NoLiC [7], respectively.

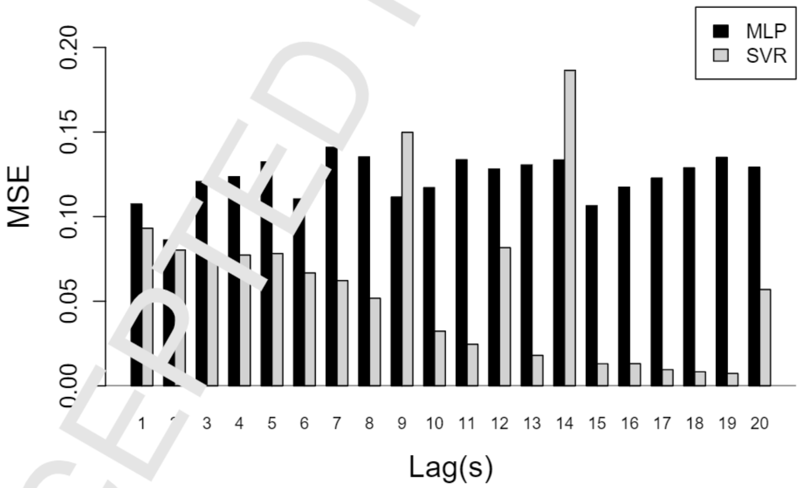


Figure 1[illegible]: Sensitivity analysis of the number of input lags with the proposed hybrid system regarding MSE values for Star series.

4.6. Star time series

Figure 13 shows the performance in terms of MSE of the 40 configurations of the proposed hybrid system for Star time series. It can be seen that the accuracy of the models varies according to the number of input lags used in the M_C model. Figure 13 shows that the MLP model reached the best performance using four inputs, two forecasts of the model M_L and two forecasts of the model M_{NL}. For SVR, the best performance is reached with nineteen lags of the model M_L and nineteen lags of the model M_{NL}. From comparison between two best combination models $SVR_{(A,M)}$ attained the smallest MSE value.

Figure 14 shows the forecasting for Star test set with linear model ARIMA and the best configuration of the proposed hybrid system, $SVR_{(A,M)}$.

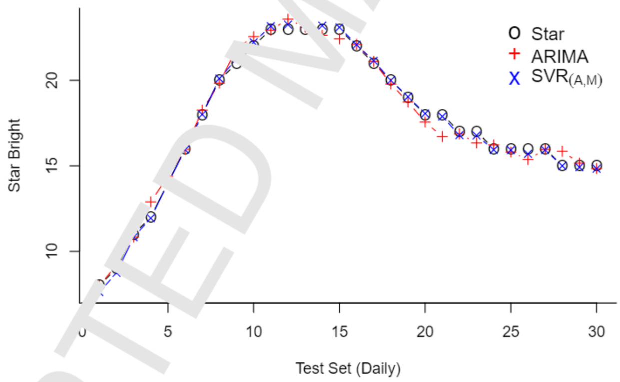


Figure 14: Forecasting for 30 first points in Star test set with ARIMA and $SVR_{(A,M)}$.

Table 9 shows the evaluation metrics MSE, MAE and MAPE reached by proposed hybrid system and models of the literature for Star test set. The proposed hybrid system ($SVR_{(A,M)}$) found the best performance in all metrics.

From the analysis of the literature models, MLP reached the best results between single models, de Oliveira and Ludermir [28] found the best performance among hybrid systems with linear combination, and among nonlinear combination hybrid systems, NoLiC [7] obtained the best accuracy.

From comparison of the proposed hybrid system with single and hybrid models of the literature, the version $SVR_{(A,M)}$ reached a percentage gain in terms of MSE, MAE and MAPE of: 95.01%, 78.45% and 81.92% over MLP, 92.42%, 72.97% and 79.02% over de Oliveira and Ludermir [28], and 92.26%, 72.71% and 78.87% over NoLiC method [7].

Table 9: Performance of the best configuration of the proposed hybrid system and other models found in the literature for Star series. The two best values for each evaluation metric are highlighted, in decreasing order, in bold and underlined.

