Research Article

Electrical load forecasting in disaggregated levels using Fuzzy ARTMAP artificial neural network and noise removal by singular spectrum analysis

M. R. Müller[1] · G. Gaio[2] · E. M. Carreno[2] · A. D. P. Lotufo[1] · L. A. Teixeira[3]

© Springer Nature Switzerland AG 2020

Abstract

Electrical load forecasting in disaggregated levels is a difficult task due to time series randomness, which leads to noise and consequently affects the quality of predictions. To mitigate this problem, noise removal using singular spectrum analysis (SSA) is used in this work in conjunction with a Fuzzy ARTMAP artificial neural network, presenting excellent results when compared with traditional methods like SARIMA. A reduction of almost 50% on the MAPE is achieved. The SSA method is preferable to other filtering methods because it has a low computational cost, depends on a small number of parameters, requires few data to present good results, and it does not cause delay into the denoised series.

1 Introduction

A time series is loosely defined as a set containing measures of a random variable of interest, ordered in time. Thus, an electric energy load curve is a time series; consequently, the time series prediction and analysis techniques can be applied to estimate the future behavior of electric power demand [1].

The capacity to forecast electrical power demand time series is associated with the efficiency and efficacy of electric energy planning, which improves the financial results of the players in the industry [2, 3]. Once the predictors are robust and accurate, the results are used as input in tasks such as expansion planning, economical operation, security analysis, and control. So, if the demand is not adequately forecast, it can lead to financial losses for the utilities as well as for the users. The utilities lose when buying electrical energy with high costs, energy losses, and possible high financial penalties, depending on market design.

The users are penalized according to their consumer class, with power interruptions and low-quality indexes.

Introducing information technologies in distribution networks, a process known as Smart Grids, will facilitate data acquisition at several aggregation levels. Electrical demand data are currently available for long time periods, with high resolution and in scenarios previously ignored, for example, individual consumers or small groups [4]. This data can be called disaggregated, as it is collected in more detailed levels than the usual; in a load curve context, it means to approach individual consumption levels in high time resolution.

There are some works on the literature dealing with these kinds of data, and they can be divided in two groups, such as residential [5–11] and nonresidential [12–19].

Thus, the electrical energy demand time series become larger, more disaggregated, and more detailed. These characteristics bring difficulties to the forecasting process, specially due the low aggregation level. It consists

✉ E. M. Carreno, edgar.franco@unioeste.br | [1]Electrical Engineering Department, UNESP-FEIS, Ilha Solteira, SP, Brazil. [2]Engineering and Exact Sciences Center - CECE, Western Parana State University – UNIOESTE, Foz do Iguaçu, PR, Brazil. [3]UNILA, Federal Latin-American Integration University, Foz do Iguaçu, PR, Brazil.

of data with high variability, resulting from the fact that it depends on the number of persons consuming simultaneously, the number of electric equipment turned on, etc. In contrast, in aggregated levels the fluctuations and noises of the individual consumers may cancel each other out while taking the sum [3].

The literature presents several prediction methods, which can, roughly, be divided in two groups: statistical and computational intelligence. The statistical includes, among others, the ARIMA family (Autoregressive integrated moving average), such as ARIMA, Seasonal ARIMA, and ARIMAX; another group contains ANNs (artificial neural networks), fuzzy logic, among others [20]. Each method presents its own advantages and disadvantages, depending on the characteristics of the time series, data quantity, and others.

For predictions in disaggregated levels, as the consumption can fluctuate heavily even in consecutive hours or consecutive days from any regular pattern [3], the principal characteristics to consider are the resolution and the data quantity, the randomness of the time series, atypical days, and noise levels.

For noise removal, this paper uses singular spectrum analysis (SSA), which is an algorithm with advantages over others used for the same task. SSA requires only two parameters to be chosen: window length and the eigenvectors used for the reconstructing. Analytical methods are not available to determine the former; nevertheless, the results are not overly sensitive to small variations of it [21]. For the latter, automatic selection strategies can be adopted based on hierarchical clustering. Unlike traditional methods, this technique works adequately for stationary signals as well as non-stationary signals. It is easily implementable and has low computational costs [22].

Noise removal is not common in load forecasting literature because most works are based on low-resolution data from large aggregated areas, where the problem is not common. However, some works show the advantages of preprocessing the data sets. For example, Ref. [23] showed accuracy in applying a moving average filter with delay corrector to the data to predict substation load demand with a general regression neural network (GRNN). In Ref. [24], an SSA filter is used to remove noise in the time series to be predicted in a feedforward neural network, thus improving the results as compared to prediction with rough data. The authors in [25] attained good results by applying SSA to further use autoregressive models for short-term load forecasting. Ref. [26] performed electrical load forecasting using an MLP (multi-layer perceptron) neural network optimized by metaheuristics and reducing noise with SSA. The authors in [27] presented good results using support vector regression optimized by the Cuckoo Search algorithm, as well as reducing the noise with SSA.

In the other hand, Fuzzy ARTMAP ANN has been shown to be a good option in terms of load forecasting. For example, references [28–32] have presented good results in comparison with other forecasting methods.

Therefore, it seems to be a logical approach to combine these two techniques aiming better forecast results. However, as there is no record in the literature of applying SSA together with Fuzzy ARTMAP ANN for load forecasting, this is an original contribution of this paper. In addition, in most of the literature, the load curves present less resolution and higher level of aggregation of electrical loads than the ones studied in this work (large geographical regions). The work herein is significantly different from previous works, in that disaggregated data are used along with a neural network based on adaptive resonance theory.

The data used in this work presents days with atypical load curves. It happens due to weekends, holydays and others. In disaggregated levels, they may cause a huge variation from the usual pattern [7]. Traditional predictors cannot incorporate the analyst's previous knowledge to the model without adaptations. According to [33], the loads of these atypical days are different from other days and cannot be predicted, thus damaging the prediction of typical days.

