Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 20, Iss. 1, pp. 17-30, 2022

A NONLINEAR OPTIMIZED HYBRID SYSTEM FOR ENERGY CONSUMPTION FORECASTING FROM SMART METERS

Emilly Pereira Alves, João Fausto Lorenzato de Oliveira, Francisco Madeiro, Manoel Henrique da Nóbrega Marinho
Universidade de Pernambuco (UPE)
epa@poli.br, fausto.lorenzato@upe.br, madeiro@poli.br, marinho75@poli.br

Abstract – Smart grids are an alternative to minimize environmental impacts, such as CO_2 emissions, and improve the efficiency of electricity consumption in buildings. Power grids enable adequate management and monitoring of consumption because of the periodic storage of measurements and easy access to them. In this scenario, an accurate prediction is a challenging task. Forecasting of consumption series is a defiant problem because data present linear and nonlinear patterns, and a dependence on external variables may be observed. Hybrid models are an alternative to mapping both patterns, which have been widely used to forecast load time series. Autoregressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR) models are used for this purpose, to map the linear and nonlinear patterns of the series, respectively. In this paper, a nonlinear optimized hybrid system based on ARIMA, SVR, and Particle Swarm Optimization (PSO) is proposed. The system can be divided into three steps. First, the linear patterns are predicted by the statistical model ARIMA. Then, the residual series is modeled using an optimized SVR, in which the parameters are selected from the PSO. One particularity from the proposal is to incorporate the choice of the topology and the inertia coefficient into the system. Lastly, the predictions are combined using the SVR. The simulations were conducted using a real database from smart meters of a building in Taiwan. To evaluate the performance of the proposed method, four related approaches were implemented and compared: a single ARIMA, two linear combination systems, and one non-linear combination system. The results show a superiority of the proposed method in terms of the metrics Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Keywords – Load forecasting, time series, smart grid, hybrid models, SVR, ARIMA.

1. INTRODUCTION

Global warming and the energetic crisis have motivated the use and development of alternative, sustainable, and clean energy sources [1]. The construction sector has shown a large energy consumption and CO_2 emission [2, 3]. Hence, buildings are critically important to the response to climate change [2]. Countries have accelerated the implementation of energy rules and regulations for various types of buildings [4].

Several studies have shown that energy efficiency can significantly reduce total energy consumption [4–6]. For these reasons, smart grids have a fundamental role in the efficient integration of energy generation and consumption. This alternative allows a closer monitoring and, consequently, a better management of the individual consumption, due to the use of the smart meters. These devices are able to record, process, and store periodic measurements of household consumption, allowing local or remote access to such information.

The advantages of a smart grid over the traditional grid does not rely only on the minimization of supply interruptions, but also on making energy consumption more efficient and adaptable to customer needs, and therefore, less costly [7]. Another convenience is that the knowledge of measurements can help reduce the peak demand. The storage of measurements allows studies on the load curve, enabling the creation of predictive models and continuous monitoring of energy consumption. The accurate forecasting of energy consumption is a critical step to improving the energy efficiency of buildings [8].

The prediction of energy consumption in buildings using smart grids has been discussed in several studies [9–12]. One of the objectives of the smart grid is to optimize the energy performance, minimize environmental impacts, and control the influence of energy costs [13]. The availability of measurements allows consumption patterns to be extracted and modeled by statistical approaches or computational intelligence techniques, which can assess how external factors impact the building's consumption, then carry out a forecast more assertively.

Electric load forecasting also has a fundamental role in distribution planning at energy companies [14, 15]. Due to the de-verticalization of the electricity sector, there is a need to minimize costs with the acquisition and distribution of electricity, increasing investments in the sector. An inaccurate estimate of the demand curve generates losses to the company and the consumer, either due to underestimation or overestimation. The first occurs when there are no sufficient resources to the load demand in some period, generating interruptions in the power supply. On the other hand, when overestimated, it can cause a waste of resources, which increases the cost of generating power [16, 17].

Electricity forecast models are developed according to the particular characteristics of each energy concessionaire, varying according to the market conditions prevailing in each region [15]. The nonlinear behavior of the load curve, as well as its dependence on external factors such as ambient weather conditions and calendar effects, can be defiance to forecasting systems [18].

Hence, the development of approaches 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 [19–25].

Time series forecasting systems can be categorized into three groups: (i) linear models, (ii) non-linear models and (iii) hybrid models. Regarding linear approaches, statistical linear models such as the Autorregressive Integrated Moving Average (ARIMA), which is a well-established Box and Jenkins methodology, have been widely used for load forecasting due to their flexibility and simplicity [24–26]. However, these approaches can only map linear patterns of the series, limiting their applicability. The vast majority of series are composed of nonlinear and linear patterns, which makes their prediction challenging [27, 28].

Computational intelligence techniques have shown to be effective to map the nonlinear patterns of the series [28]. Among the nonlinear models, approaches based on Support Vector Machine (SVM) have shown significant results in the forecasting scenario [22, 28–31]. Some advantages of SVMs are the capability to solve a linear constrained quadratic programming problem and its wide applicability, which is related to the principle of structural risk minimization. With the introduction of Vapnik's ε-insensitive loss function [32] and kernel functions, the SVM has also been extended to nonlinear regression problems, in the form of Support Vector Regression (SVR).

There is no optimal model in the literature which can handle all series [33]. However, hybrid models have stood out due to their ability to predict both linear and nonlinear patterns of time series. In this architecture, statistical models and computational intelligence techniques are combined to improve the quality of prediction [34]. In general, hybrid forecasting systems consist of three steps: (i) original series prediction, (ii) residual series prediction, and (iii) the combination of predictions. The first step consists in modeling the linear patterns using a linear technique. Then, the residual series is obtained, given by the difference between the original series and the forecast of the linear model. The residual series is modeled by a nonlinear technique, resulting in the nonlinear forecast. Finally, the combination of predictions, from the original and the residuals series, is performed, resulting in the final prediction of the system.

