DOI: 10.48011/sbse.v1i1.2459

Data Preprocessing for Load Forecasting using Artificial Neural Network

Rafael S. F. Ferraz * **Renato S. F. Ferraz** **
Lucas F. S. Azeredo ** **Benemar A. de Souza** *

* *Programa de Pós-Graduação em Engenharia Elétrica. Departamento de Engenharia Elétrica. Universidade Federal de Campina Grande, PB, (e-mails:rafael.ferraz@ee.ufcg.edu.br, benemar@dee.ufcg.edu.br).*
** *Programa de Pós-Graduação em Engenharia Elétrica. Departamento de Engenharia Elétrica. Universidade Federal do Espírito Santo, ES, (e-mail:renato.s.ferraz@aluno.ufes.br, lucas.azeredo@aluno.ufes.br).*

Abstract: An accurate demand forecasting is essential for planning the electric dispatch in power system, contributing financially to electricity companies and helping in the security and continuity of electricity supply. In addition, it is evident that the distributed energy resource integration in the electric power system has been increasing recently, mostly from the photovoltaic generation, resulting in a gradual change of the load curve profile. Therefore, the 24 hours ahead prediction of the electrical demand of Campina Grande, Brazil, was realized from artificial neural network with a focus on the data preprocessing. Thus, the time series variations, such as hourly, diary and seasonal, were reduced in order to obtain a better demand prediction. Finally, it was compared the results between the forecasting with the preprocessing application and the prediction without the preprocessing stage. Based on the results, the first methodology presented lower mean absolute percentage error with 7.95% against 10.33% of the second one.

Resumo: Uma previsão precisa da demanda é essencial para o planejamento do despacho econômico no sistema elétrico de potência, contribuindo financeiramente com as companhias elétricas e auxiliando na segurança e continuidade do serviço. Ademais, é evidente que a integração de geração distribuída no sistema elétrico tem aumentado recentemente, principalmente através da inserção de geração fotovoltaica, resultando em uma modificação gradual do perfil de demanda. Portanto, a previsão 24 horas à frente da demanda da cidade de Campina Grande, Brasil, foi realizada utilizando redes neurais artificiais com um destaque para a etapa de pré-processamento dos dados. Com isso, variações nessas séries temporais, como variações horárias, diárias e sazonais, foram reduzidas a fim de obter uma melhor previsão da demanda. Finalmente, comparou-se a previsão realizada com os dados com e sem pré-processamento, sendo obtido um erro absoluto médio percentual de 7,95% e 10,33%, respectivamente.

Keywords: artificial neural network, data preprocessing, load forecasting, stationarity.

Palavras-chaves: estacionariedade, pré-processamento dos dados, previsão de carga, rede neural artificial.

1. INTRODUCTION

The electricity sector is fundamental to the economy of a country, since it is the one that provides electricity to the users, with appropriate quality, to the extent that it is requested (Kagan et al., 2005). The Brazilian electricity sector had small private facilities until 1900, totaling more than 10 MW of installed capacity. At the beginning of the twentieth century, Brazil presented a significant population growth in the same period that the world glimpsed a period of technological innovations in the electric sector. As a result, it was necessary for foreign companies to invest in the country in order to meet the demand in public and private lighting.

After the 1929 crisis, there was an expansion of the State's role in this sector with measures that ensured greater regulation of public services. However, there was an increase in the indebtedness of the electric sector with the uncontrolled increase in inflation in the country in the early 1980s, resulting in a migration from the state monopoly to a market model, in addition to the creation of new control agents for distribution. In that way, there was a deverticalization of the electric energy sectors: generation, transmission, distribution and commercialization, and many distribution companies were privatised (Fiorotti, 2015).

After the unbundling of the electricity sectors, and the consequent search for greater efficiency and profit, the electric energy companies aimed to be successful in the energy dispatch process and in the expansion planning of the system, in order to maintain the operation of the

⋆ This paper was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

electric sector without interruptions. Therefore, prediction techniques with different horizons have been increasingly improved to meet the growing demand. Moreover, this forecasting study has become more challenging after the insertion of distributed energy resource (DER) from intermittent sources in the electric sector.

