A Nonlinear Optimized PSO-SVR Hybrid System for Time Series Forecasting with ARIMA 3 rd Manoel Henrique da Nóbrega Marinho

4 authors, including:

Joao Fausto Lorenzato de Oliveira
Universidade de Pernambuco
47 PUBLICATIONS   **413** CITATIONS

SEE PROFILE

Manoel H. N. Marinho
University of Pernambuco
43 PUBLICATIONS   **159** CITATIONS

SEE PROFILE

Francisco Madeiro
Universidade de Pernambuco
178 PUBLICATIONS   **755** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Previsão de Séries Temporais Utilizando Sistemas Inteligentes Híbridos View project

Desenvolvimento de um Portal de Análise de Imagens View project

A Nonlinear Optimized PSO-SVR Hybrid System for Time Series Forecasting with ARIMA

1st Emily Pereira Alves
Escola Politécnica de Pernambuco (POLI)
Universidade de Pernambuco (UPE)
Recife, Brazil
epa@poli.br

2nd João Fausto Lorenzato de Oliveira
Escola Politécnica de Pernambuco (POLI)
Universidade de Pernambuco (UPE)
Recife, Brazil
fausto.lorenzato@upe.br

3rd Manoel Henrique da Nóbrega Marinho
Escola Politécnica de Pernambuco (POLI)
Universidade de Pernambuco (UPE)
Recife, Brazil
marinho75@poli.br

4th Francisco Madeiro
Escola Politécnica de Pernambuco (POLI)
Universidade de Pernambuco (UPE)
Recife, Brazil
madeiro@poli.br

Abstract—In the forecasting time series field, the combination of techniques to aid in predicting different patterns has been the subject of several studies. Hybrid models have been widely applied in this scenario, where the vast majority of series are composed of linear and nonlinear patterns. The Autoregressive Integrated Moving Average (ARIMA) presents satisfactory results in a linear pattern prediction but can not capture nonlinear ones. In dealing with nonlinear patterns, the Support Vector Regression (SVR) has shown promising results. In order to map both patterns, an optimized nonlinear combination model based on SVR and ARIMA is proposed. The main difference in comparison with other works is the use of an interactive Particle Swarm Optimization (PSO) to increase the prediction performance. To the experimental setup, six well-known datasets of the literature is used. The performance is assessed by the metrics Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). The results show the proposed system attains better outcomes when compared to the other tested techniques, for most of the used data.

Index Terms—Time series forecasting, hybrid models, ARIMA, SVR

I. Introduction

The prediction of electricity load has a fundamental role in the planning of electrical energy distribution systems [1], [2]. An inaccurate estimate of the demand curve generates waste of resources, either due to underestimation or overestimation [3]. Underestimation occurs when there are no sufficient resources to the load demand in some period, causing a power outage and increasing customer dissatisfaction. On the other hand, the overestimation leads to a waste of resources, which increases the cost of generating power [4].

The non-linear nature of the load curve's behavior and its dependence on external factors, such as time, day, weather conditions, calendar effects, and economic activities, for example, are obstacles in forecasting systems [5]. Therefore, the development of systems capable of satisfactorily modeling temporal phenomena has proved to be a relevant and challenging task and has been the subject of several studies in the past decades [6]–[12].

Traditionally, Box and Jenkins' statistical models have been widely used for load forecasting due to their flexibility and simplicity [11]–[13], such as Autoregressive Integrated Moving Average (ARIMA). However, regarding time series, these approaches can only map linear patterns, limiting their applicability. Modeling a real world time series becomes a challenging problem because it is usually made up of a combination of linear and non-linear patterns [14], [15].

In the literature, there is no single approach that can predict well both patterns. Due to this, hybrid proposals, capable of handling different models, have been highlighted in the literature. This architecture combines statistical models and computational intelligence techniques to improve the quality of prediction and is more robust to changes in time series patterns [16]. Commonly, statistical models are used for predicting the linear patterns of the series.

The prediction process generates prediction errors, also called residuals errors, calculated by the difference of the original data and the predicted one. The residual series contains the nonlinear components that were not modeled by the first model. To improve the final forecast, the residual series is modeled by a machine learning technique. Combining the prediction of the original series with the residual one can increase the performance of the prediction system [17].

Among the nonlinear models, approaches based on Support Vector Machine (SVM) have shown significant results in the forecasting scenario [9], [15], [18]–[20]. SVMs are successfully used in time series forecasting due to the structural risk minimization principle and the capability to solve a linear constrained quadratic programming problem, returning a globally optimal solution [18]. With the introduction of Vapnik's ε-insensitive loss function [21] and kernel functions, the SVM has also been extended to nonlinear regression problems.

To improve the accuracy of the forecasts it is necessary to

find the optimal set of parameters of the chosen model. It may be a challenge, and there are several studies to find the best way for this purpose. In the approaches which use Support Vector Regression (SVR), one traditional form proposed by Hsu et al. [22] is to combine all the possibilities of a pre-established range of parameters, performing a grid search to find the optimal or suboptimal combination of the parameter's values. However, depending on the number of parameters, this strategy can generate a high time cost. Some works use optimization algorithms, such as Particle Swarm Optimization (PSO), to select the parameters of machine learning techniques [19], [23]–[26].