Regarding the way of combining forecasts, two possibilities are commonly used in the literature. Zhang assumes that a time series can be decomposed into its linear and nonlinear patterns through a linear function (summation) [34]. In contrast to [34], Kashei and Bijari [35] assume that the linear and nonlinear components of the predictions have a nonlinear relation, then the combination is made by a function that combines the linear and nonlinear predictions.

There is no ideal methodology for all problems. Searching for the best way to make the combination of the predictions is an open problem. Nonlinear models based on Machine Learning (ML) may be applied for this purpose. However, they can show vulnerabilities as the sensitivity 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.

The choice of an optimal, or sub-optimal, set of parameters for the techniques improves the accuracy of the forecasts. Moreover, the residual series may present irregular patterns due to hereroscedasticity and noise, thus modelling residual series is challenging in the context of selecting the most appropriate model [19]. For this reason, a new hybrid technique is proposed for time series forecasting 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 optimized 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 the PSO algorithm is used. The choice of topology and inertia coefficient is incorporated into the system, which can change according to the characteristics of each database. All combinations, among a set of established parameters, are tested, and the best model, evaluated in terms of MSE, is returned. To combine the forecast from the first and the second model, the SVR is also used as a nonlinear model combination.

The remaining of this paper is organized as follows: Section 2 presents the related works that are the basis for the proposal; Section 3 describes the proposed hybrid method; Section 4 shows the experimental setup; Section 5 presents the results obtained; the discussion of the proposed method and the concluding remarks are provided in Section 6.

2. RELATED WORKS

Time series forecasting has proven to be a challenging field of research for some reasons: (i) the serial dependency between past and present values of the database samples, and (ii) the dependence on exogenous variables such as environmental conditions and human activities [36]. Real-world time series can be purely linear or purely nonlinear, but in the majority of cases, it presents a combination of linear and nonlinear patterns [37]. For that reason, in some cases, single models may not perform as well as hybrid models, which can handle different combination of patterns.

In 2003, Zhang [34] proposes a hybrid approach based on ARIMA and Artificial Neural Networks (ANN). These techniques are used jointly, aiming to capture different forms of relationships in time series data. Another way to improve the accuracy of the system is to deal separately with time series and residuals series [38]. Other hybrid models have been proposed assuming that the time series can be decomposed into low volatility and high volatility. Such models used the ARIMA [39–41] to predict linear patterns, combined with neural networks and SVM for predicting residuals.

In 2016, Oliveira and Ludermir [41] proposed an evolutionary hybrid ARIMA-SVR model based on the decomposition of the time series according to the volatility. For this, the authors used a Simple Exponential Smoothing (SES) to decompose the series into low and high volatility components. Once this is done, the ARIMA model is based on the components of low volatility, while those of high volatility are decomposed by the AR-SVR model. The PSO algorithm was used to choose the SVR parameters and the lags, which define how many past instants are needed to predict a future one.

In 2017, Panigrahi and Behera [37] proposed a methodology similar to [34], replacing ARIMA by an Exponential Smoothing (ETS). In the same year, Mattos Neto *et al.* [31] proposed a nonlinear combination method, referred to as Nonlinear Combination (NoLiC), to combine the predictions using a Multi Layer Perceptron (MLP). To predict the linear patterns of the series, they used a Hybrid Intelligent System (HIS) [42], which is based on the use of a genetic algorithm to find the best MLP parameter setting.

In 2018, Rahimi and Khashei [43] proposed a hybrid model incorporating ARIMA and MLP in parallel, in order to map linear and nonlinear patterns simultaneously. In addition, a least squares process was used to find the optimal weights of each component of the prediction system, improving the efficiency of the hybrid model. Siami-Namini *et al.* [44] showed that algorithms based on deep learning, specifically the Long-Short Term Memory (LSTM) [45], which is a variation of ANN based on feed forwarding mechanisms, have a better performance in forecasting time series when compared to the ARIMA model.

In 2019, Domingos *et al.* [22] proposed a hybrid system based on ARIMA, SVR and MLP, whose differential from the techniques in the literature is the search for the best function to combine the predictions of the original series and the residual series. In this system, a combination was performed based on time series data, seeking for the most suitable function for the model, as well as for the number of predictions that improve the performance of the combination.

In 2020, Hajirahimi and Khashei [46] demonstrated that using a non-linear intelligent model as the first component in the sequential modeling system leads to more accurate results when compared to models whose first step is to use a linear technique. Thus, they proposed a Nonlinear-Linear approach with the SVM and MLP techniques. In the same year, Mattos Neto *et al.* [47] proposed a hybrid model based on a search technique for the most suitable combination between the forecasts. For the linear components, the ARIMA and ARIMAX models were used, combined with Radial Basis Function (RBF), SVR and MLP, to predict the non-linear components. Kazemzadeh *et al.* [48] proposed a hybrid approach using three residual series, resulting from the prediction of three different techniques: (i) RNA, (ii) ARIMA and (iii) SVR-PSO. The proposed method prioritizes each forecasting method based on the resulting error over data, establishing a weight for each technique.

In 2021, de Oliveira *et al.* [19] proposed a hybrid system based on a dynamic selection algorithm capable of: (i) deciding on the most suitable machine learning model for the series and (ii) evaluating which models offer the greatest accuracy when linearly combined. Another aspect evaluated by the authors is the reduction of uncertainty in the selection of the model, in order to avoid deterioration in the forecast of the time series.

Regarding the techniques that use the SVR, one traditional form to choose the set of parameters, proposed by Hsu *et al.* [49], is to combine all the possibilities of a pre-established range of parameters, performing a grid search to find the optimal or sub-optimal combination of the parameter's values. However, depending on the number of parameters, this strategy can be costly. Some works use optimization algorithms, such as PSO, to select the parameters of machine learning techniques [20,30,48,50–52].

Shamsuddin and Sallehuddin [52] proposed a hybrid model 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 [30], 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 [20] the PSO is used to find the best hyperparameters for SVR models to perform residual forecasting in a hybrid proposal. In order to avoid problems from the optimization, a subset of the population is selected and combined to perform forecasts.

Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a traditional forecasting method introduced by Box and Jenkins [24]. According to Zhang [34], 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,

$$y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}, \tag{1}$$

in which y_t and ε_t are the actual and error values at time period t, respectively. The model coefficients are μ, $\theta_i (i = 1, 2, ..., q)$ 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 [34].