Many papers have already been developed in the area of electrical demand forecasting for some decades in the past, as observed in Park et al. (1991) and Chow and Leung (1996) which were developed in 1991 and 1996 respectively. It is worth mentioning that Park et al. (1991) carried out the forecast with a horizon of 24 hours, using the artificial neural network (ANN), however, it was limited to the days of the week, thus, excluding holidays and weekends; while Chow and Leung (1996) presented a technique for electric load forecasting based on neural weather compensation. In Brito et al. (2006), the authors developed an ANN for each day of the week, adopting as input parameters, time in binary format, six historical data and the value of the demand of the previous week at the same time, obtaining as a mean absolute error (MAE) 3.70% for the prediction of one step ahead.

Other papers compared the application of ANN and autoregressive integrated of moving average (ARIMA), as in Kandananond (2011) and Dube et al. (2017), and ANN presented better results. In addition, most demand forecasting researches performed a data preprocessing just in order to identify and correct some errors in the database. These errors can be mainly caused by lack of measurements, data repetition and null values as observed in Brito et al. (2006). Other papers simply perform data normalization or separate data on weekdays and weekends as a preprocessing step (Singh et al., 2016, 2017).

This paper intends to forecast the 24th hour ahead of demand data, investing in the preprocessing stage and applying only one ANN for every day of the week. In addition to identifying and correcting erroneous database values as proposed in Brito et al. (2006), this paper seeks to reduce the hourly, daily and annual variations of the time series as well as in Siwek et al. (2011). Thus, it is possible to analyze whether the reduction of the time series deterministic variations results in a minimization of the prediction error. It is noteworthy that the selection of the ANN optimization method is also important to obtain lower errors. For this reason, Levenberg-Marquardt (LM) and resilient back-propagation (RPROP) methods have been widely used for load prediction (Brenna et al., 2017; Turai et al., 2017; Jin et al., 2006). Therefore, this paper compared both methods in order to verify the best prediction performance.

In Sections 2 and 3, the necessary theory of time series and ANN respectively are presented. Sections 4 and 5 show the methodology applied and the results, respectively. Finally, concluding remarks are provided in Section 6.

2. TIME SERIES

Time series is defined as a set of observations about a variable, ordered in time (Morettin and Toloi, 1986). According to Mendenhall and Sincich (1989), the time series can present the following components: trend, cycle, seasonality and a random signal. Therefore, it is worth mentioning that it is necessary to analyze the deterministic variations, from the identification of non-random patterns of data. In case of the demand data, it is known that there is a characteristic curve for the weekdays and another for the weekends and holidays. In addition, there are variations during the year, which may be an increase or a decrease. For example: according to EPE (2017), the northeast region of Brazil showed a 1.6% growth in demand in 2016 compared to 2015. Unlike other time series such as irradiance, where curve models are proposed to eliminate their variations as in Benmouiza and Cheknane (2015), it is difficult to achieve the same methodology for electricity demand database since it presents a growth rate per year. As a way to mitigate these demand curve variations, it is possible to calculate different indexes related to the hours of the day, days of the week, and the days of the year (Siwek et al., 2011).

2.1 Siwek Model

The method adopted by Siwek et al. (2011) consisted in 3 stages. At first, the weekdays variations were reduced, and for this reason, the average demand ($P_m(i)$) for day i of a week was divided by the total demand average, P_m, in order to obtain the coefficient related to each day of the week, as presented in (1).

$$\alpha_{dw}(i) = \left(\frac{P_m(i)}{P_m}\right) \tag{1}$$

For example, the average demand of all Mondays ($i = 2$) was divided by the total demand average. Thus, it was possible to obtain the Monday coefficient, being represented by $\alpha_{dw}(2)$. The same procedure is done for the other days of the week, obtaining another 6 coefficients, $\alpha_{dw}(i)$, where i ranges from 1 to 7, representing each day of the week. Finally, each value $P(i)$ of the database is divided by the coefficient corresponding to its day of the week resulting in a new database (P_1), as shown in (2). For example, all Mondays data would be divided by $\alpha_{dw}(2)$.

$$P_1 = \frac{P(i)}{\alpha_{dw}(i)} \tag{2}$$

Then, the hourly variation was reduced applying a similar methodology. As a result, the hourly coefficients were obtained, $\alpha_h(j)$, where j ranges from 1 to 24. Finally, the coefficients referring to the annual variation were calculated, thus, $\alpha_s(m)$ were obtained, where m were values between 1 and 365 for each day of the year. After the application of the prediction model, the data in the original scale could be obtained from (3), where $P(i, j, m)$ is the demand data with trend and seasonality and $P_{detrend}$ is the demand data after the reduction of the daily and hourly trend and seasonality.