Depending on the time series characteristics, linear combination methods can be effective and have been widely used in the literature [27]–[29]. However, these models have some limitations, since real-world time series are rarely pure linear. They often contain both linear and nonlinear patterns which linear models can not deal with. Khashei and Bijari [30] introduced an approach that uses a nonlinear combination between the predictions of linear and nonlinear patterns. The choice of approaches will depend on the characteristics of the time series used. There is no methodology defined as ideal for all problems. Searching for the best combination is an open problem. If the linear model is well specified, there must be no linear correlation in the residual series. Therefore, nonlinear models based on machine learning may be applied. A challenge with these techniques is the vulnerability to overfitting or underfitting.

The essence of ML techniques is to capture the dominant behavior and to fit the data according to that behavior. When the model is overtrained, secondary patterns that may not be necessary for the generalization of the model can be mapped, causing overfitting. On the other hand, when the model does not learn enough patterns from the training data, not capturing the dominant trend, the underfitting occurs. Besides that, in sequential forecasting systems, whose prediction of the first model serves as the basis for the second, and so on, the error of the forecasts ends up spreading through the stages of the system.

In this paper, a new hybrid technique is proposed for forecasting time series based on the ARIMA, SVR, and PSO algorithm. In consonance with other methods of the literature, ARIMA is used to predict the linear components of the series. Then, an optimzated SVR is used to forecast the residual series and combine the predictions. The main difference between the proposed method and techniques of the literature is how PSO algorithm is used. The choice of topology and inertia coefficient is part of the forecasting system, which can change according to the characteristics of each dataset. To combine the forecast from the first and the second model, the SVR is also used as a nonlinear model combination.

The present work is organized as follows: Section II presents the related works that are the basis for the proposal; Section III describes the proposed hybrid method; Section IV shows the experimental setup and the results obtained; the discussion of the proposed method and the concluding remarks are provided in Section V.

II. RELATED WORKS

Real-world time series can be purely linear or purely non-linear, but in the majority of cases, it presents a combination of linear and nonlinear patterns [31]. For that reason, single models can not perform as well as hybrid models, which can handle different combination of patterns. As Zhang has shown [16], modeling separately linear and nonlinear patterns using two predictors improves the performance of single models. Another way to improve the accuracy of the system is to deal separately with time series and residuals series [27]. In the works that use the SVR, some authors have proposed different approaches to find the optimal set of parameters to improve the forecast result [19], [23]–[26].

Shamsuddin and Sallehuddin proposed a hybrid model [25] that combines SVR and ARIMA to crime rate forecasting. The PSO was used by them in two executions: one to estimate the parameters of the SVR and the other to estimate ARIMA parameters. Oliveira and Ludemir made a different approach [19], using the PSO algorithm to improve the quality of predictions. They proposed a distributed ARIMA-SVR hybrid system simultaneously optimized by a discrete PSO and a continuous-valued PSO. The discrete PSO selects the parameters of the ARIMA model, while the continuous version selects the SVR ones. In [7] the PSO is used to find the best hyperparameters for SVR models to perform residual forecasting in a combinational hybrid proposal.

A. Autoregressive Integrated Moving Average (ARIMA)

The ARIMA is a traditional forecasting method introduced by Box and Jenkins [11]. According to Zhang [16], it is one of the most important and widely used model to time series forecasting. The ARIMA model (p, d, q) consists in an Auto Regressive (AR) model of order p, a Moving Average (MA) model of order q and a differentiation step d to make the series stationary. In an ARIMA model, the values of the prediction are supposed to be a linear combination of past values and past residual errors, that is,

$$
\begin{aligned}
y_t = {} & \theta_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} \\
& + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q},
\end{aligned}
\tag{1}
$$

in which y_t and ε_t are the actual and error values at time period t, respectively. The model coefficients are $\theta_i(i = 1, 2, ..., q)$ and $\phi_j(j = 1, 2, ..., p)$, where p and q are the degrees of the autoregressive and moving average polynomial functions.

The major limitation of ARIMA is the inability to capture nonlinear patterns, which ends up limiting its applicability in real-world problems when used alone.

B. Support Vector Regression

In most linear regression models, the main objective is to minimize the squared errors. However, the Support Vector Regression (SVR), proposed by Vapnik [32], is based on the Structured Risk Minimization (SRM) principle to overcome overfitting, estimating a function that minimizes the upper

limit of the generalization error. In the SVR, a deviation of the real value can be acceptable as long as they are less than a previously established value.