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 [53], 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 it is 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

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

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

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

where C is a regularization factor, $||.||^2$ is a L2 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

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

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

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

subject to

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

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

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

in which α and α^* are Lagrange multipliers, $k(\mathbf{x}_i, \mathbf{x})$ is a kernel function, and 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

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

where γ is a parameter of the RBF kernel.

Particle Swarm Optimization

Particle Swarm Optimization (PSO) [54] 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, the best personal position is calculated, called $pbest$. After that, in the next iteration, the particles move according to the velocity function, as

$$v_i(t+1) = w \times v_i(t) + c_1 \times r_1 \times (pbest_i - x_i(t)) \; + c_2 \times r_2 \times (gbest_i - x_i(t)), \tag{9}$$

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 i-th element of the current solution x.

3. PROPOSED APPROACH

In this paper a new alternative is proposed for time series forecasting using a nonlinear approach optimized by the PSO with different topologies, called SVR(ARIMA, PSOSVR). The hybrid forecasting system can be divided in two phases: the prediction stage and the combination one. Furthermore, we can subdivide the first stage into two groups: linear and nonlinear modeling. 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 residuals, which are modeled by a nonlinear technique. The second stage of the system consists of combining the forecasts, from the original and the residual series. This combination can be linear, adding both forecasts, or nonlinear, performed by some machine learning algorithm, for example. In this paper, the combination of linear and non-linear predictions is made by a non-linear technique.

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 also 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 [55]. Due to this, the choice of topology and coefficient of inertia to be used is incorporated into the model, which may vary depending on the characteristics of the dataset. All combinations, among a set of established parameters, are tested, and the best model is returned, that is, the one that leads to the smallest MSE. This approach makes the hybrid system more flexible to different datasets since different topologies can be applied to the model, expanding the range of possibilities to be tested in defining the hyperparameters of the nonlinear technique. The architecture of the first stage proposal is shown in Figure 1.

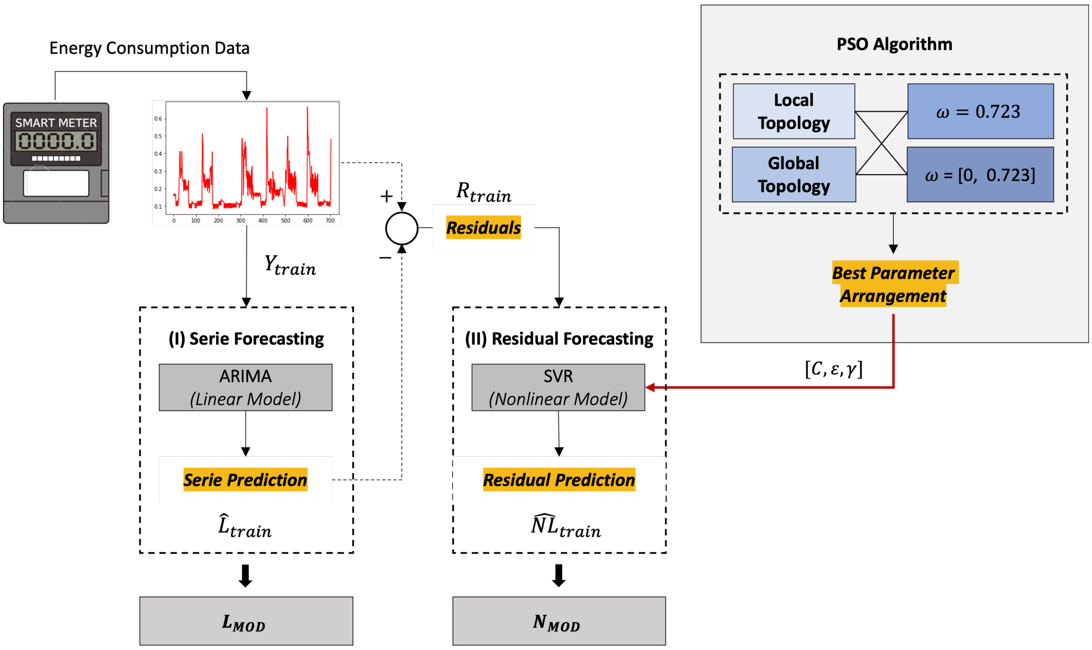


Figure 1: First stage of the proposed architecture (Steps I and II).

In the proposed method, the Autoregressive Integrated Moving Average (ARIMA) is used as the first model, performing the linear patterns prediction ($\hat{L}_{train}$). After that, the residuals (R_{train}) are calculated from the difference between the original data (Y_{train}) and the prediction one ($\hat{L}_{train}$). Then, using the residual series (R_{train}), an SVR is trained to model the nonlinear patterns that were not modeled by the ARIMA, resulting in the residual prediction, denoted by $\hat{N}_{train}$. The SVR parameters are chosen by the PSO algorithm, according to the lowest mean squared error of the prediction. Since there is not a universal optimal model for all databases, the choice of the topology used by the PSO is incorporated into the system, varying between local and global topology, and with or without linear decay.

As shown in Figure 1, the output of Step (I) is an estimated linear model (L_{MOD}), that will be used in the test stage. The output of Step (II) is an estimated nonlinear model (N_{MOD}). To perform the residual forecasting, R_{train} is subdivided into training and validation sets used to SVR parameters estimation by the PSO. The purpose of Step (III) is to make the most suitable combination between the linear forecast ($\hat{L}_{train}$) and nonlinear one ($\hat{N}_{train}$), resulting in the final prediction ($\hat{Y}_{train}$). For this, the SVR technique is used, as shown in Figure 2. The SVR parameters are obtained by an exhaustive search, using grid search. This step results in the estimated combination model, denoted by C_{MOD}.