Ref. [33] showed that the treatment of atypical days is not common in the literature, and sometimes they are substituted on the original series by forecasting within the sample for that day or a pre-determined day, manipulating the data to consider it as a typical day [34, 35]. This approach, however, does not consider the information contained that possibly could aid modeling the behavior of the following days. Another approach consists of separating atypical days from weekends and labor days by introducing a dummy variable, according to [33, 36, 37].

However, the causes of such behavior are not always known; therefore, it is important that the predictor is able to mitigate this situation.

This work compares three forecasting methods: a statistical model, SARIMA (seasonal autoregressive integrated moving average), because the data is seasonal; and two others from computational intelligence, i.e., MLP and Fuzzy ARTMAP, applied to short-term load forecasting (1, 3, and 7 days), with and without noise removal. The data originates from an intelligent microgrid (smart micro grid) in Brazil. Identifying weekends and atypical days is a task that the Fuzzy ARTMAP performs very well and is superior compared to the other methodologies tested, which cannot identify a normal day from an atypical day or a weekend. The noise removal with SSA for further modeling by the Fuzzy ARTMAP improves the forecasted results.

2 Scenario of study

This work uses real data from a smart microgrid in Foz do Iguassu, Brazil, and the load curve is composed of commercial sites, startups, research centers, laboratories, offices, and services.

Two scenarios of the same smart microgrid were studied, one composed by 46 daily measurements with interval of 30 min between them, and another, with 96 daily data, with an interval of 15 min. The patterns sets correspond to January, February, and March of 2017 (from January 1 to March 31). The data is available upon request from the authors.

The labor days are different and present variations; otherwise, the weekend demands are lower than that of labor days. In addition, atypical days exist in the time series used. Figure 1 presents the week prior to a national holiday in Brazil (Carnival), the week of the holiday, and the week following to the holiday (February 19 to March 11).

3 Fuzzy ARTMAP artificial neural network

The first mathematical neuron based on a biological neuron was presented in [38], leading to the first artificial neural network. This work forms the basis of further computational implementations [39], based on the ability of live organisms to assimilate knowledge when facing new situations.

ANNs can present answers even with data different from the training sets, which is the capacity to generalize.

Different ANN architectures have been proposed to improve beyond the limits of previous ones or to perform new tasks [40].

Adaptive resonance theory (ART) neural networks are examples of an ANN development. ART theory solves the plasticity/stability problem, where the stability assures that every element has a category created, being plasticity the capacity to assume new patterns without losing the previously acquired knowledge.

Fuzzy ARTMAP ANN pertains to the supervised training paradigm, with the calculus based on fuzzy logic. The architecture is composed of two Fuzzy ART modules (ART_a and ART_b), where the ART_a module processes the input vector and the ART_b module processes the desired output vector.

The associative memory module Inter-ART receives the inputs of the ART_a module (referred to as the associative connection $J \rightarrow K$, where J and K are the active categories in modules ART_a and ART_b, respectively. This module contains an auto-regulator mechanism called Match Tracking, which matches the categories of ART_a with ART_b [17]. The Fuzzy ARTMAP architecture is shown in Fig. 2.

First, none of the categories are active; therefore, the weight matrix (w) is 1. As the pairs in ART_a and ART_b are confirmed (according to *Match-Tracking*), the categories become active.

Below, the steps of Fuzzy ARTMAP ANN are shown:

- *Initial Module*:
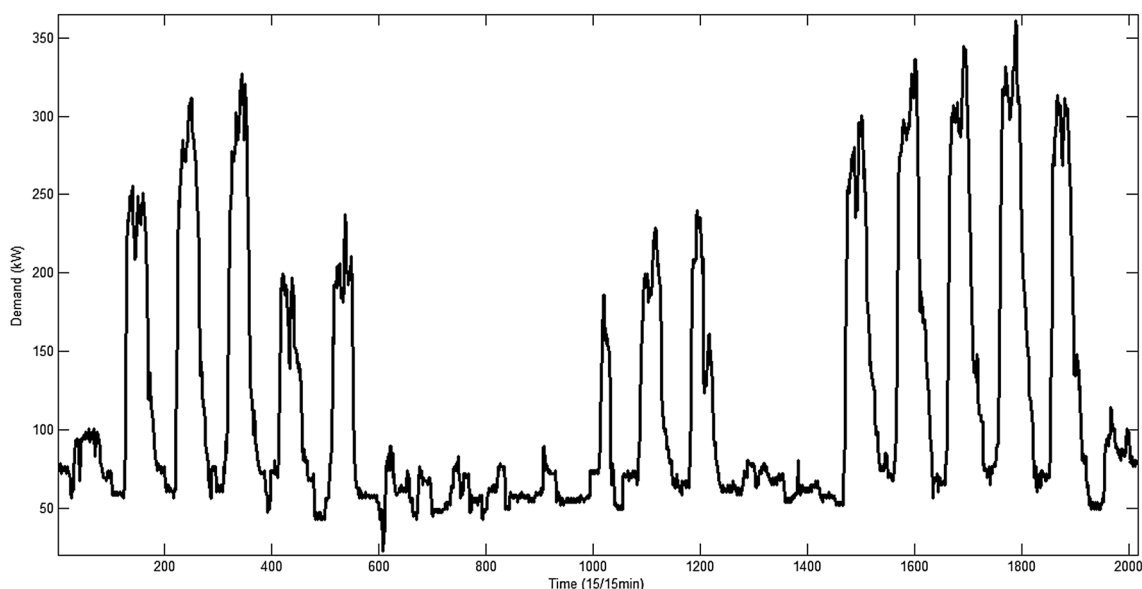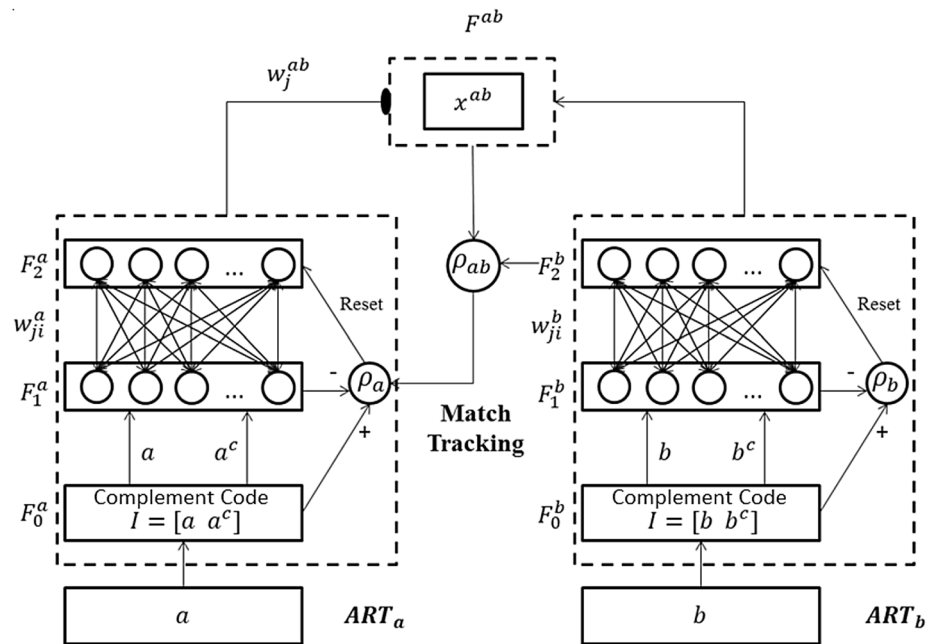- Reading the input (*A*) and output (*B*) patterns for training:

Fig. 1 Load curves prior to, during, and following Carnival

Fig. 2 Fuzzy ARTMAP ANN [14]

$A = [a_1, a_2, \ldots, a_n]$ and $B = [b_1, b_2, \ldots, b_n]$;

- Normalization of input (A) and output (B) vectors:

$$\bar{a} = \frac{a}{|a|} \text{ and } \bar{b} = \frac{b}{|b|}, \text{ where } |a| = \sum_i |a_i| \qquad (1)$$

- Input (A) and output (B) complement:

$$\bar{a}_i^c = 1 - a_i \text{ and } \bar{b}_i^c = 1 - b_i; \qquad (2)$$

- Input (A) and output (B) vectors normalized and complemented:

$$I^e = [a\bar{a}] \text{ and } I^s = [b\bar{b}]; \qquad (3)$$

- Initialization of weight matrices (value 1):

$$w_j^a = 1; w_k^b = 1; w_j^{ab} = 1; \text{ showing that there is no active category.}$$

- Reading the parameters:

$$\alpha, \beta, \rho_a, \rho_b, \rho_{ab} \text{ and } \varepsilon;$$

- *Fuzzy ART$_b$ Module:*
- Calculus of function T_k^b:

$$T_k^b(I^s) = \frac{|I^s \wedge w_k^b|}{\alpha + w_k^b}; \qquad (4)$$

- Choose the category (K) for module *Fuzzy ART$_b$*:

$$T_k^b = \max\{T_k^b : k = 1, \ldots, N_b\}; \qquad (5)$$

- Verify if the vigilance criterion for *Fuzzy ART$_b$* is attended:

$$|x^b| = \frac{|I^s \wedge w_k^b|}{|I^s|} \geq \rho_b; \qquad (6)$$

- If not: Reset: $T_k^b = 0$;
- If yes:
  Resonance, adaptation of *Fuzzy ART$_b$* weights

$$w_K^{new} = \beta(I^s \wedge w_K^{old}) + (1 - \beta)w_K^{old}; \qquad (7)$$

and

Calculus of the activity vector in F_2:

$$y_k^b = [y_1^b, y_2^b, \ldots, y_N^b], \text{ where} : y_k^b = \begin{cases} 1, & \text{if } k = K \\ 0, & \text{if } k \neq K \end{cases} \qquad (8)$$

- Go to the *Match Tracking* module;

- *Fuzzy ART$_a$ Module:*
- Calculus of function T_j^a:

$$T_j^a(I^e) = \frac{|I^e \wedge w_j^a|}{\alpha + w_j^a} \qquad (9)$$

- Choose the category (J) *Fuzzy ART$_a$ Module:*

$$T_j^a = \max\{T_j^a : j = 1, \ldots, N_a\} \qquad (10)$$

- Verify the vigilance criterion of *Fuzzy ART$_a$ Module:*

$$|x^a| = \frac{\left|I^e \wedge w_j^a\right|}{|I^e|} \geq \rho_a \tag{11}$$

- If yes: Calculus of activity vector in F_2: $y_j^a = \left[y_1^a, y_2^a, \ldots, y_N^a\right]$, where:

$$y_j^a = \begin{cases} 1, & \text{if } j = J \\ 0, & \text{if } j \neq J \end{cases} \tag{12}$$

- Go to the *Match Tracking* module;

- *Match Tracking module*:
- Verify if the vigilance criterion of Inter-ART module is attended:

$$\left|x^{ab}\right| = \frac{\left|y^b \wedge w_j^{ab}\right|}{|y^b|} \geq \rho_{ab} \tag{13}$$

- If not: Increment the vigilance parameter:

$$\rho_a = \frac{\left|I^e \wedge w_j^a\right|}{|I^e|} + \varepsilon \tag{14}$$

- Reset: $T_j^a = 0$;
- If *Match Tracking* is attended:
- Resonance, adaptation of *Fuzzy* ART$_a$ weights:

$$w_J^{\text{new}} = \beta\left(I^e \wedge w_J^{\text{old}}\right) + (1 - \beta)w_J^{\text{old}} \tag{15}$$

- Update the Inter-ART module weights:

$$W_{JK}^{ab} = \left[y_1^{ab}, y_2^{ab}, \ldots, y_N^{ab}\right], \text{where :}$$

$$y_{jk}^{ab} = \begin{cases} 1, & \text{if } j = J; k = K \\ 0, & \text{if } j \neq J; k \neq K \end{cases} \tag{16}$$

- Verify if all training pairs were processed;
- If not: $\rho_a = \rho_a^{\text{original}}$ and continue in *Fuzzy* ART$_a$ and *Fuzzy* ART$_b$;
- If yes: end training.
- where α: chosen parameter ($\alpha > 0$); β: training rate [0, 1]; ρ_a:vigilance parameter of ART$_a$ module (0, 1]; ρ_b :vigilance parameter of ART$_b$ module (0, 1]; ρ_{ab}:vigilance parameter of ART$_{ab}$ module (0, 1]; ε: increment of parameter ρ_a; n: quantity of input patterns.