Consider a training set $(\mathbf{x}_i, y_i)$, where $\mathbf{x} \in \mathbb{R}^d$ is the input vector, $y_i \in \mathbb{R}$ is the ith predicted value, and d is the embedding dimension of the time series. The goal of SVR is to find the best function in the form

$$\{f | f(x) = \mathbf{w}^T\mathbf{x} + \mathbf{b} \text{ with } w \in \mathbb{R}, b \in \mathbb{R}\}, \quad (2)$$

where w is the weight vector estimated by minimizing the regularized risk function and $\mathbf{b}$ is a previously established threshold. One can describe this as an optimization problem, which needs to

$$\text{minimize} \quad \frac{1}{2}||\mathbf{x}||^2 + C \sum_{i=1}^{L} L(y_i, f(x_i)), \quad (3)$$

where C is a regularization factor, $||.||$ is a $2-$norm, and $L(\cdot, \cdot)$ is a loss function. To induce sparsity in SVR, one creates an ε-tube allowing some predictions with deviation within these limits. For this, the ε-insensitive loss function is subject to

$$L(y, f(\mathbf{x})) = \begin{cases} 0, & |f(\mathbf{x}) - y| < \varepsilon \\ |f(\mathbf{x}) - y| - \varepsilon, & \text{otherwise.} \end{cases} \quad (4)$$

Another particularity of SVR is to introduce slack variables, denoted by ξ and ξ_i^*, to measure the errors that occurs by values outside the limits of ε-tube. So, the SVR function is rewritten to

$$\text{minimize} \quad \frac{1}{2}||\mathbf{x}||^2 + C \sum_{i=1}^{L} (\xi - \xi^*), \quad (5)$$

subject to

$$\begin{cases} \mathbf{w}\mathbf{x}_i + \mathbf{b} - y_i \leq \varepsilon + \xi_i \\ y_i - \mathbf{w}\mathbf{x}_i - \mathbf{b} \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0. \end{cases} \quad (6)$$

In SVR, the employment of kernels allows performing nonlinear mappings into a higher dimensionality space. The regression procedure is expressed as

$$f(\mathbf{x}) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}) + \mathbf{b}, \quad (7)$$

in which α and α^* are Lagrange multipliers, $k(\mathbf{x}_i, \mathbf{x})$ is a kernel function, and $\mathbf{b}$ is a previously established threshold. In the present work, the kernel used is the radial basis function (RBF), also known as Gaussian kernel, which takes the form

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{2\gamma^2})\right), \quad (8)$$

where γ is a parameter of the RBF kernel.

C. Particle Swarm Optimization

Particle Swarm Optimization (PSO) [33] is a population-based stochastic algorithm inspired by the collective behavior of a bird flocking. In the PSO, each particle represents a potential solution to the problem, and the swarm is the population of solutions. As the metaphor suggests, the search for the best position is carried out based on the interaction between the birds. Each particle determines its movement combining the historical information about its own positions and of other particles. After these interactions, the swarm tends to move close to the best global location already found.

At each iteration t of the algorithm, the best global position, called $gbest$, is calculated, and for each particle i of the swarm P, the best personal position is calculated, called $pbest$. After that, in the next iteration, the particles move according to the velocity function, as

$$\begin{aligned} v_i(t+1) = w \times v_i(t) + c_1 \times r_1 \times (pbest_i - x_i(t)) + \\ c_2 \times r_2 \times (gbest_i - x_i(t)), \end{aligned} \quad (9)$$

where w is the inertia weight, v_i is the actual velocity, c_1 and c_2 are acceleration coefficients, r_1 and r_2 are random values in the range [0,1] and x_i is the ith element of the current solution x.

III. PROPOSED METHOD

The main objective of this paper is to propose a new alternative for time series forecasting using a nonlinear approach with different PSO topologies, called SVR(ARIMA, PSOSVR). Also, the proposed approach is compared to other techniques in the literature, such as traditional linear and nonlinear forecasting techniques using ARIMA and SVR. In general, the hybrid forecasting systems are divided into three sequential steps. First, the linear patterns are modeled by a linear technique. Thus, it is expected that the correlated components, such as trend and seasonality, are not observed in the residues, avoiding multicollinearity. After that, the difference between the original data and the first forecast is calculated generating the residual base, which is modeled by a nonlinear technique. Lastly, the combination of the forecasts is carried out. This combination can be linear, adding both forecasts, or non-linear, performed by some machine learning algorithm, for example.

Besides using traditional approaches, the proposed method uses the PSO to find the optimal or sub-optimal set of parameters for the technique used in the second model; specifically in the present paper the SVR is used. The performance of the PSO is strongly related to its topology, which should be chosen according to the characteristics of the problem. An improper choice of the topology may lead the PSO to premature convergence or may lead to low search efficacy [34]. Due to this, the choice of topology and coefficient of inertia to be used was incorporated into the model, which may vary depending on the characteristics of the base. The architecture of the proposal is shown in Figure 1.