In the second stage, shown in Figure 3, the energy consumption data from the smart meters are the input data to the hybrid system, which has the trained models. The steps are the same as in the training stage. First, the data are modeled by a linear model ($\hat{L}_{test}$), and the residual series is obtained by the difference between the original data and the linear prediction. Therefore,

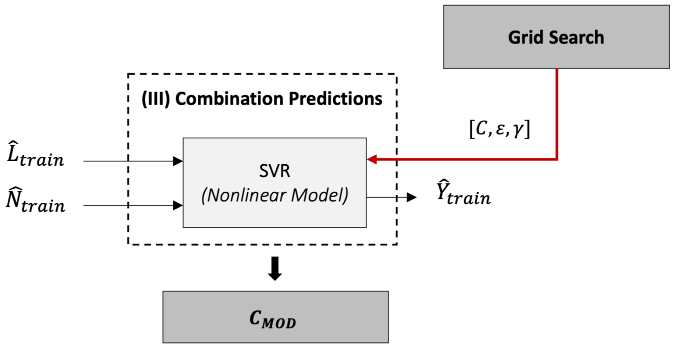


Figure 2: First stage of the proposed architecture (Step III).

the nonlinear patterns are captured by a nonlinear model ($\hat{N}_{test}$), resulting in the second prediction. Lastly, the predictions are combined using a nonlinear model, giving the final prediction of the system.

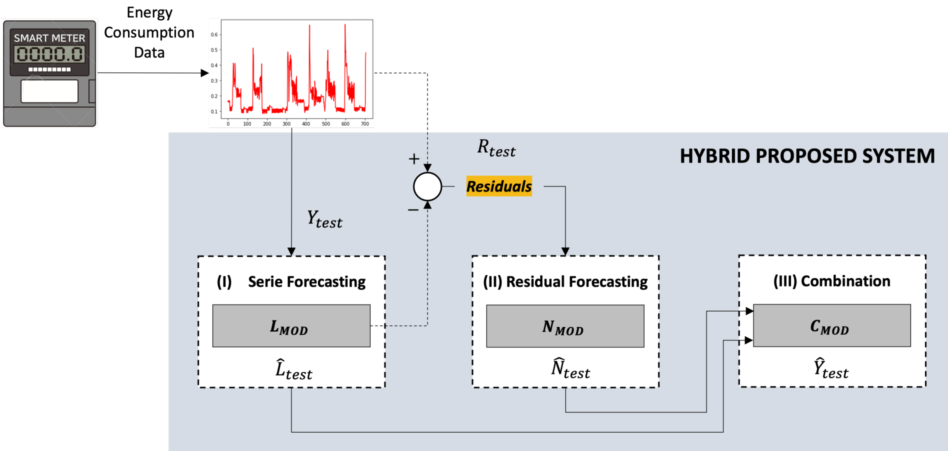


Figure 3: Second stage of the proposed architecture.

4. EXPERIMENTAL SETUP

Data. The data series is composed by the energy consumption of a household in the Xindian district, in Taiwan, for approximately one month [8]. The residence has three floors, totaling $350\mathbf{m^2}$, and its occupancy during the measurements was five people: a couple of adults and three children. The data were obtained through a smart grid infrastructure, in which smart meters were installed to record the energy consumption in each room. Data were collected at 15-minute intervals, totaling 96 times per day. For this paper, the total consumption of the residence was considered, that is, the sum of the individual measurements of each room. The consumption curve, containing 2.879 samples, is shown in Figure 4. In addition to energy consumption, two external information also compose the database: day of the week and temperature. In this work, the exogenous variable of temperature is not considered. For simulation purposes, the database was partitioned into seven series according to the day of the week: Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday.

The databases were partitioned with approximately 50% of the data for training, 25% for validation, and the remaining 25% for testing. In the approaches that use the PSO meta-heuristic, the fitness is performed by calculating the mean squared error of the residual prediction, from the SVR, for each particle. To avoid overlapping samples at different stages of the system, the training set was divided into two series: training and validation. All the databases were normalized, that is, represented in the range [0,1].

In the 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. These values are configured according to [20, 22, 30, 41]. The description and partitioning of the data are shown in Table 1.

Selection of Parameters. The parameter selection for the linear and non-linear approaches is made in different ways. To the ARIMA model, the choice of parameters (p, q, d) was made according to the Hyndman methodology [56], through the **autoarima** function, which chooses them to minimize the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [57]. For the SVR, the parameters can be set in two ways. In the optimized approaches, the choice of these values

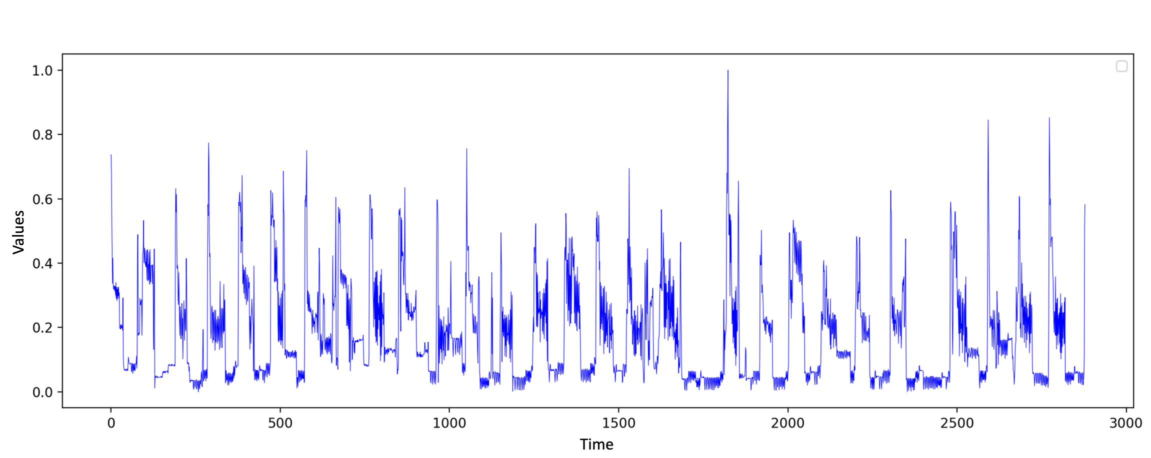


Figure 4: Normalized Energy Consumption curve obtained by smart meters of a household in Taiwan.

Table 1: Description and partitioning of databases.