The *Match-Tracking* allows the increase in the vigilance parameter ρ_a (ART$_a$ module) to correct the error in ART$_b$ module. Hence, when a wrong prognostic is performed, generalization is maximized, whereas an error is minimized. The ART$_a$ module begins the search for a correct prognostic, or a new category, that is created for the current input.

The Fuzzy ARTMAP ANN, used in this work, is implemented according to references [41–44]. Other studies, also apply architectures based on ARTMAP artificial neural networks, to predict electricity consumption at more aggregate levels [29, 31]. ARTMAP ANN is also applied to other forecasting tasks [45–47].

4 Singular spectrum analysis: SSA

The singular spectrum analysis is a powerful tool for time series analysis. This technique incorporates the elements of multivariate statistics, classical time series analysis, multivariate geometry, dynamical systems, and signal processing. SSA can decompose the original time series in a small number of components, such as the tendency with low frequency variation, oscillatory components, and white noise.

Hence, it can extract complex and nonlinear tendencies, detect changes in a regime, remove seasonality with variable amplitude and periods, and smoothen time series. The advantages of SSA include the fact that, in general, short univariate time series are sufficient to obtain good results, depends on a small number of parameters, requires only basic knowledge on the series from the analyst, has low computational cost and the denoised series does not present delay in comparison with the original series [48, 49]. Other studies also apply SSA as a data denoising method, for example references [50–52].

The SSA algorithm is divided in two steps: decomposition and reconstruction [48, 53, 54]. Each of these two steps evolves into two other sub-steps. Thus, the decomposition is formed by incorporating and decomposing in single values, whereas the reconstruction includes the clustering and diagonal averaging.

During the sub-step incorporation, let $y_t(t = 1, 2, 3, \ldots, N)$ be an N dimensional time series to be analyzed using SSA. y_t is transformed in a sequence of K L-dimension vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ (L being the size of the window), which are used to form the trajectory matrix X as

$$X = \left[\mathbf{x}_1, \ldots, \mathbf{x}_K\right] = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_K \\ \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \cdots & \mathbf{y}_N \end{pmatrix} \tag{17}$$

This matrix has size L by K, where $K = N - L + 1$. L must be contained in the interval $2 \leq L \leq N - 1$. A method to find the optimal value for the parameter L is still an open discussion in the literature.

The decomposition in single values consists of transforming the trajectory matrix in a sum of matrices with rank 1. Let $S = XX^T$, $\lambda_1, \lambda_2, \ldots, \lambda_L$ be the respective eigenvalues in decreasing order, and $U_1, U_2, \ldots, U_L$ be the eigenvectors associated to $\lambda_1, \lambda_2, \ldots, \lambda_L$. If $V_i = X^T U_i / \sqrt{\lambda_i}$, the trajectory matrix X can be written as $X = X_1 + X_2 + \ldots + X_L$, where $X_i = \lambda_i U_i V_i^T$. The set $\left(\lambda_i, U_i, V_i\right)$ is called the ith eigentriple of matrix X.

The clustering sub-step combines the eigentriples, such that, the group with less variability is adopted as tendency, one or more components are described as oscillatory with variable periods and, finally, the last one, with higher variability, is considered random noise. Thus, the L eigentriples are separated into m disjoint sets. Let $I = \left\{i_1, i_2, \ldots, i_p\right\}$, the eigentriples index group in one of the m desired subsets; therefore, the resulting matrix X_I is determined as the sum of the corresponding matrices X_i, i.e., $X_I = X_{i_1} + X_{i_2} + \cdots + X_{i_p}$.

The diagonal averaging sub-step transforms each of the m matrices X_I in a univariate time subseries of size N, called an SSA component. If $L^* = \min(L, K)$, $K^* = \max(L, K)$, and $y_{I_t}(t = 1, 2, \ldots, N)$ is an SSA component, each element of y_{I_t} is given by:

$$y_{I_t} \begin{cases} \frac{\sum_{i=1}^{t} x_{I_{(i,t-i+1)}}}{t}, & \forall 1 \leq t < L^* \\ \frac{\sum_{i=1}^{L^*} x_{I_{(i,t-i+1)}}}{L^*}, & \forall L^* \leq t \leq K^* \\ \frac{\sum_{i=t-K^*}^{N-K^*} x_{I_{(i,t-i+1)}}}{N-K^*}, & \forall K^* < t \leq N \end{cases} \tag{18}$$

This step consists of calculating the average of the elements X_I, called $x_{I_{(i,j)}}$, where $i + j = t + 1$. In this expression, t equals 1 gives $y_{I_1} = x_{I_{(1,1)}}$, for t equals 2 then $y_{I_2} = \frac{x_{I_{(1,2)}} + x_{I_{(2,1)}}}{2}$, and so on.

5 Methodology

In summary, the methodology used in this paper consists in denoising the time series with SSA and, after that, using the denoised series as training set for the Fuzzy ARTMAP ANN which will then provide the forecast. Herein each step will be detailed.

5.1 SSA decomposition

This step and the next one consist in applying the methodology described in the SSA—singular spectrum analysis section of this paper.