In the proposed method, the ARIMA is used in the first stage, performing the linear patterns prediction. The difference
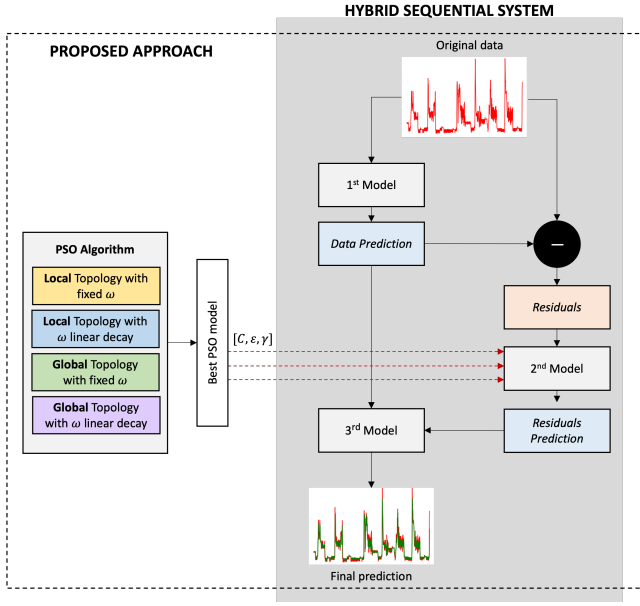
Fig. 1. Proposed hybrid architecture: SVR(ARIMA, PSOSVR).

Fig. 2. Time series dataset used.

between the original and predicted data results in a new series, known as residuals. The second stage of the proposed system is responsible to model the residuals. For this, one uses the SVR, whose parameters are optimized by the PSO algorithm, which returns the lowest mean squared error of the prediction. Since there is not a universal optimal model for all datasets, the choice of the topology used by the PSO was incorporated to the system, varying between local and global topology, and with or without linear decay. In order to compare the results, other previously published methods are implemented, as shown in Table I. According to [17], the combination of the first prediction with the residual serie can improve the precision of the whole system. The nonlinear combination, introduced by Khashei and Bijari [30], that was used as basis for the Nonlinear Combination Method [20], has provided good results for time series prediction in terms of accuracy [20]. Due to this, the SVR is used as the combination model for the forecasts.

IV. EXPERIMENTAL SETUP AND RESULTS

In this section the experimental setup is detailed along with the presentation of results.

A. Experimental Setup

In this paper, six datasets from the Time Series Data Library [36] are used to evaluate the performance of the methods. The description of each series is shown in Table II. These time series, presented in Figure 2, are commonly used in the literature [9], [19], [31], [37], [38] and are from different application areas with distinct characteristics.

Approximately 50% of the data was used as a training set and the remaining 50% for the test set. Specifically for machine learning techniques, the training set was subdivided
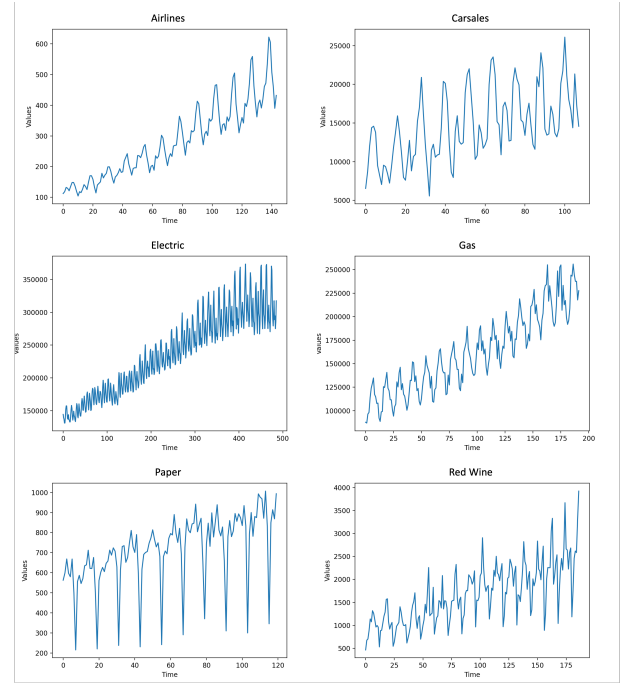
for the training and validation steps. All data sets were normalized, that is, represented in the range [0,1], according to the equation

$$x = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{10}$$

in which x_{min} is the lower value of the series and x_{max} is the highest one.

The evaluation of methods was performed using three well-known performance measures in the literature [9], [16], [39]: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE), given by the equations 11, 12, 13, respectively,

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^{N} (y_t - y'_t)^2, \tag{11}$$

$$\text{MAPE} = \frac{100}{N} \sum_{t=1}^{N} \left| \frac{y_t - y'_t}{y_t} \right|, \tag{12}$$

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - y'_t|, \tag{13}$$

in which N is the dataset size, y_t is the real value at time t and y'_t is the forecast value at the same time. For these metrics, the lower is the value, the better is the accuracy. The MSE is commonly used as an error measure in time series forecasting domain. However, this metric has some limitations such as being sensitive to outlier data and being scale-dependent [40]. On the other hand, the MAPE metric is not scale-dependent and outputs a percentage error. In terms of sensitivity to

Approach	Models	Reference	Short descriptions
Single model	ARIMA	Zhang [16]	Single ARIMA prediction.
Linear combination	ARIMA + SVR	de Oliveira and Ludermir [35]	Sum of ARIMA and SVR predictions.
	ARIMA + PSOSVR	de Olanda and de Oliveira [7]	PSO to set the SVR parameters in a linear combination.
Nonlinear combination	SVR(ARIMA, SVR)	de Mattos Neto *et. al* [20]	SVR to perform the combination of predictions.
	SVR(ARIMA, PSOSVR)	Proposed Method	PSO to set the SVR parameters in a nonlinear combination.