Approach	**Model**	**MSE**	**MAE**	**MAPE**
Single Models	ARIMA [10, 14, 15]	0.2418	0.3871	4.43
	MLP [10, 14, 15]	0.1445	0.3110	3.32
Hybrid System (Linear combination)	Zhang [10]	0.0951	0.2479	2.86
	de Oliveira and Ludermir [28]	<u>0.0512</u>	<u>0.1750</u>	3.55
	Babu and Reddy [19]	0.0972	0.2479	4.10
	de Oliveira and Ludermir [24]	0.3331	0.4682	6.18
	Panigrahi and Behera [17]	0.1987	0.3424	7.02
Hybrid System (Nonlinear combination)	Khashei and Bijari [14]	0.1482	0.3196	3.65
	Khashei and Bijari [15]	0.1542	0.3249	3.51
	NoLiC [7]	0.0931	0.2456	2.84
Proposed Hybrid System	$MLP_{(A,S)}$	0.0861	0.2325	<u>2.72</u>
	$SVR_{(A,M)}$	**0.0072**	**0.0670**	**0.60**

5. Discussion

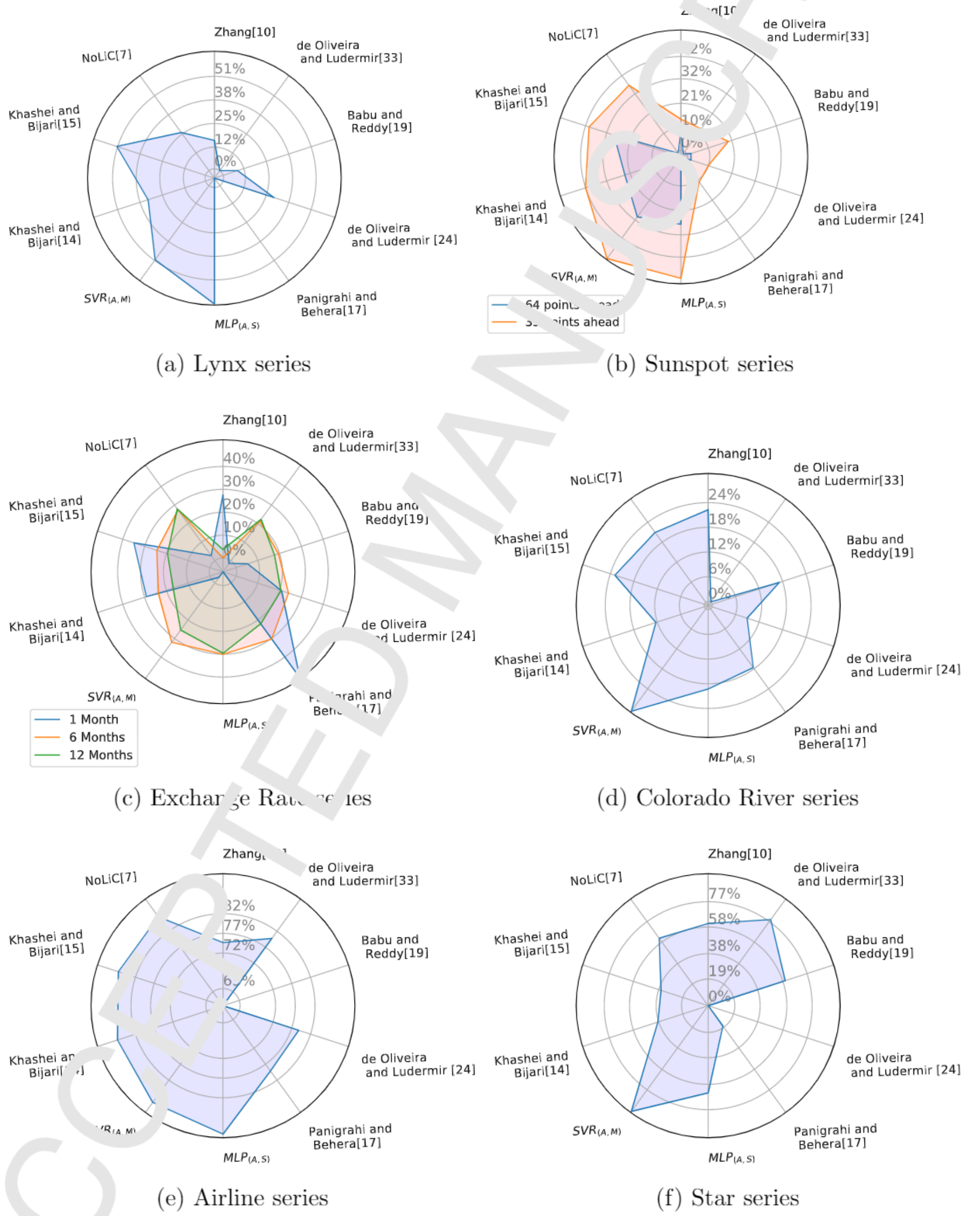
This work proposes a hybrid system that performs time series forecasting in three steps: time series modeling using M_L model; residuals forecasting with M_{NL} model; and the combination of the forecasts of M_L and M_{NL} using the M_C model. M_C is an intelligent model that searches the most suitable function to combine the forecasts of M_L and M_{NL}. The traditional ARIMA model was adopted as M_L, and two intelligent techniques — MLP neural network and SVR — were used alternately as M_{NL} and M_C, generating two versions of the proposed hybrid system, $MLP_{(A,S)}$ and $SVR_{(A,M)}$. However, other intelligent models could be used to search for a more suitable combination function.