SSA is used to reduce the noise level of the series by ignoring the component classified as noise when reconstructing. The only relevant parameter for this step of the method is the window length, set as two times the data periodicity in this work. According to [22, 48], using window lengths multiple of periodicity attains better separability of the data. In this step, just the data prior to the forecasting horizon was used.

5.2 Hierarchical clustering of SSA components and reconstruction

In this work, the SSA components are hierarchically clustered in three groups using the weighted-correlations matrix as distance criterion. Such groups are interpreted as the trend, the oscillatory components and noise, which are not considered when the smoothed series is reconstructed.

5.3 Fuzzy ARTMAP training and parameter optimization

Following that, the rebuilt series is presented to the Fuzzy ARTMAP ANN for training, as data to be forecasted was not included in the previous step, the training set does not contain the period to be forecasted. The Fuzzy ARTMAP ANN parameters were chosen considering appropriate values [43, 55, 56], the parameters ρ_a, ρ_b, and α were fine tuned by a grid search, and the search is realized between values 0.93 and 0.97, 0.995 and 0.999, and 0.003 and 1.323, respectively, with variations of 0.01 for ρ_a, 0.001 for ρ_b, and 0.66 for α. The parameters β and ρ_{ab} were set to 1 and were not changed. The dimension of the input vectors used for the prediction phase is defined according to the results obtained during the validation phase, where tests were done for 1–7 days. Mean absolute percent error (MAPE) is the criterion for parameter selection and the lower value is chosen.

The figure below illustrates the described method (Fig. 3).

To test the efficiency of the proposal, two well-known methods are used as benchmark: the SARIMA models and the feedforward MLP, with training done using the backpropagation algorithm [40]. The time series is forecasted with and without noise using SSA. The training set, for both the benchmarks and the Fuzzy ARTMAP ANN, was composed by the available data prior to the forecast horizon.

For choosing the SARIMA model order and calculating the coefficients, the automatic method described by [57] was used. This method is implemented by the auto.arima function of the R language forecast package.

The feedforward MLP weight calculations were made utilizing the mlp function of the RSNNS package, also available for the R language. Just one hidden layer with sigmoid activation function was used, as it is sufficient
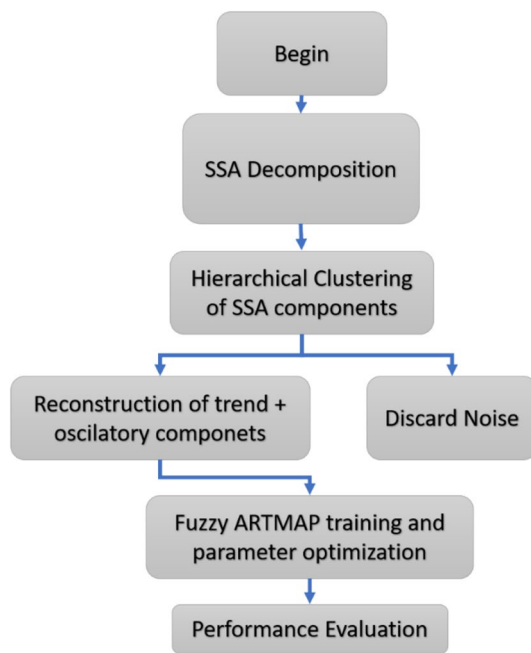
Fig. 3 Fuzzy ARTMAP ANN flowchart

for the majority of the forecasting problems [58]. In the output layer just one neuron with linear activation function was used, the multi-step forecast is made iteratively, presenting the output of the previous step as an input for the next one. The number of neurons in the input and hidden layers was chosen by testing various architectures, and selecting the one that presented the minimal in-sample MAPE. For the input layer, the number of neurons tested was $n = ks$, $n = ks + 2$ or $n = ks - 2$, where n is the number of neurons in the input layer, k an natural number lower than 5, and s the seasonality of the time series. The number of neurons in the hidden layer was a natural number greater than $\frac{n}{2} - 3$ and lower than $\frac{n}{2} + 3$.

The results were also compared with those provided by Fuzzy ARTMAP ANN without using the SSA. The forecast horizons are for 1, 3 and 7 days for all methods.

5.4 Performance evaluation

MAPE is used to evaluate the performance of the forecasts, according to Eq. 19:

$$\text{MAPE} = \frac{1}{n_e} \sum_{t=1}^{n_e} \frac{\left| L_t - F_t \right|}{L_t} \times 100 \tag{19}$$

where L_t: real electrical load at time t; F_t: forecasted electrical load at time t;

n_e: quantity of inputs presented for the diagnosis phase.

In this paper, MAPE was used with the main purpose of measuring the forecast quality from the several methods. Aiming this, all the time series were split into two different sets, one for training and another for testing. The test set was not used for modeling, and was the reference in the forecasts MAPE calculations, being compared with the forecasts.

MAPE is one of the most popular measures of forecast accuracy, is scale-independent and easy to interpret, making it popular in the industrial sector, including the electrical energy companies, where it is a pattern to evaluate the performance of load forecasting [59, 60]. MAPE is recommended in most textbooks [61, 62]. Lower MAPE numbers indicate a more accurate result [63].

MAPE is not indicated when the real values are zero or close to zero. This situation is not verified in the data set used in this paper.

6 Results

Table 1 presents the MAPE scores for the forecasts provided by each method, for the series with 48 daily data using three different predictive horizons, 1, 3 and 7 days, Table 2, in its turn, shows the MAPEs calculated using the same methods and horizons, but for the series with 96 daily data. Bold values represent the best MAPE for each predictive horizon.