TABLE II
DATASET DESCRIPTION.

Dataset	Size	Description
Airline	144	Monthly International airline passengers
Carsales	108	Monthly car sales in Quebec
Electric	486	Australian Electricity: Electricity end use
Gas	192	Monthly gas demand in Ontario
Paper	120	Industry sales for printing and writing paper
Wine	187	Monthly Australian wine sales

TABLE IV
BEST EMBEDDING DIMENSION FOR EACH DATASET.

Dataset	Lag
Airline	12
Carsales	12
Electric	12
Gas	12
Paper	20
Redwine	12

outliers, an alternative is the MAE metric, which is best suited for being unambiguous and more robust to outliers [41].

The values of the parameters for the techniques used are presented in Table III. For ARIMA, the choice of parameters (p, q, d) was made using the **autoarima** function, which chooses them in order to minimize the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [42]. For the models which do not use the PSO algorithm, the parameters were defined through an exhaustive search, which combines all the values of a pre-set discrete range and finds the optimal arrangement. Regarding the PSO technique, the parameters used are well-established in the literature.

In time series forecasting domain, an important parameter to be defined is the number of lags that will be used, that is, how many past instants will be used to predict a future one. The number of lags is an overriding factor in making good predictions. The number of lags used for each database was configured according to [7], [9], [19], [38], and are shown in the Table IV.

TABLE III
LIST OF PARAMETERS.

Model	Parameters	Values
ARIMA	p, d, q	*Hyndman [36] methodology*
SVR	Kernel	*Radial Basis Function* (RBF)
	ϵ	$[10^{-8}, 10^{-7}, \cdots, 10^{-2}]$
	C	$[0.01, 0.1, 1, 10, 100, 1000, 10000]$
	γ	$[0.01, 0.1, 1, 10, 100, 1000]$
PSO	c_1 and c_2	2.05
	w	0.723 and [1, 0.723]
	Particles	30
	Iteration	5000
	Topology	Local and global

The analysis of the results is divided as follows: the first, second and third subsection are based on the MSE, MAPE, and MAE error measures, respectively, for all the techniques. In the last subsection comparisons are presented between the ARIMA and the proposal method. For the approaches that use the PSO, which is a non-deterministic technique, 20 simulations for each database were performed. In these cases, the results are presented as $a \pm d$, where a is the average and d is the standard deviation.

B. Analysis based on MSE

The results concerning the MSE are presented in Table V, where the best results are highlighted in bold. According to the results, the proposed method achieved the lowest MSE values on five datasets: Airline, Carsales, Electric, Paper, and Redwine. Regarding the standard deviation, it can be noted that the values, in most cases, are close.

C. Analysis based on MAPE

The results are also analyzed under the MAPE criterion, which is very important in time series forecasting domain. The standard deviation is presented in percentage points (p.p.). According to this metric, the proposed method outperformed all techniques in four cases: Airline, Electric, Paper and, Redwine, as one can see in Table VI. Concerning the standard deviation, the results obtained for the Carsales and Redwine databases present a greater difference between the techniques. For the Carsales database, observing only the average of the results, the method ARIMA + PSOSVR is superior. However, when evaluating the standard deviation, one notice a smaller deviation for the results obtained by the proposed method. For the Redwine database, the proposed method obtains a better performance in terms of average and deviation.

D. Analysis based on MAE

Another analysis is made according to the MAE metric, and the results are shown in Table VII. The proposed method presented the best results considering Airline and Electric datasets. Analyzing the standard deviation, for the Carsales and Redwine dataset, we can notice a significant difference between the ARIMA + PSOSVR technique and the proposed

TABLE V

MEAN SQUARE ERRORS (MSE) FOR ALL DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Dataset	ARIMA	ARIMA + SVR	SVR(ARIMA, SVR)	ARIMA + PSOSVR	SVR(ARIMA, PSOSVR)
Airline	0.00943	0.00138	0.00138	0.00132 ± 0.00051	**0.00127** ± 0.00052
Carsales	0.03824	0.01398	0.01427	0.01443 ± 0.00531	**0.01373** ± 0.00260
Electric	0.00549	0.00198	0.00194	0.00163 ± 0.00061	**0.00142** ± 0.00051
Gas	0.01224	0.00323	0.00322	**0.00341** ± 0.00108	0.00369 ± 0.00133
Paper	0.08310	0.00873	0.01172	0.00821 ± 0.00262	**0.00804** ± 0.00261
Redwine	0.02857	0.00946	0.00973	0.00838 ± 0.00281	**0.00814** ± 0.00011

TABLE VI

MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) FOR ALL DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Dataset	ARIMA	ARIMA + SVR	SVR(ARIMA, SVR)	ARIMA + PSOSVR	SVR(ARIMA, PSOSVR)
Airline	13.13%	4.48%	4.49%	4.56% ± 1.53 p.p.	**4.41%** ± 1.53 p.p.
Carsales	24.86%	17.95%	18.20%	**17.89%** ± 5.90 p.p.	18.55% ± 1.12 p.p.
Electric	8.38%	4.75%	4.72%	4.28% ± 1.41 p.p.	**4.00%** ± 1.31 p.p.
Gas	11.75%	6.08%	6.07%	**6.02%** ± 1.91 p.p.	6.11% ± 1.95 p.p.
Paper	63.73%	16.99%	17.06%	15.88% ± 5.06 p.p.	**15.55%** ± 4.97 p.p.
Redwine	36.20%	17.89%	17.83%	15.87% ± 5.16 p.p.	**15.42%** ± 0.26 p.p.

method. According to these values, it can be observed a smaller dispersion in the proposed method, resulting in a better performance when compared in terms of deviation.