Database	Description	Total	Train	Validation	Test	Lags
Sunday	Energy Consumption from a household on Sundays	384	192	96	96	12
Monday	Energy Consumption from a household on Mondays	384	192	96	96	12
Tuesday	Energy Consumption from a household on Tuesdays	383	191	96	96	12
Wednesday	Energy Consumption from a household on Wednesdays	480	240	120	120	12
Thursday	Energy Consumption from a household on Thursdays	480	240	120	120	12
Friday	Energy Consumption from a household on Fridays	384	192	96	96	12
Saturday	Energy Consumption from a household on Saturdays	384	192	96	96	12

is performed by the PSO. On the other hand, 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. The parameters of the SVR model are defined by PSO according to the options described in Table 2. Regarding the PSO technique, the parameters used are well-established in the literature [58]. They are also shown in Table 2.

Table 2: List of Parameters.

Model	Parameters	Values
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
	Number of simulations	20

Comparison Papers. To carry out a comparative analysis, four methods were considered: (i) A single ARIMA approach, proposed by Zhang [34]; (ii) A linear combination using ARIMA and SVR, by de Oliveira and Ludermir [30]; (iii) the optimized version proposed by de Holanda and de Oliveira [20]; (iv) A nonlinear combination using SVR, presented by de Mattos Neto *et al.* [31]. The previously mentioned approaches are represented by the acronyms presented in Table 3.

Simulations. The implementation of the system was carried out using the R and python languages. For the estimation of the ARIMA parameters, the **autoarima** function was used in the forecast package, available in the R libraries [56]. The other

Table 3: Acronyms of the comparison approaches.

Approach	Acronyms
(i)	ARIMA
(ii)	ARIMA + SVR
(iii)	ARIMA + PSO-SVR
(iv)	SVR(ARIMA,SVR)

implementations were performed in python. Concerning the SVR, the Sklearn library was used. The PSO was fully implemented by the authors.

Evaluation Metrics. The evaluation of the methods was performed by using three well-known performance measures in the literature [22,34,35]: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE), given by the equations 10, 11, 12, respectively.

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}(y_t - y'_t)^2, \tag{10}$$

$$\text{MAPE} = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{y_t - y'_t}{y_t}\right|, \tag{11}$$

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

in which N is the database 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.

To make a direct comparison from the proposed approach with other methods, the percentage difference (PD) is calculated between the proposed hybrid system and literature models,

$$PD = 100 \times \frac{Model_L - Model_P}{Model_L}, \tag{13}$$

where $Model_P$ refers to the results from the proposed method and $Model_L$ to the literature ones. Positive PD values indicate that the result of the proposed technique was better than the ones of the compared technique, while negative values indicate that the proposal obtained worse results.

To evaluate the variability of the results from techniques that used the PSO, the coefficient of variation was calculated, according to

$$CV = 100 \times \frac{\sigma}{a}, \tag{14}$$

in which σ is the standard deviation and a is the average. The coefficient of variation is an attractive statistical metric because it permits the comparison of variables free from scale effects, there is, it is dimensionless [59].

5. RESULTS

In order to compare the performance of the proposed method with other techniques from the literature, analyzes were performed in terms of three metrics: MSE, MAE, and MAPE. The percentages differences between the proposal and the other approaches also are evaluated. In addition, the coefficient of variation was evaluated for approaches that use the PSO. As it is a non-deterministic technique, 20 simulations were performed for each date, whose results are presented by their mean and standard deviation.

The results concerning the MSE are presented in Table 4, in which the best results are highlighted in bold. In terms of this metric, the proposed method outperformed the others in four out of seven cases: Monday, Tuesday, Wednesday, and Saturday. One may note that the proposed method overperforms ARIMA to all data, with the exception of Friday.

Results of MAE are shown in Table 5. The proposed method presented superiority in most database. More specifically, it obtained the best results in four out of seven cases: Monday, Tuesday, Wednesday, and Thursday. For the Sunday data, the result from the proposal is close to that of the ARIMA+PSO-SVR, with a difference of 0.0004.

The results are also analyzed under the MAPE criterion. The standard deviation is presented in percentage points (p.p.). According to this metric, the proposed method outperformed all techniques in six out of seven cases: Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday, as one can see in Table 6.

Table 4: Mean Square Errors (MSE) for all databases. The best results are highlighted in bold.

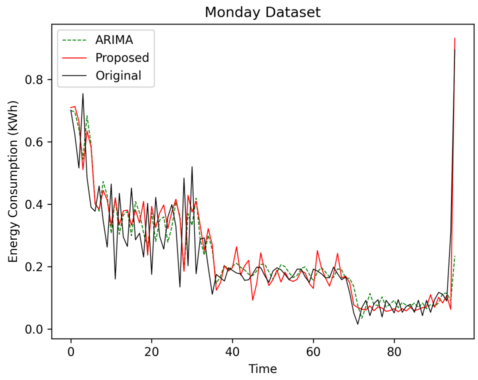
database	ARIMA	ARIMA+SVR	SVR(ARIMA,SVR)	ARIMA+PSO-SVR	Proposed
Monday	0.01373	0.01162	0.01162	0.01042 ± 0.00021	$\mathbf{0.00969} \pm 0.00004$
Tuesday	0.00830	0.00860	0.00849	0.00829 ± 0.00005	$\mathbf{0.00810} \pm 0.00010$
Wednesday	0.00940	0.00910	0.00903	0.00860 ± 0.00018	$\mathbf{0.00851} \pm 0.00016$
Thursday	0.01258	$\mathbf{0.01142}$	$\mathbf{0.01142}$	0.01233 ± 0.00011	0.01223 ± 0.00015
Friday	0.00926	$\mathbf{0.00911}$	0.00925	0.00958 ± 0.00000	0.00945 ± 0.00000
Saturday	0.00712	0.00664	0.00665	0.00661 ± 0.00018	$\mathbf{0.00633} \pm 0.00026$
Sunday	0.00451	0.00427	0.00427	$\mathbf{0.00423} \pm 0.00003$	0.00425 ± 0.00005

Table 5: Mean Absolute Errors (MAE) for all databases. The best results are highlighted in bold.