The performance of the Fuzzy ARTMAP ANN was improved by noise removal, reducing the MAPE indicator in all scenarios. In addition, it can be perceived that for

Table 1 MAPE of the predictions for load curve with 48 daily data

Model	1 day	3 days	7 days	Average
SARIMA	**0.0831**	0.1925	0.3636	0.2130
Feedforward MLP	0.1550	0.3760	0.5770	0.3693
Fuzzy ARTMAP	0.1167	0.1284	0.4504	0.2318
SARIMA + SSA	0.0883	0.1668	0.3494	0.2015
Feedforward MLP + SSA	0.2430	0.1610	0.4800	0.2946
Fuzzy ARTMAP + SSA	0.1143	**0.1256**	**0.1689**	**0.1921**

Table 2 MAPE of the predictions for load curve with 96 daily data

Model	1 day	3 days	7 days	Average
SARIMA	**0.1090**	0.2260	0.3860	0.2403
Feedforward MLP	0.1410	0.3560	0.5600	0.3523
Fuzzy ARTMAP	0.1147	0.1473	0.2138	0.1586
SARIMA + SSA	0.2250	0.1760	0.3073	0.2360
Feedforward MLP + SSA	0.1890	0.2730	0.6570	0.3730
Fuzzy ARTMAP + SSA	0.1144	**0.1421**	**0.1513**	**0.1397**

the horizons of 3 and 7 days, no method had a superior performance in MAPE terms and, for the 1 day horizon, only SARIMA presented a smaller indicator.

Table 2 shows a similar scenario, being Fuzzy ARTMAP ANN plus SSA combination the method capable of providing the most assertive forecasts for the 3 and 7 days horizons. This demonstrates the power of maintenance of generalization capacity of this kind of ANN, even for larger forecasts horizons.

When comparing the two artificial neural network architectures used in this article, feedforward MLP and Fuzzy ARTMAP, it is not possible to generalize the statement that ANNs behave better in this type of scenario, since among all the models tested, feedforward MLP had some of the worst results, while Fuzzy ARTMAP ANN presented most of the best. Furthermore, it is also not possible to say if statistical or artificial intelligence methods are superior, since SARIMA obtained the best results in some of the scenarios.

Considering all the tested scenarios, ARTMAP Fuzzy ANN proved to be the most consistent predictor, since its results varied very little in all different scenarios, attesting to the robustness of the method. In addition, in various cases, the MAPE of the ARTMAP Fuzzy ANN was the best, showing its capacity in providing assertive results.

Figure 4 shows the performance of the three methods for 7 days in advance, where the load curve contains 96 measures. Considering this scenario, it is important to highlight that only the Fuzzy ARTMAP ANN is able to represent the real load shape, discerning labor days from weekends without any previous knowledge inserted by the researcher. By just extracting information from the data, it has the potential of being very useful when this kind of information is not available for the analyst.

7 Conclusions

The use of the Fuzzy ARTMAP ANN with noise removal by SSA presents the best results when compared to other methods considered as the benchmark in literature, except for the results for a horizon of 1 day.

It is emphasized that the modeling is difficult due to the nature of the data. This occurs according to the aggregation level and the presence of atypical days in the series. The former causes a relative increase in randomness of the model, thereby increasing the amplitude of the noise due to unpredictable decisions. The latter is due to the fact that classical modeling methodologies, very often, do not capture such characteristics and are not prepared to incorporate previous knowledge to the forecasts. In this situation, Fuzzy ARTMAP ANN presents superior performance and, when enhanced with the use of the SSA, provides more generalization.

The use of noise removal with the other methods presents ambiguous results; however, for the load curves formed by 48 data, the results are promising, according to the MAPE for the 3 days and 7 days horizon. Nevertheless, the results for 7 days in advance must be carefully analyzed considering the presence of labor days and weekends in the interval of prediction. SARIMA and feedforward MLP cannot adapt to this situation; therefore, their prediction capacity is reduced, and noise removal does not matter.
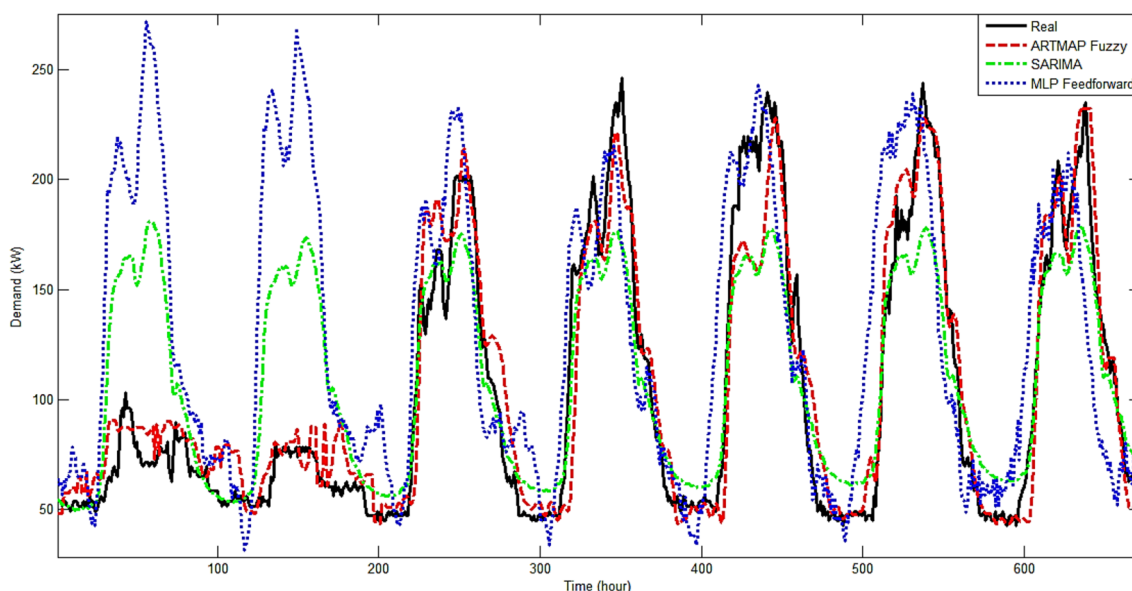

Fig. 4 Seven days in advance forecast (daily curve with 96 measures) without noise removal

The good results obtained by SARIMA for 1 day in advance are due to the behavior correlated with the previous days, consequently providing a good approximation by linear combination, which is characteristic of this model.