In the time lag sensitivity analysis, presented in Figures 3, 5, 7, 9, 11 and 13, it can be seen that the MSE values obtained by the proposed method vary as a function of the number of lags. In fact, the number of delays plays an essential role in time series forecasting systems [37]. As can be observed from the analysis, in general, the SVR presented higher variations in MSE due to its hyper-parameter sensitivity. On the other hand, the MLP is less sensitive to parameters variation. Lower values of MSE indicate a possible number of lags most suitable to be employed in the forecasting performed by M_C.

Table 10 shows the ranking obtained by versions of the proposed hybrid system for each study case. The ranking takes into account the performance regarding the jointly used measures in the evaluation. In light of the adopted performance measures, comparing with single and hybrid models from literature, at least one of the proposed hybrid systems achieved the best performance in terms of MSE, MAE and MAPE for Lynx, Sunspot, Airline and Star time series. For Colorado River, the version $SVR_{(A,M)}$ reached the best MSE and MAE values. So, in our comparison of the two versions of the proposed hybrid system, regarding all study cases, we have observed that $SVR_{(A,M)}$ reached the best result in five out of nine cases and the second rank in three out nine cases. $MLP_{(A,S)}$ reached the best rank in three out of nine cases and the second rank in two out of nine cases. In the case of Exchange Rate, the proposed system did not obtain relevant results in the first month of the test set, but considering 6 and 12 months the proposed systems achieved the best metric results. Therefore, the proposed hybrid system was not able to improve the forecasting of the first points of the Exchange Rate test set, worsening the forecasting of the ARIMA model. However, when the overall performance is analyzed, the proposed system reaches better results than the models in the literature.

Furthermore, the $MLP_{(A,S)}$ achieved best results in times series that present trend patterns such as Lynx and Airline. The $SVR_{(A,M)}$ outperformed the $MLP_{(A,S)}$ in time series which presents seasonal/cyclic patterns, as can be observed in Sunspot, Colorado River and Star data set. A performance analysis of SVR and MLP methods based on time series patterns (trend, seasonality and their variations) was also carried in [38] and their results corroborate with the accuracy obtained by the two versions of the proposed hybrid system.

Figure 15 shows radar charts of the percentage difference in terms of MSE between hybrid systems and ARIMA model for Canadian Lynx, Sunspot, Exchange Rate, Colorado River, Airline and Star series. Each corner of radar represents a hybrid system. The closer the line is to the corresponding



(a) Lynx series (b) Sunspot series

(c) Exchange Rate series (d) Colorado River series

(e) Airline series (f) Star series

Figure 15: Percentage difference in terms of MSE between hybrid systems and ARIMA model for all data base.