$$P(i, j, m) = P_{detrend}(i, j, m).\alpha_{dw}(i).\alpha_h(j).\alpha_s(m) \tag{3}$$

3. ARTIFICIAL NEURAL NETWORK

The ANNs are inspired by the human nervous system, since both are based on several connections between the elements, which acquire knowledge through experience. In addition, connecting forces between neurons, known as synaptic weights are used to store the acquired knowledge (Haykin, 2007). One of the greatest benefits of the ANN is its ability to generalize, since it is able to produce suitable outputs for inputs that were not present during the training stage. Moreover, ANN also has other advantages such as adaptability, since it has the capacity to adapt the synaptic weights according to modifications of the environment.

An artificial neuron is represented in Figure 1, in which an input signal x_j can be analyzed at the input synapse j which is connected to neuron k that is multiplied by the synaptic weight w_{kj}. Then, there is a linear combiner responsible for summing a bias, b_k, and the input signals, weighted by the respective synaptic strengths of the neuron. Finally, there is the activation function, v_k, that restricts the output amplitude of a neuron, being generally within the interval [0,1] or [-1,1] (Haykin, 2007).

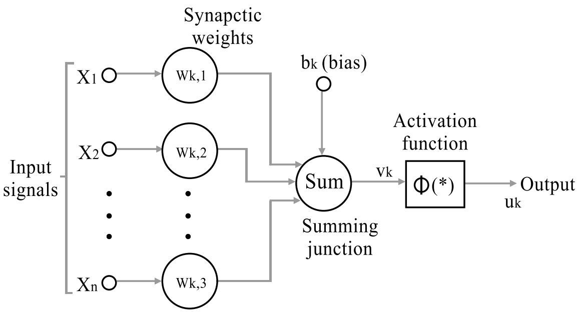

Figure 1. Representation of neuron.

The weighted sum of input signals multiplied by synaptic weights, with the exclusion of bias, can be observed in (4).

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{4}$$

The sum of u_k and the bias b_k is given by the term: induced local field (v_k). Therefore, the output signal (y_k) of neuron k, after activation function ϕ(*), is expressed in (5).

$$y_k = \phi(v_k) \tag{5}$$

The ANN application can be divided into 3 stages: training, validation and test. The first one has the purpose of minimizing a cost function $L[w]$ in relation to the vector of weights w, and this function aims to calculate the error committed by neural network, determining if the model fit was satisfactory. For this weight adjustment, optimization methods should be used, and RPROP and LM are present in many forecasting papers (Brenna et al., 2017; Turai et al., 2017; Jin et al., 2006).

3.1 Resilient Backpropagation

The RPROP algorithm was proposed by Riedmiller and Braun in Riedmiller and Braun (1993) and the main advantage is the elimination of the harmfull impact of the right step of the size of the partial derivatives. Therefore, only the symbols of the derivative $\Delta_{ij}^{(t)}$ are considered for the direction of updating the right as presented in (6), where $\frac{\partial E^{(t)}}{\partial wij}$ is the sum of gradient information for all the patterns at the time t (Pan et al., 2013).

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, if \dfrac{\partial E^{(t)}}{\partial wij} > 0 \\ +\Delta_{ij}^{(t)}, if \dfrac{\partial E^{(t)}}{\partial wij} < 0 \\ 0, else \end{cases} \tag{6}$$

As result, the algorithm has fast speed and reaches the best convergence without selecting parameters.

3.2 Levenberg Marquardt

The LM algorithm is basically inspired from both the gradient descent and the Gauss-Newton method. The gradient descent is good for optimizing problems however, it suffers from speeds whereas the Gauss-Newton method has relatively good speed and avoids the calculations of second derivate but may fall into local minimum (Awan et al., 2018).

The weight update by LM algorithm is given by adding Δw to previous weights and calculated from (7), where J is Jacobian matrix, i.e., first derivative of training error vector and μ is combination coefficient (Singh et al., 2016).

$$\Delta w = \left(J^T J + \mu I\right)^{-1} J^T e \tag{7}$$

4. METHODOLOGY

The methodology applied in this paper can be divided into 3 stages, as observed in Fig. 2. The first one consists of the data preprocessing in which the model proposed by Siwek et al. (2011) is applied, reducing the deterministic variations of the demand series, and then the data are normalized.