E. Final analysis

Comparing with the literature methods, regarding the average results, for the Airline and Electric datasets, the proposed method presented better results in terms of the three metrics used. Considering the standard deviation results, the proposed method presented superior results for the Carsales, Electric and Redwine.

When comparing to ARIMA, the proposed method obtained better results for all datasets. Table VIII presents superiority in terms of MSE, MAPE and MAE for each dataset. Figures 3, 4 and 5 present the curves referring to the original time series, the ARIMA prediction and the proposed approach prediction, for the Carsales, Electric and Redwine datasets, respectively. One may note that the proposed method's curve is closer to the original one than the curve of the ARIMA method.

V. CONCLUSION

In this paper, an optimized hybrid system for forecasting time series was proposed. It is assumed that, in general, time series are composed of linear and nonlinear patterns. The first step of the system consists of capturing the linear patterns using a linear model. In the second one, a machine learning technique is used to predict the generated residual errors, and finally, a nonlinear combination of the obtained predictions is accomplished. The parameter optimization was performed using PSO algorithm to improve the predictions. The choice of the PSO topology was incorporated into the system, increasing the flexibility of the model to different types of data. The proposed model obtained promising results when compared to other techniques in the literature.

In order to carry out a comparative analysis, four methods were considered: Zhang [16], de Oliveira [19], de Holanda

[7] and de Mattos Neto *et al.* [20], which uses ARIMA, SVR and, PSO, in single, linear and nonlinear approaches. Based on the experimental results obtained, it is possible to note that no technique obtained systematic superiority for all datasets. However, the proposed method outperformed the others in five out of six datasets used, concerning the MSE, and four out of six concerning MAPE and MAE metrics. In relation to [7], which uses the PSO algorithm to optimize the parameters, our results suggest that allowing the system to choose the best combination of the PSO parameters improves the accuracy of the system. It may occur because the model fits the different characteristics of the time series.

As future works, we aim to develop a multiobjective-optimized hybrid system to reduce the error propagated in sequential models, using multiobjective swarm algorithms, for example. Also we intend to investigate the performance of other machine learning techniques to improve the combination of the linear and nonlinear forecasting models. Other swarm algorithms could be analysed to enhance the system's parameter choice.

REFERENCES

[1] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Advanced Engineering Informatics*, vol. 35, pp. 1–16, 2018.

[2] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 431–438, 2015.

[3] Y. Lin, H. Luo, D. Wang, H. Guo, and K. Zhu, "An ensemble model based on machine learning methods and data preprocessing for short-term electric load forecasting," *Energies*, vol. 10, no. 8, p. 1186, 2017.

[4] A. B. Nassif, "Short term power demand prediction using stochastic gradient boosting," in *5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*. IEEE, 2016, pp. 1–4.

[5] N. Neeraj, J. Mathew, M. Agarwal, and R. K. Behera, "Long short-term memory-singular spectrum analysis-based model for electric load forecasting," *Electrical Engineering*, pp. 1–16, 2020.