Table 10: Ranking of the versions of the proposed hybrid system. Lower ranks represent better results, considering all metric values used for each study case.

Data set		Ranking	
		$MLP_{(A,S)}$	$SVR_{(A,M)}$
Lynx		**1**	2
Sunspot	35 points ahead	2	**1**
	64 points ahead	2	**1**
Exchange Rate	1 month	12	11
	6 month	3	**1**
	12 month	[illegible]	2
Colorado River		5	**1**
Airline		**1**	2
Star		3	**1**

corner, the better is the obtained percentage enhancement in relation to ARIMA model.

Figure 15a shows that in the Canadian Lynx series, the $MLP_{(A,S)}$ obtained the best percentage gain in relation to the ARIMA model, followed by Khashei and Bijari [15], and $SVR_{(A,M)}$, respectively. This figure also shows that Panigrahi and Behera [17], and de Oliveira and Ludermir [28] did not reach a percentage improvement in relation to ARIMA model.

Figure 15b shows, that for the first 35 points ahead in the Sunspot time series, all hybrid systems reached a positive percentage gain. $SVR_{(A,M)}$ obtained the highest value followed by $MLP_{(A,S)}$, and Khashei and Bijari [14], respectively. For the 64 points ahead, $SVR_{(A,M)}$ also achieved the highest value. Panigrahi and Behera [17], de Oliveira and Ludermir [28], and NoLiC [7] did not reach a percentage improvement in relation to ARIMA model.

Figure 15c shows the results in the Exchange Rate series for 1 month, 6 months and 12 months ahead. For one month, Panigrahi and Behera [17] obtained the best percentage improvement, followed by Khashei and Bijari [15], and Khashei and Bijari [14], respectively. For this study case, both proposed systems did not achieve improvements. For six months, all hybrid systems obtained a percentage improvement in relation to the ARIMA model. For this case, $SVR_{(A,M)}$ reached the best percentage improvement, followed by Panigrahi and Behera [17] and $MLP_{(A,S)}$, respectively. For twelve months,

$\mathrm{MLP}_{(\mathrm{A,S})}$ obtained the best percentage improvement, followed by $\mathrm{SVR}_{(\mathrm{A,M})}$ and Panigrahi and Behera [17].

Figure 15d shows that in the Colorado River series, the $\mathrm{SVR}_{(\mathrm{A,M})}$ obtained the best percentage gain in relation to the ARIMA model, followed by Khashei and Bijari [15], and Zhang [10], respectively. The hybrid system proposed by de Oliveira and Ludermir [28] reached the worse percent of improvement (0.085%).

Figure 15e shows that in Airline series, all hybrid systems reached a positive percentage gain. $\mathrm{MLP}_{(\mathrm{A,S})}$ obtained the highest value followed by $\mathrm{SVR}_{(\mathrm{A,M})}$, and Khashei and Bijari [14].

Figure 15f shows that in the Star series, $\mathrm{SVR}_{(\mathrm{A,M})}$ obtained the best percentage gain in relation to the ARIMA model, followed by de Oliveira and Ludermir [28], and $\mathrm{MLP}_{(\mathrm{A,S})}$. For this series de Oliveira and Ludermir [24] do not reached percentage improvements.

6. Conclusion

In this work, a hybrid system which searches for the most suitable function to combine the linear forecasts of the time series and the nonlinear forecasts of the respective residuals is proposed. The hybrid system performs: the forecasting of the series using a linear model ($\mathrm{M_L}$), error forecasting using a nonlinear model ($\mathrm{M_{NL}}$), and a combination of the forecasts of the series and respective residuals using $\mathrm{M_C}$ (denoted by $\mathrm{M_{C(M_L,M_{NL})}}$). To maximize the accuracy, the proposed system performs a sensitivity analysis of the inputs of $\mathrm{M_C}$ (forecasts of $\mathrm{M_L}$ and $\mathrm{M_{NL}}$) and searches a suitable function to combine these estimates.