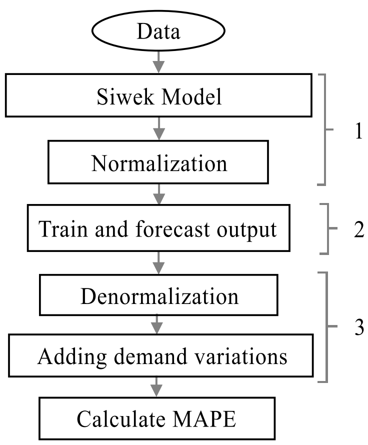

Figure 2. Methodology flowchart.

The second stage is based on the data processing, in which the ANN training is performed, comparing two optimization methods: RPROP and LM, followed by the forecast 24 hours ahead of demand. At last, there is the post-processing that consists of denormalizing the output data of the network, and the deterministic variations of demand curve are added to the denormalized data from (3). Finally, the mean absolute percentage error (MAPE) of the load forecast was calculated. This error corresponds to the metric that offers more information in relation to the predictive quality (Mendenhall and Sincich, 1989). Thus, it is possible do compare the efficiency of both optimization methods.

In addition, the results were compared with the prediction performed applying the same methodology but using the data without reducing the time series variations. As result, it is possible to verify the relevance of the application of the stage of database preprocessing.

5. RESULTS AND DISCUSSIONS

For this paper, it was used a database with sampling interval of one hour. These data are from the city of Campina Grande, Paraíba and were provided by Energisa. For the training of the ANN, data referring to the years 2014, 2015 and 2016 were used. And, 2017 data was used in order to test the ANN. Table 1 shows the mean, minimum and maximum values, and the standard deviation of training and test data. The average of the test data is higher than the training data due to the annual growth rate of demand which was similar to Northeast region (EPE, 2017).

Table 1. Training and Testing Parameters.

	Training (kW)	**Test (kW)**
Mean	2.6246	2.7274
Minimun	1.2461	1.2940
Maximum	5.3256	5.4319
Standard deviation	1.0314	1.0702

Fig. 3 shows the demand data for 2017, where it is possible to observe the variation during the day and the annual variation. In the middle of the year, average monthly demand values are lower than the other months of October to March.

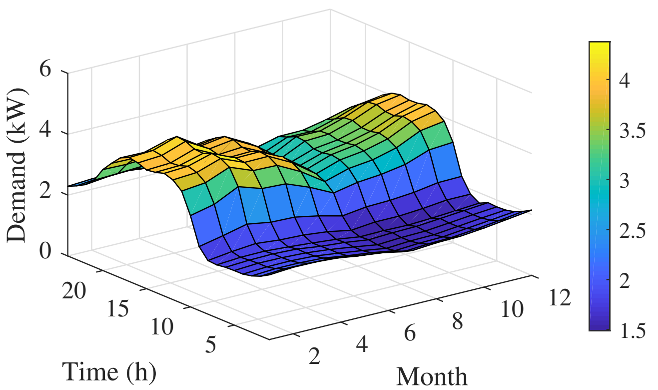

Figure 3. Demand curves during a year.

First, ADF test was applied to database and the result was the alternative hypothesis which means that the time series is stationarity or trend-stationarity; in other words, the data does not present stochastic variations. However, it is known that the demand time series presents deterministic variations, thus, it was applied the methodology proposed in Siwek et al. (2011) as explained in Section 2.

The first step consisted in calculating the indexes for the days of the week. It is possible to notice in Fig. 4 that the indexes 2 to 6, referring to the weekdays (Monday to Friday), presented close values, between 1.0624 and 1.0830; whereas the indexes referring to the weekend (Saturday and Sunday), presented values lower than 1, since the demand data at these days are lower than the other days. Therefore, when applying (2), the weekend demand values for the new database will be higher than the original one. After calculating the new demand values, two more indexes were calculated, one refers to the holidays on weekdays and the other is related to holidays on Saturdays. As a result, 0.6883 and 0.7496 were obtained respectively, and the new demand database was again calculated, similar to (2).