| Dataset	ARIMA	ARIMA + SVR	SVR(ARIMA, SVR)	ARIMA + PSOSVR	SVR(ARIMA, PSOSVR)
Airline	0.0817	0.0281	0.0281	0.0269 ± 0.0089	**0.0260** ± 0.0091
Carsales	0.1498	0.0933	0.0936	**0.0900** ± 0.0295	0.0919 ± 0.0059
Electric	0.0579	0.0346	0.0343	0.0308 ± 0.0102	**0.0286** ± 0.0094
Gas	0.0876	0.0466	0.0466	**0.0448** ± 0.0142	0.0453 ± 0.0144
Paper	0.1991	**0.0686**	0.0805	0.0699 ± 0.0226	0.0707 ± 0.0225
Redwine	0.1261	0.0754	0.0752	**0.0695** ± 0.0223	0.0726 ± 0.0005
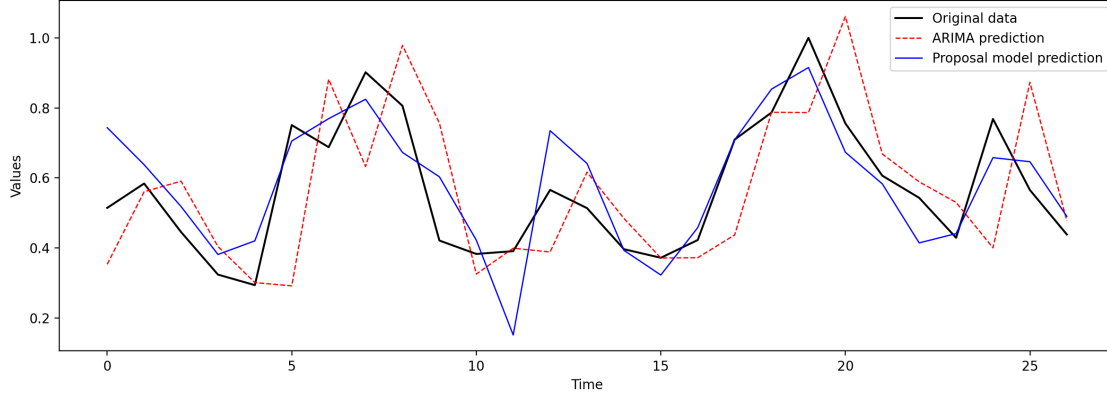
Fig. 3. Curves of the original data and the predictions from ARIMA and the proposal method, for the Carsales dataset.
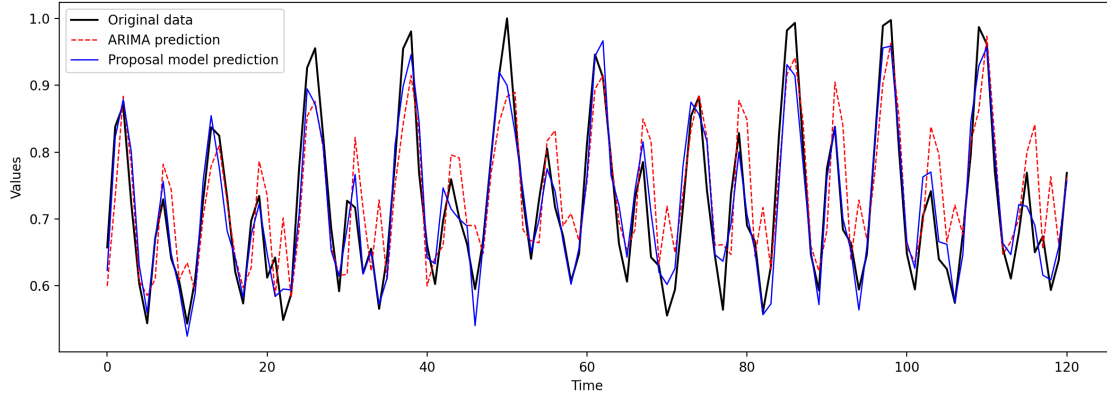

Fig. 4. Curves of the original data and the predictions from ARIMA and the proposal method, for the Eletric dataset.

TABLE VIII
PERCENTAGE DIFFERENCE IN TERMS OF MSE, MAE AND MAPE
BETWEEN THE PROPOSAL MODEL AND ARIMA, FOR ALL DATASETS.

Dataset	MSE	MAPE	MAE
Airline	86.50%	68.23%	66.41%
Carsales	64.11%	38.63%	25.38%
Electric	74.18%	50.58%	52.23%
Gas	69.87%	48.34%	47.97%
Paper	90.32%	64.48%	75.60%
Redwine	71.52%	42.42%	57.41%

[6] J. F. L. de Oliveira, E. G. Silva, and P. S. G. de Mattos Neto, "A hybrid system based on dynamic selection for time series forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[7] R. B. B. de Holanda and J. F. de Oliveira, "Swarm based ensembles for time series residual forecasting," in *IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2020, pp. 595–602.

[8] Z. Tan, J. Zhang, Y. He, Y. Zhang, G. Xiong, and Y. Liu, "Short-term load forecasting based on integration of SVR and stacking," *IEEE Access*, vol. 8, pp. 227 719 – 227 728, 2020.

[9] S. O. Domingos, J. F. 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*, vol. 175, pp. 72–86, 2019.

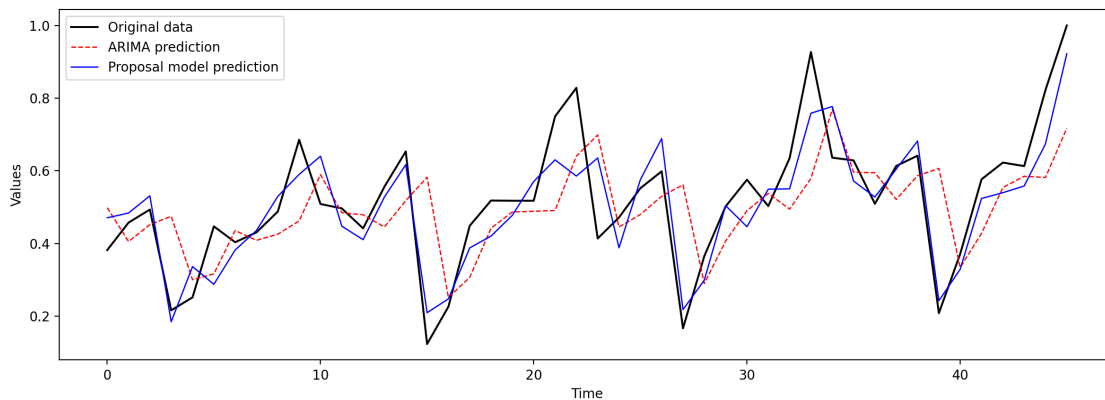[10] L. Xiao, W. Shao, M. Yu, J. Ma, and C. Jin, "Research and application

Fig. 5. Curves of the original data and the predictions from ARIMA and the proposal method, for the Redwine dataset.

of a combined model based on multi-objective optimization for electrical load forecasting," *Energy*, vol. 119, pp. 1057–1074, 2017.