Variations of the proposed hybrid system were evaluated considering two traditional nonlinear intelligent models: MLP and SVR. For both configurations, the ARIMA model is used as $\mathrm{M_L}$. Each intelligent model is employed as $\mathrm{M_{NL}}$ and $\mathrm{M_C}$, generating the configurations $\mathrm{MLP}_{(\mathrm{A,S})}$ and $\mathrm{SVR}_{(\mathrm{A,M})}$, where A, M, and S represent the ARIMA, MLP and SVR models, respectively.

An experimental evaluation using three evaluation metrics is performed with well-known time series from the literature: Canadian Lynx, Sunspot, British pound/US dollar Exchange Rate, Colorado River, Airline Passengers and Star Brightness. The experimental results show that the proposed hybrid system reaches a better performance than other systems of the literature [5, 10, 14, 15, 19, 24, 28] in most case studies. Our results suggest that the framework of the proposed hybrid system leads to higher accuracy because it is able to model separately the linear and nonlinear patterns of the data through M_L and M_{NL} components. Moreover, it employs an exclusive step for searching an underlying function that is more suitable to combine these components considering their temporal relationship.

In our future work, we aim to develop a method based on meta-heuristic algorithm to automatically define the parameters and temporal features used by combination model. Hybrid systems that combine ML models should be investigated because there are works in the literature that reach relevant results by employing nonlinear models in all of the steps of the modeling. [7, 16, 18]. In addition, other ML models can be investigated, such as: deep learning [39, 40], echo state networks [41] and decision trees for regression [42] with the objective to improve the combination of the linear and nonlinear models.

Acknowledgments

This work was partially supported by Brazilian agencies: CNPq and Facepe. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We would like to thank professor F. Madeiro for their suggestions and contributions.

References

[1] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[2] G. Zhang, B. E. Patuwo, M. Y. Hu, Forecasting with artificial neural networks: The state of the art, International Journal of Forecasting 14 (1) (1998) 35–62.

[3] P. R. A. Firmino, P. S. de Mattos Neto, T. A. Ferreira, Correcting and combining time series forecasters, Neural Networks 50 (2014) 1–11.

[4] N. Kourentzes, D. K. Barrow, S. F. Crone, Neural network ensemble operators for time series forecasting, Expert Systems with Applications 41 (9) (2014) 4235–4244.

[5] S. Panigrahi, H. Behera, A hybrid ETS–ANN model for time series forecasting, Engineering Applications of Artificial Intelligence 66 (2017) 49–59.

[6] P.-F. Pai, C.-S. Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, Omega 33 (6) (2005) 497–505.

[7] P. S. de Mattos Neto, G. D. Cavalcanti, F. Madeiro, Nonlinear combination method of forecasters applied to PM time series, Pattern Recognition Letters 95 (2017) 65–72.

[8] R. Chandra, S. Chand, Evaluation of co-evolutionary neural network architectures for time series prediction with mobile application in finance, Applied Soft Computing 49 (2016) 462–473.

[9] T.-M. Choi, Y. Yu, K.-F. Au, A hybrid SARIMA wavelet transform method for sales forecasting, Decision Support Systems 51 (1) (2011) 130 – 140.

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

[11] D. Ö. Faruk, A hybrid neural network and ARIMA model for water quality time series prediction, Engineering Applications of Artificial Intelligence 23 (4) (2010) 586–594.

[12] T. Koutroumanidis, K. Ioannou, G. Arabatzis, Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model, Energy Policy 37 (9) (2009) 3627–3634.

[13] P. R. A. Firmino, P. S. de Mattos Neto, T. A. Ferreira, Error modeling approach to improve time series forecasters, Neurocomputing 153 (2015) 242–254.

[14] M. Khashei, M. Bijari, An artificial neural network (p, d, q) model for time series forecasting, Expert Systems with Applications 37 (1) (2010) 479–489.

[15] M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Applied Soft Computing 11 (2) (2011) 2664–2675.

[16] P. S. de Mattos Neto, T. A. Ferreira, A. R. Lima, G. C. Vasconcelos, G. D. Cavalcanti, A perturbative approach for enhancing the performance of time series forecasting, Neural Networks 88 (2017) 114–124.

[17] S. Panigrahi, H. Behera, A hybrid ETS–ANN model for time series forecasting, Engineering Applications of Artificial Intelligence 66 (2017) 49–59.