8 Future works

In future work, the authors will continue to incorporate data preprocessing as well as including previous information to the forecast methods. One approach is to cluster the data according to the nature of the days, and then use a specific model for each subset.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Morettin PA, Tolói CMC (2006) Time series analysis, 2nd edn. Edgard Blücher, São Paulo
2. Berriel RF, Lopes AT, Rodrigues A, Varejão FM, Oliveira-Santos T (2017) Monthly energy consumption forecast: a deep learning approach. In: 2017 international joint conference on neural networks (IJCNN), Anchorage, AK, pp 4283–4290
3. Bandyopadhyay S, Ganu T, Khadilkar H, Arya V (2015) Individual and aggregate electrical load forecasting. In: The 2015 ACM sixth international conference, pp 121–130
4. Agarwal Y, Weng T, Gupta RK (2011) Understanding the role of buildings in a smart microgrid. In: Design, automation & test in Europe conference & exhibition. Grenoble, pp 1–6
5. Zheng Z, Chen H, Luo XA (2019) Kalman filter-based bottom-up approach for household short-term load forecast. Appl Energy 250:882–894
6. Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y (2019) Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans Smart Grid 10(1):841–851
7. Lusis P, Khalilpour KR, Andrew L, Liebman A (2017) Short-term residential load forecasting: impact of calendar effects and forecast granularity. Appl Energy 205:654–669
8. Kong W, Dong ZY, Hill DJ, Luo F, Xu Y (2018) Short-term residential load forecasting based on resident behaviour learning. IEEE Trans Power Syst 33(1):1087–1088
9. Zeng A, Liu S, Yu Y (2019) Comparative study of data driven methods in building electricity use prediction. Energy Build 194:289–300
10. Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C (2019) Probabilistic individual load forecasting using pinball loss guided LSTM. Appl Energy 235:10–20
11. Xu L, Wang S, Tang R (2019) Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. Appl Energy 237:180–195
12. Kim Yunsun, Son Heung-gu, Kim Sahm (2019) Short term electricity load forecasting for institutional buildings. Energy Rep 5:1270–1280
13. Cretu M, Ceclan A, Czumbil L, Şteţ D, Bârgăuan B, Micu DD (2019) Key performance indicators (KPIs) for the evaluation of the demand response in the Technical University of Cluj-Napoca Buildings. In: 8th international conference on modern power systems (MPS), Cluj Napoca, Romania, pp 1–4
14. Glavan M, Gradišar D, Moscariello S, Juricic D, Vrancic D (2019) Demand-side improvement of short-term load forecasting using a proactive load management a supermarket use case. Energy Build 186:186–194
15. Amber KP, Ahmad R, Aslam MW, Kousar A, Usman M, Khan MS (2018) Intelligent techniques for forecasting electricity consumption of buildings. Energy 157:886–893
16. Ribeiro M, Grolinger K, ElYamany HF, Higashino WA, Capretz MAM (2018) Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. Energy Build 165:352–363
17. Khoshbakht M, Gou Z, Dupre K (2018) Energy use characteristics and benchmarking for higher education buildings. Energy Build 164:61–76
18. Ke X, Jiang A, Lu N (2016) Load profile analysis and short-term building load forecast for a university campus. In: 2016 PES general meeting
19. Deba C, Eanga LS, Yanga J, Santamourisa M (2016) Forecasting diurnal cooling energy load for institutional buildings using artificial neural networks. Energy Build 121:284–297
20. Wang Jue, Li Xiang, Hong Tao, Wang Shouyang (2018) A semi-heterogeneous approach to combining crude oil price forecasts. Inf Sci 460–461:279–292
21. Alonso FJ, Del Castillo JM, Pintado P (2005) Application of singular spectrum analysis to the smoothing of raw kinematic signals. J Biomech 38:1085–1092
22. Golyandina N, Korobeynikov A (2014) Basic singular spectrum analysis and forecasting with R. Comput Stat Data Anal 71:934–954
23. Nose-Filho K, Lotufo ADP, Minussi CR (2011) Preprocessing data for short-term load forecasting with a general regression neural network and a moving average filter. In: PowerTech 2011 IEEE Trondheim, pp 1–7
24. Lisi F, Nicolis O, Sandri M (1995) Combining singular-spectrum analysis and neural networks for time series forecasting. Neural Process Lett 2(4):6–10
25. Li H, Cui L, Guo S (2014) A hybrid short-term power load forecasting model based on the singular spectrum analysis and autoregressive models. Adv Electr Eng 2014:42478. https://doi.org/10.1155/2014/424781
26. Niu M, Sun S, Wu J, Wang J (2016) An innovative integrated model using the singular spectrum analysis and nonlinear multi-layer perceptron network optimized by hybrid intelligent algorithm for short-term load forecasting. Appl Math Model 40(5–6):4079–4093. https://doi.org/10.1016/j.apm.2015.11.030
27. Zhang X, Wang J, Zhang K (2017) Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. Electr Power Syst Res 146:270–285
28. Abreu T, Amorima AJ, Santos-Junior CR, Lotufo ADP, Minussi CR (2018) Multinodal load forecasting for distribution systems using a fuzzy-artmap neural network. Appl Soft Comput 71:307–316
29. Ferreira ABA, Minussi CR, Lotufo ADP, Lopes MLM, Chavarette FR, Abreu TA (2019) Multinodal load forecast using euclidean

ARTMAP Neural network. In: 2019 IEEE PES innovative smart grid technologies conference—Latin America (ISGT Latin America), Gramado, Brazil, pp 1–6

30. Yap KS, Abidin IZ, Lim CP, Shah MS (2006) Short term load forecasting using a hybrid neural network. In: 2006 IEEE international power and energy conference, Putra Jaya, pp 123–128