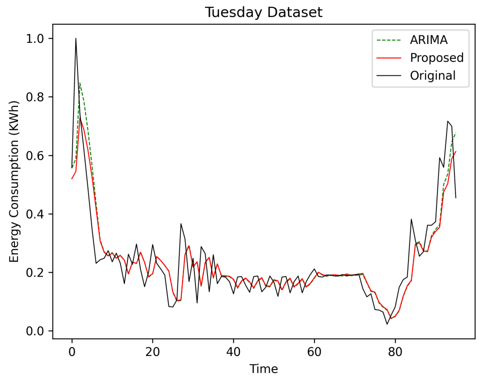
database	ARIMA	ARIMA+SVR	SVR(ARIMA,SVR)	ARIMA+PSO-SVR	Proposed
Monday	0.0726	0.0709	0.0709	0.0745 ± 0.0019	$\mathbf{0.0682} \pm 0.0001$
Tuesday	0.0586	0.0605	0.0608	0.0588 ± 0.0005	$\mathbf{0.0580} \pm 0.0005$
Wednesday	0.0574	0.0555	0.0558	0.0557 ± 0.0020	$\mathbf{0.0553} \pm 0.0008$
Thursday	0.0731	0.0703	0.0703	0.0708 ± 0.0006	$\mathbf{0.0701} \pm 0.0008$
Friday	0.0500	$\mathbf{0.0447}$	0.0468	0.0553 ± 0.0000	0.0464 ± 0.0000
Saturday	0.0554	$\mathbf{0.0485}$	0.0501	0.0498 ± 0.0009	0.0500 ± 0.0010
Sunday	0.0315	0.0293	0.0294	$\mathbf{0.0290} \pm 0.0009$	0.0294 ± 0.0010

Table 6: Mean Absolure Percentage Errors (MAPE) for all databases. The best results are highlighted in bold.

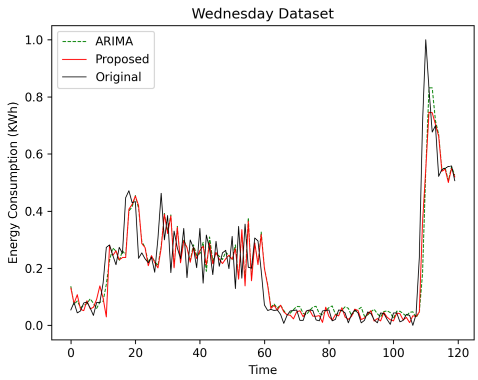
database	ARIMA	ARIMA+SVR	SVR(ARIMA,SVR)	ARIMA+PSO-SVR	Proposed
Monday	39.80%	38.94%	38.94%	$50.15\% \pm 1.98p.p$	$\mathbf{37.17}\% \pm 0.06p.p$
Tuesday	27.89%	29.39%	29.72%	$27.93\% \pm 0.31p.p$	$\mathbf{27.78}\% \pm 0.30p.p$
Wednesday	74.22%	53.76%	52.49%	$52.86\% \pm 2.28p.p$	$\mathbf{51.02}\% \pm 1.81p.p$
Thursday	81.63%	74.38%	74.36%	$74.29\% \pm 2.66p.p$	$\mathbf{71.38}\% \pm 2.86p.p$
Friday	66.28%	52.36%	52.55%	$77.20\% \pm 0.01p.p$	$\mathbf{46.90}\% \pm 0.04p.p$
Saturday	133.02%	72.61%	76.83%	$67.80\% \pm 4.91p.p$	$\mathbf{66.67}\% \pm 3.79p.p$
Sunday	179.02%	131.13%	$\mathbf{130.83}\%$	$141.89\% \pm 11.75p.p$	$137.49\% \pm 13.48p.p$



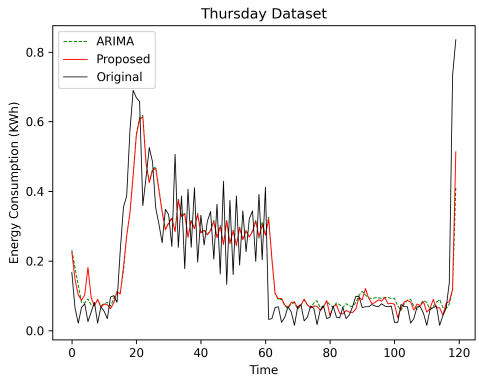
(a) *Monday*



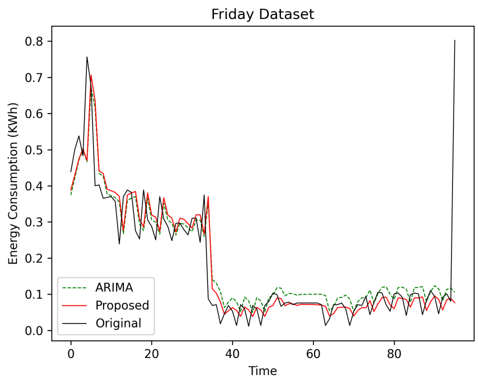
(b) *Tuesday*



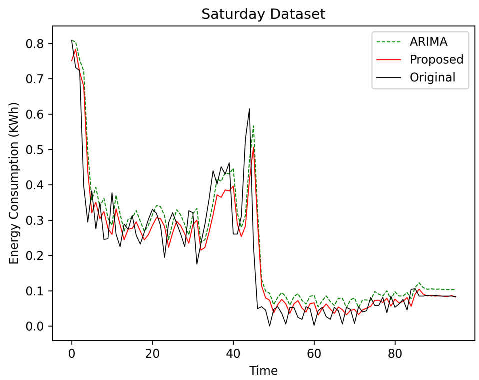
(c) *Wednesday*



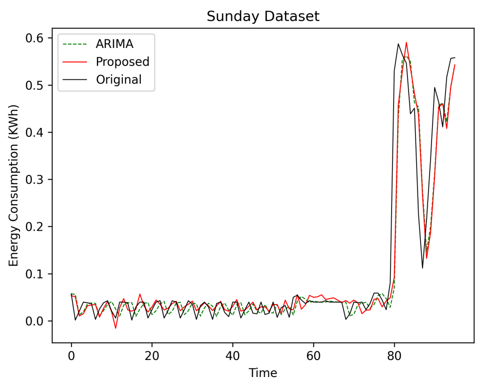
(d) *Thursday*



(e) *Friday*



(f) *Saturday*



(g) *Sunday*

Figure 5: Forecast curves for the test set with ARIMA and the proposed method.