[18] I. Ginzburg, D. Horn, Combined neural networks for time series analysis, in: Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS'93, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993, pp. 224–231.

[19] C. N. Babu, B. E. Reddy, A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data, Applied Soft Computing 23 (2014) 27–38.

[20] L. Zhou, L. Yu, Y. Wang, Z. Lu, L. Tian, L. Tan, Y. Shi, S. Nie, L. Liu, A hybrid model for predicting the prevalence of schistosomiasis in humans of Qianjiang City, China, PloS One 9 (8) (2014) e104875.

[21] J. Ruiz-Aguilar, I. Turias, M. Jiménez-Come, Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting, Transportation Research Part E: Logistics and Transportation Review 67 (2014) 1–13.

[22] R. Wongsathan, I. Seedadan, A hybrid ARIMA and Neural Networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area, Procedia Computer Science 86 (2016) 273–276.

[23] E. Cadenas, W. Rivera, Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model, Renewable Energy 35 (12) (2010) 2732–2738.

[24] J. F. de Oliveira, T. B. Ludermir, A hybrid evolutionary decomposition system for time series forecasting, Neurocomputing 180 (2016) 27–34.

[25] B. Zhu, Y. Wei, Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology, Omega 41 (3) (2013) 517–524.

[26] T. Taskaya-Temizel, M. C. Casey, A comparative study of autoregressive neural network hybrids, Neural Networks 18 (5) (2005) 781 – 789.

[27] M. Khashei, M. Bijari, Which Methodology is Better for Combining Linear and Nonlinear Models for Time Series Forecasting?, Journal of Industrial and Systems Engineering 4 (4) (2011) 265–285.

[28] J. F. L. de Oliveira, T. B. Ludermir, A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting, in: Intelligent Systems (BRACIS), 2014 Brazilian Conference on, IEEE, 2014, pp. 73–78.

[29] J. M. Bates, C. W. Granger, The combination of forecasts, Journal of the Operational Research Society 20 (4) (1969) 451–468.

[30] R. T. Clemen, Combining forecasts: A review and annotated bibliography, International Journal of Forecasting 5 (4) (1989) 559–583.

[31] X. Qiu, P. N. Suganthan, G. A. Amaratunga, Ensemble incremental learning random vector functional link network for short-term electric load forecasting, Knowledge-Based Systems 145 (2018) 182 – 196.

[32] L. Yu, L. Zhou, L. Tan, H. Jiang, Y. Wang, S. Wei, S. Nie, Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China, PloS One 9 (6) (2014) e98241.

[33] K.-Y. Chen, C.-H. Wang, A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan, Expert Systems with Applications 32 (1) (2007) 254–264.

[34] R. J. Hyndman, M. Akram, Time series data library, Available from Internet: http://robjhyndman. com/TSDL.

[35] P. S. de Mattos Neto, G. D. Cavalcanti, F. Madeiro, T. A. Ferreira, An approach to improve the performance of PM forecasters, PloS One 10 (9) (2015) e0138507.

[36] R. Hyndman, Y. Khandakar, Automatic time series forecasting: The forecast package for r, Journal of Statistical Software, Articles 27 (3) (2008) 1–22.

[37] F. Takens, Detecting strange attractors in turbulence, in: Dynamical systems and turbulence, Warwick 1980, Springer, 1981, pp. 366–381.

[38] S. F. Crone, J. Guajardo, R. Weber, A study on the ability of support vector regression and neural networks to forecast basic time series patterns, in: IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, 2006, pp. 149–158.

[39] G. Song, Q. Dai, A novel double deep ELMs ensemble system for time series forecasting, Knowledge-Based Systems 134 (2017) 31–49.

[40] M. Qin, Z. Li, Z. Du, Red tide time series forecasting by combining ARIMA and deep belief network, Knowledge-Based Systems 125 (2017) 39 – 52.

[41] H. Jaeger, The echo state approach to analysing and training recurrent neural networks-with an erratum note, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148 (34) (2001) 13.

[42] D. Steinberg, P. Colla, CART: classification and regression trees, Vol. 9, CRC Press, 2009.