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

[12] V. Bianco, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy*, vol. 34, no. 9, pp. 1413–1421, 2009.

[13] R. Abdel-Aal and A. Z. Al-Garni, "Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis," *Energy*, vol. 22, no. 11, pp. 1059–1069, 1997.

[14] P. R. A. Firmino, P. S. G. de Mattos Neto, and T. A. E. Ferreira, "Error modeling approach to improve time series forecasters," *Neurocomputing*, vol. 153, pp. 242–254, 2015.

[15] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[16] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[17] P. S. G. de Mattos Neto, G. D. C. Cavalcanti, F. Madeiro, and T. A. E. Ferreira, "An approach to improve the performance of PM forecasters," *PloS One*, vol. 10, no. 9, p. e0138507, 2015.

[18] W.-C. Hong, "Electric load forecasting by support vector model," *Applied Mathematical Modelling*, vol. 33, no. 5, pp. 2444–2454, 2009.

[19] J. F. L. de Oliveira and T. B. Ludermir, "A distributed PSO–ARIMA–SVR hybrid system for time series forecasting," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2014, pp. 3867–3872.

[20] P. S. G. de Mattos Neto, G. D. C. Cavalcanti, and F. Madeiro, "Nonlinear combination method of forecasters applied to PM time series," *Pattern Recognition Letters*, vol. 95, pp. 65–72, 2017.

[21] V. Vapnik, S. E. Golowich, A. Smola *et al.*, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in Neural Information Processing Systems*, pp. 281–287, 1997.

[22] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," http://aiweb.techfak.uni-bielefeld.de/content/bworld-robot-control-software/, 2003, [Online; accessed 12-June-2021].

[23] C.-L. Huang and J.-F. Dun, "A distributed PSO–SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing*, vol. 8, no. 4, pp. 1381–1391, 2008.

[24] J. Che, J. Wang, and G. Wang, "An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting," *Energy*, vol. 37, no. 1, pp. 657–664, 2012.

[25] R. Alwee, S. M. Shamsuddin, and R. Sallehuddin, "Hybrid support vector regression and autoregressive integrated moving average models improved by particle swarm optimization for property crime rates forecasting with economic indicators," *The Scientific World Journal*, vol. 2013, 2013.

[26] M.-R. Kazemzadeh, A. Amjadian, and T. Amraee, "A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting," *Energy*, vol. 204, p. 117948, 2020.

[27] P.-F. Pai and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[28] E. Cadenas and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model," *Renewable Energy*, vol. 35, no. 12, pp. 2732–2738, 2010.

[29] P. R. A. Firmino, P. S. G. de Mattos Neto, and T. A. E. Ferreira, "Correcting and combining time series forecasters," *Neural Networks*, vol. 50, pp. 1–11, 2014.

[30] M. Khashei and M. Bijari, "Which methodology is better for combining linear and nonlinear models for time series forecasting?" *Journal of Industrial and System Engineering (JISE)*, 2012.

[31] S. Panigrahi and H. S. Behera, "A hybrid ETS–ANN model for time series forecasting," *Engineering Applications of Artificial Intelligence*, vol. 66, pp. 49–59, 2017.

[32] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[33] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *International Conference on Neural Networks (ICNN)*, vol. 4. IEEE, 1995, pp. 1942–1948.

[34] S. Cheng, Y. Shi, and Q. Qin, "Population diversity based study on search information propagation in particle swarm optimization," in *IEEE Congress on Evolutionary Computation*. IEEE, 2012, pp. 1–8.

[35] P. S. G. de Mattos Neto, F. Madeiro, T. A. E. Ferreira, and G. D. C. Cavalcanti, "Hybrid intelligent system for air quality forecasting using phase adjustment," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 185–191, 2014.

[36] R. Hyndman, "Time series data library," 2015, [Online; accessed 12-June-2021].

[37] P. S. G. de Mattos Neto, J. F. L. de Oliveira, H. V. Júnior, Domingos Savio de Oliveira Santos Siqueira, M. H. D. N. Marinho, and F. Madeiro, "A hybrid nonlinear combination system for monthly wind speed forecasting," *IEEE Access*, vol. 8, pp. 191 365–191 377, 2020.

[38] J. F. de Oliveira and T. B. Ludermir, "A hybrid evolutionary decomposition system for time series forecasting," *Neurocomputing*, vol. 180, pp. 27–34, 2016.

[39] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Applied Soft Computing*, vol. 11, no. 2, pp. 2664–2675, 2011.

[40] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, vol. 8, no. 1, pp. 69–80, 1992.

[41] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.

[42] R. J. Hyndman, Y. Khandakar *et al.*, "Automatic time series forecasting: the forecast package for r," *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.