Figure 5 presents the curves referring to the original time series, the ARIMA prediction and the proposed approach prediction, for Taiwan's building database, for each day of the week. One may note that the proposed method's curve is closer to the original one than the curve of the ARIMA method.

Table 7 shows the percentage difference between the proposed method and literature models in terms of MSE, for all databases used. Compared to ARIMA, the greatest results can be noted for the Monday and Saturday data. For these cases, the proposal outperformed ARIMA, in terms of percentage difference, 29.42% and 11.05%, respectively. To Monday data, in comparison to the ARIMA+SVR and SVR(ARIMA, SVR), the proposed method surpass them with an improvement of 16.63% and 16.61%, respectively. Regarding Sunday data, the proposal overperformed all techniques, with the exception of ARIMA+PSO-SVR, with an inferiority of 0.02 percentual points.

Table 7: The percentage difference (%) between the proposed method and literature models according to Equation (13), in terms of MSE.

database	ARIMA	ARIMA+SVR	SVR(ARIMA,SVR)	ARIMA+PSO-SVR
Monday	29.42	16.63	16.61	7.01
Tuesday	2.42	5.77	4.65	2.34
Wednesday	9.42	6.44	5.74	0.97
Thursday	2.81	-7.04	-7.05	0.86
Friday	-2.08	-3.72	-2.25	1.30
Saturday	11.05	4.54	4.75	4.22
Sunday	5.79	0.41	0.59	-0.57

According to the coefficient of variation (CV) shown in Table 8, the proposed approach presents less variability for two out of seven days of the week and has the same performance for the Friday data.

Table 8: Coefficient of Variation, in percentage (%), of the techniques which use PSO.

Model	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
ARIMA + PSO-SVR	2.06	**0.66**	2.12	**0.88**	**0.03**	**2.75**	**0.80**
Proposed Method	**0.37**	1.18	**1.87**	1.26	**0.03**	4.03	1.25

6 CONCLUSION

Energy generation and distribution systems have undergone several changes over the past few years. One of the motivations for these changes is the need to optimize energy consumption in order to reduce impacts on the environment, such as high carbon emissions. In this context, smart grids have proven to be a great alternative. One of the main advantages of this type of network is the facility of monitoring measurements, allowing for better management and knowledge of the consumption pattern, which result in a more rational use of energy. For this reason, the development of load curve prediction models has become an important tool for planning power distribution systems.

In this paper, an optimized hybrid system is proposed for load forecasting from smart meters. In the first stage of the system, the technique performs a training step, split into three sequential phases. Initially, the original series is modeled by ARIMA to map the linear patterns. The difference between the original values and the prediction of this model results in the residual series. Residuals are modeled by an optimized version of the SVR, which is the innovation of the proposed approach. The choice of SVR parameters is carried out by the PSO algorithm. Also, the choice of topology and inertia coefficient used in PSO are incorporated into the model. The predictions of the original and residual series are combined by the SVR. In this step, the SVR parameters are chosen by an exhaustive search.

The experiments were carried out on real data of a smart grid in a building in Taiwan. Measurements were taken at 15-minute intervals for approximately one month. For simulation purposes, the full database was subdivided in seven small database, according to the days of the week. In order to assess the performance of the proposal, the results were compared with four techniques well-established in the literature for time series forecasting.

The results show that the proposed technique achieved the best overall results in terms of all the metrics used. Compared to the ARIMA model, the proposed method presents better results for all seven days of the week in all three metrics, with the exception of Friday, concerning the MSE. It is important to highlight the performance of the proposed method in terms of MSE percentage difference, for the databases Monday and Saturday, in which the proposal outperformed ARIMA in 29.42% and 11.05%, respectively. When compared with all the other evaluated techniques, specifically for MAPE, the proposal performed better in six out of seven databases. The proposed technique obtained better performance in four out of seven databases, in terms of MSE and MAE. These results show that the use of optimized models that vary their topology according to each database can improve the accuracy of predictions. In addition, it is important to emphasize that the choice of parameters leads to more assertive forecasts.

Concerning consumption prediction, overestimating can result in a waste of energy and, consequently, in a higher acquisition cost. In the scenario of smart meters and contracted energy on demand, the estimate is expected to be as close as possible, avoiding interruptions in supply or waste of resources. The evaluation of the acceptable error may vary according to each application, however the proposed model aim to be as assertive as possible.

For future works, it is intended to investigate the use of exogenous variables, such as temperature, in order to improve the prediction results, as well as to carry out an investigation regarding the choice of lags. In addition, the choice of different computational intelligence techniques may also be explored.

7 ACKNOWLEDGEMENTS

The authors would like to thank the Foundation for Science and Technology Support from Pernambuco (FACEPE) - Process No. APQ-1252-1.03/21, for the financial support.

REFERENCES

[1] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin and Z. Hu. “Photovoltaic and Solar Power Forecasting for Smart Grid Energy Management”. *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38–46, 2015.

[2] R. Yao and K. Steemers. “A Method of Formulating Energy Load Profile for Domestic Buildings in the UK”. *Energy and Buildings*, vol. 37, no. 6, pp. 663–671, 2005.

[3] A. Allouhi, Y. El Fouih, T. Kousksou, A. Jamil, Y. Zeraouli and Y. Mourad. “Energy Consumption and Efficiency in Buildings: Current Status and Future Trends”. *Journal of Cleaner Production*, vol. 109, pp. 118–130, 2015.

[4] J.-S. Chou and D.-S. Tran. “Forecasting Energy Consumption Time Series Using Machine Learning Techniques Based on Usage Patterns of Residential Householders”. *Energy*, vol. 165, pp. 709–726, 2018.

[5] C. Reinisch, M. Kofler, F. Iglesias and W. Kastner. “Thinkhome Energy Efficiency in Future Smart Homes”. *EURASIP Journal on Embedded Systems*, vol. 2011, pp. 1–18, 2011.