31. Abreu T, Moreira JR, Minussi CR, Lotufo ADP, Lopes MLM (2019) Short-term multinodal load forecasting using a Fuzzy-ARTMAP neural network. In: 2019 IEEE PES innovative smart grid technologies conference—Latin America (ISGT Latin America), Gramado, Brazil, pp 1–6

32. Antunes JF, de Souza Araújo NV, Minussi CR (2013) Multinodal load forecasting using an ART-ARTMAP-fuzzy neural network and PSO strategy. In: 2013 IEEE Grenoble conference, Grenoble, pp 1–6

33. Arora S, Taylor JW (2018) Rule-based autoregressive moving average models for forecasting load on special days: a case study for France. Eur J Oper Res 226–1:259–268

34. Hippert HS, Bunn DW, Souza RC (2005) Neural networks for short-term load forecasting: a review and evaluation. Int J Forecast 21:425–434

35. Taylor JW (2010) Triple seasonal methods for short-term electricity demand forecasting. Eur J Oper Res 204:139–152

36. Arora S, Taylor JW (2013) Short-term forecasting of anomalous load using rule-based triple seasonal methods. IEEE Trans Power Syst 28(3):3235–3242

37. Cancelo JR, Espasa A, Grafe R (2008) Forecasting from one day to one week ahead for the Spanish system operator. Int J Forecast 24:588–602

38. McCulloch WS, Pitts WH (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133

39. Eberhart RC, Dobbins RW (1990) Neural network PC tools: a practical guide. Academic Press, San Diego

40. Haykin S (2001) Neural networks, principles and practice, 2nd edn. Bookman, New York, p 900

41. Carpenter GA, Grossberg S (1992) A self-organizing neural network for supervised learning, recognition, and prediction. IEEE Commun Mag 30(9):38–49

42. Carpenter GA, Grossberg S, Iizuka K (1992) Comparative performance measures of Fuzzy ARTMAP, learned vector quantization, and back propagation for handwritten character recognition. In: International joint conference on neural networks—IJCNN, vol 1, pp 794–799

43. Carpenter GA, Grossberg S, Markuzon N, Reynolds JH, Rosen DB (1992) *Fuzzy ARTMAP:* "A neural network architecture for incremental supervised learning of analog multidimensional maps". IEEE Trans Neural Netw 3:698–713

44. Lopes MLM, Minussi CR, Lotufo ADP (2005) Electric load forecasting using a fuzzy ART&ARTMAP neural network. Appl Soft Comput 5(2):235–244

45. Alves PB, Lotufo ADP, Lopes MLL, Maciel GF (2020) Impact wave predictions by a Fuzzy ARTMAP neural network. Ocean Eng 202:10716. https://doi.org/10.1016/j.oceaneng.2020.107165

46. Phookronghin K, Srikaew A, Attakitmongcol K, Kumsawat P (2018) 2 level simplified Fuzzy ARTMAP for grape leaf disease system using color imagery and gray level co-occurrence matrix. In: 2018 international electrical engineering Congress (iEECON), Krabi, Thailand, pp 1–4

47. Yang R (2018) Application and analysis of Fuzzy ARTMAP network in computer servo system. In: 2018 3rd international conference on mechanical, control and computer engineering (ICMCCE), Huhhot, pp 627–631. https://doi.org/10.1109/ICMCCE.2018.00138

48. Golyadina N, Nekrutkin V, Zhigliavsky AA (2001) Analysis of time series structure: SSA and related techniques, 1st edn. CRC Press, Boca Raton

49. Elsner J, Tsonis A (1996) Singular spectrum analysis: a new tool in time series analysis, 1st edn. Plenum Press, New York

50. Rodrigues PC, Mahmoudvand R (2016) Correlation analysis in contaminated data by singular spectrum analysis. Qual Reliab Eng Int 32:2127–2137

51. Hassani H, Mahmoudvand R, Yarmohammadi M (2010) Filtering and denoising in the linear regression models. Fluct Noise Lett 9(4):343–358

52. Guo T, Zhang L, Liu Z, Wang J (2020) A combined strategy for wind speed forecasting using data preprocessing and weight coefficients optimization calculation. IEEE Access 8:33039–33059

53. Marques CAF, Ferreira JA, Rocha A, Castanheira JM, Melo-Gonçalves P, Vaz N, Dias JM (2006) Singular spectrum analysis and forecasting of hydrological time series. Phys Chem Earth 31:1172–1179

54. Wu CL, Chau KW (2011) Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. J Hydrol 399:394–409

55. Weenink D (1997) Category ART: a variation on adaptive resonance theory neural networks. In: Institute of phonetic sciences—University of Amsterdam, IFA Proceedings, vol 21, pp 117–129

56. Georgiopoulos M, Fernlund H, Bebis G, Heileman GL (1996) Order of search in fuzzy ART and fuzzy ARTMAP: effect of the choice parameter. Neural Netw 9:1541–1559

57. Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: the forecast package for R. J Stat Softw 27(3):1–22

58. Moreno JJM, Pol AP, Gracia PM (2011) Artificial neural networks applied to forecasting time series. Psicothema 23(2):322–329

59. Sungil K, Heeyoung K (2016) A new metric of absolute percentage error for intermittent demand forecasts. Int J Forecast 32:669–679

60. Byrne RF (2012) Beyond traditional time-series: using demand-sensing to improve forecasts in volatile times. J Bus Forecast 31:13

61. Dobler CP, Anderson-Cook C (2005) Chapter 6, Time series regression. In: Bowerman BL, O'Connell RT, Koehler AB (eds) Forecasting, time series, and regression: an applied approach, 4th edn. The American Statistician, Alexandria, p 278

62. Hanke JE, Reitsch AG, Wichern DW (2001) Business forecasting. Prentice Hall, Boston

63. Hyndmana RJ, Koehlerb AB (2006) Another look at measures of forecast accuracy. Int J Forecast 22(4):679–688