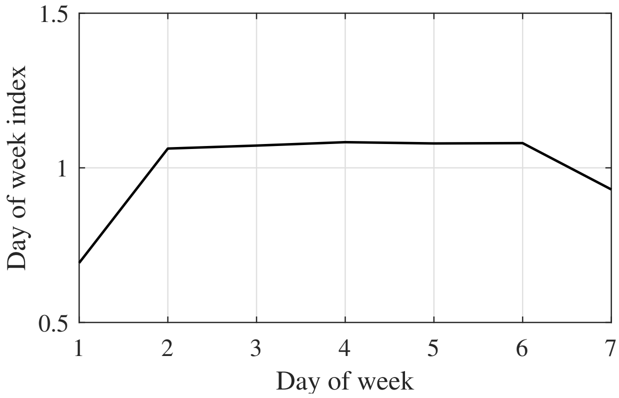

Figure 4. Days of week indexes.

The same previous procedure was applied to determine indexes obtained for the hours of the day and for the days of the year, and the results can be observed in Figs. 5 and 6. In Fig. 6, it can be observed that in the middle of the year, the indexes present values are lower than 1, making it possible to eliminate the annual variation.

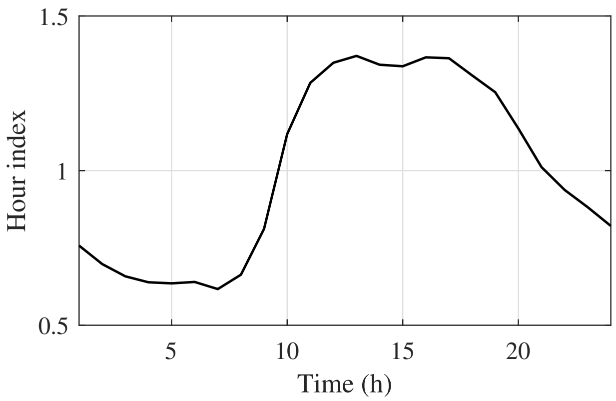

Figure 5. Hours indexes.

After the reduction of the variations of the demand, the difference between the original database and the data after the application of the preprocessing stage can be graphically observed in Fig. 7. In addition, it can be seen that the preprocessed demand data varies around an average value, which is approximately 2.8 kW.

In addition, the measure of variability of each step, n, of the methodology applied by Siwek et al. (2011) was

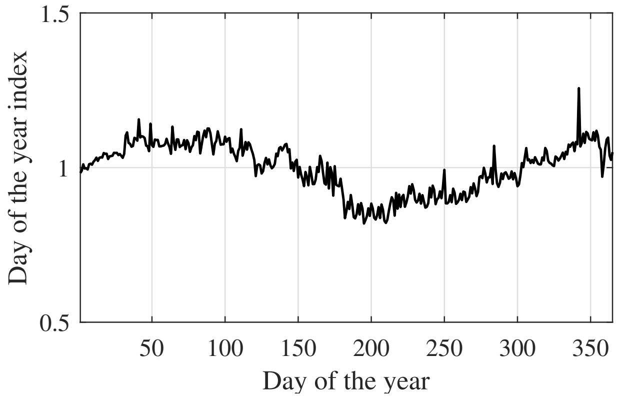

Figure 6. Days of the year indexes.

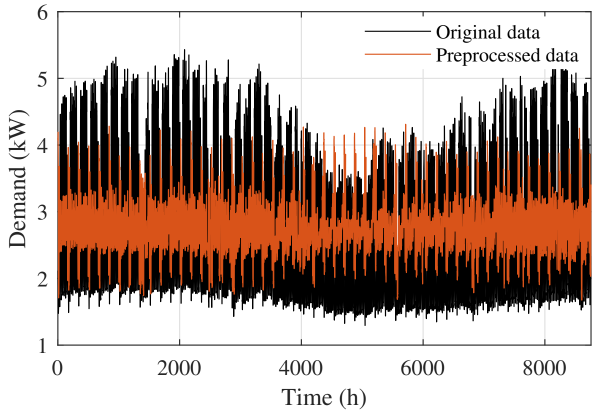

Figure 7. Database before and after the preprocessing stage.

calculated, from the division of the standard deviation of the data by the mean value of the data. In Table 2, it can be observed that there was a reduction in the measure of variability for each preprocessing step sequentially: the variation of the days of the week (k_1), the variation in function of the holidays in the weekdays and on Saturday (k_2), the hourly (k_3) and the year (k_4) variation.

Table 2. Measure of variability.