[6] Y. M. Lee, L. An, F. Liu, R. Horesh, Y. T. Chae and R. Zhang. “Applying Science and Mathematics to Big Data for Smarter Buildings”. *Annals of the NEW YORK Academy of Sciences*, vol. 1295, no. 1, pp. 18–25, 2013.

[7] M. E. El-Hawary. “The Smart Grid State of the Art and Future Trends”. *Electric Power Components and Systems*, vol. 42, no. 3-4, pp. 239–250, 2014.

[8] J.-S. Chou and N.-T. Ngo. “Time Series Analytics Using Sliding Window Metaheuristic Optimization-Based Machine Learning System for Identifying Building Energy Consumption Patterns”. *Applied Energy*, vol. 177, pp. 751–770, 2016.

[9] K. Gajowniczek and T. Zkabkowski. “Short Term Electricity Forecasting Using Individual Smart Meter Data”. *Procedia Computer Science*, vol. 35, pp. 589–597, 2014.

[10] C.-N. Yu, P. Mirowski and T. K. Ho. “A Sparse Coding Approach to Household Electricity Demand Forecasting in Smart Grids”. *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 738–748, 2016.

[11] D. M. Izidio, P. S. de Mattos Neto, L. Barbosa, J. F. de Oliveira, M. H. N. Marinho and G. F. Rissi. “Evolutionary Hybrid System for Energy Consumption Forecasting for Smart Meters”. *Energies*, vol. 14, no. 7, pp. 1794, 2021.

[12] M. N. Fekri, H. Patel, K. Grolinger and V. Sharma. “Deep Learning for Load Forecasting with Smart Meter Data: Online Adaptive Recurrent Neural Network”. *Applied Energy*, vol. 282, pp. 116177, 2021.

[13] H. X. Zhao and F. Magoulès. “A Review on the Prediction of Building Energy Consumption”. *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.

[14] 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.

[15] 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.

[16] 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, pp. 1186, 2017.

[17] A. B. Nassif. “Short Term Power Demand Prediction Using Stochastic Gradient Boosting”. In *5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, pp. 1–4. IEEE, 2016.

[18] 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.

[19] 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*, vol. 32, pp. 1 – 13, 2021.

[20] 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)*, pp. 595–602. IEEE, 2020.

[21] 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. 227719 – 227728, 2020.

[22] 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.

[23] L. Xiao, W. Shao, M. Yu, J. Ma and C. Jin. “Research and Application of a Combined Model Based on Multi-Objective Optimization for Electrical Load Forecasting”. *Energy*, vol. 119, pp. 1057–1074, 2017.

[24] G. E. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] W.-C. Hong. “Electric Load Forecasting by Support Vector Model”. *Applied Mathematical Modelling*, vol. 33, no. 5, pp. 2444–2454, 2009.

[30] 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)*, pp. 3867–3872. IEEE, 2014.

[31] 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.

[32] V. Vapnik, S. E. Golowich and A. Smola. “Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing”. *Advances in Neural Information Processing Systems*, pp. 281–287, 1997.

[33] D. H. Wolpert and W. G. Macready. “No Free Lunch Theorems for Optimization”. *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.

[34] G. P. Zhang. “Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model”. *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[35] 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.

[36] B. Wang, N. Tai, H. Zhai, J. Ye, J. Zhu and L. Qi. “A new ARMAX Model Based on Evolutionary Algorithm and Particle Swarm Optimization for Short-Term Load Forecasting”. *Electric Power Systems Research*, vol. 78, no. 10, pp. 1679–1685, 2008.

[37] 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.

[38] 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.

[39] D. Ö. Faruk. “A Hybrid Neural Network and ARIMA Model for Water Quality Time Series Prediction”. *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586–594, 2010.

[40] C. N. Babu and B. E. Reddy. "A Moving-Average Filter Based Hybrid ARIMA–ANN Model for Forecasting Time Series Data". *Applied Soft Computing*, vol. 23, pp. 27–38, 2014.

[41] J. F. L. de Oliveira and T. B. Ludermir. "A Hybrid Evolutionary Decomposition System for Time Series Forecasting". *Neurocomputing*, vol. 180, pp. 27–34, 2016.

[42] 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.

[43] Z. H. Rahimi and M. Khashei. "A Least Squares-Based Parallel Hybridization of Statistical and Intelligent Models for Time Series Forecasting". *Computers & Industrial Engineering*, vol. 118, pp. 44–53, 2018.

[44] S. Siami-Namini, N. Tavakoli and A. S. Namin. "A Comparison of ARIMA and LSTM in Forecasting Time Series". In *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1394–1401. IEEE, 2018.

[45] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[46] Z. Hajirahimi and M. Khashei. "Sequence in Hybridization of Statistical and Intelligent Models in Time Series Forecasting". *Neural Processing Letters*, pp. 1–21, 2020.

[47] P. S. de Mattos Neto, J. F. L. de Oliveira, D. S. de Oliveira, H. V. Siqueira, M. H. N. Marinho and F. Madeiro. "A Hybrid Nonlinear Combination System for Monthly Wind Speed Forecasting". *IEEE Access*, vol. 8, pp. 191365–191377, 2020.

[48] 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, pp. 117948, 2020.

[49] 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].

[50] 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.

[51] 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.

[52] 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.

[53] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2013.

[54] J. Kennedy and R. Eberhart. "Particle Swarm Optimization". In *International Conference on Neural Networks (ICNN)*, volume 4, pp. 1942–1948. IEEE, 1995.

[55] 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*, pp. 1–8. IEEE, 2012.

[56] R. Hyndman. "Time Series Data Library". 2015. [Online; accessed 12-June-2021].

[57] 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.

[58] Y. Shi and R. C. Eberhart. "Parameter Selection in Particle Swarm Optimization". In *International Conference on Evolutionary Programming*, pp. 591–600. Springer, 1998.

[59] C. E. Brown. "Coefficient of Variation". In *Applied Multivariate Statistics in Geohydrology and Related Sciences*, pp. 155–157. Springer, 1998.