Stages	Variability
Initial data	0.3930
k_1	0.3498
k_2	0.3410
k_3	0.1719
k_4	0.1495

After data preprocessing, ANN input parameters were selected from historical data. Many prediction papers selected the input based on the significance of the data from the autocorrelation function (ACF) (Flores et al., 2012; Sfetsos and Coonick, 2000). Fig. 8 presents the ACF of the database without the preprocessing stage in black on the chart and the ACF of the database with preprocessing application in orange on the chart.

It can be noticed that the original data ACF has many historical data with high significance, thus it is more difficult to select the quantity of lags as ANN input, whereas in orange on the chart (Fig. 8), it can be seen that the most recent historical data present a high autocorrelation, and historical data between 165 and 175 also show a high autocorrelation. As result, the delay intervals of demand selected as input of the network were: 1-12 and 166-170. Moreover, data 24 hours ahead of the input were selected as the network output for the training stage.

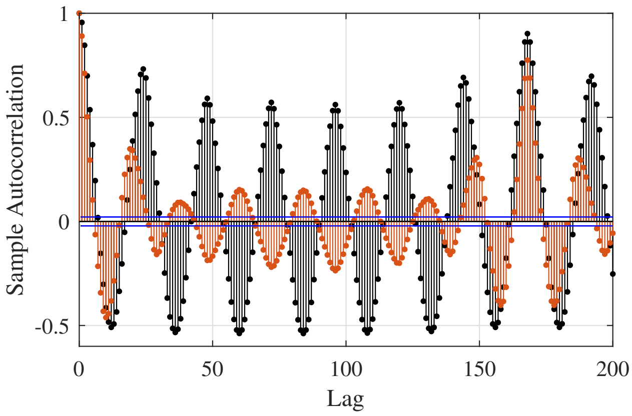

Figure 8. Demand ACFs.

Moreover, the number of neurons in the hidden layer was also determined , and from tests, 25 neurons were chosen. In addition, the activation function used in the hidden layer was the logarithmic of sigmoid and in the output layer, the linear function was applied. After the modeling of the ANN, two predictions of demand were made, in which the RPROP optimization method was used for the first one, and for the second, LM was applied. Finally, the same procedure was performed, however at this time, the input and output of the ANN training were the original database (without applying preprocessing). It is important to highlight that the same amount of historical data was selected for the ANN input, since the original data ACF does not contribute to its selection. In addition, determining the same number of parameters allows checking whether the ANN can recognize the time series patterns for the original data.

In Table 3, it can be observed that the application of the preprocessing of the data contributed to obtain a lower MAPE, and it presented more significant reduction than in Siwek et al. (2011), since it was possible to notice a higher minimization in the measure variability of the database. It is essential to affirm that the advantages of the preprocessing are: the reduction in the effort of the ANN to search for the time series patterns, since the deterministic variations were reduced and consequently, fewer historical data are needed in the input; besides the ease of selection of the input parameters based on the ACF. In addition, it is worth emphasizing the importance of the determination of the ANN optimization method. LM presented better result than the RPROP as well as in Maxwell (2014), which concluded that LM converges faster and give more accurate results for a prediction application.

CONCLUSION

In this paper, it was analyzed the importance of the data preprocessing application and the comparison between two optimization methods for the ANN training in order to

Table 3. MAPE for different methodologies.

	Optimization methods	MAPE (%)
with pre-processing	Resilient backpropagation	10.33
	Levenberg-Marquardt	7.95
without pre-processing	Resilient backpropagation	13.12
	Levenberg-Marquardt	10.62

reduce the MAPE. From the results, it was observed that the reduction of the deterministic variations of the demand curve promoted a decrease of the MAPE considering the application of the same optimization method. In addition, a good selection of the optimization method is also essential to obtain a reduced MAPE. It is evident that the LM method presented lower MAPE for both predictions: with the application of the preprocessing and without the application of the preprocessing. Therefore, allying this optimization method with the reduction of the variations of the demand curve which were inserted in the neural network as input for the training stage can guarantee good prediction results for 24 hours ahead. For the data of Campina Grande, it was observed that when applying the LM and reducing the variations of the series of demand, an MAPE of 7.95% was obtained. This result was approximately 3% lower than the forecast without applying the preprocessing. In this regard, this study contributes to the electricity companies, since it becomes easier to precisely plan the electric dispatch in the power system with an accurate prediction.

REFERENCES

Awan, S.M., Khan, Z.A., and Aslam, M. (2018). Solar generation forecasting by recurrent neural networks optimized by levenberg-marquardt algorithm. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 276–281. doi:10.1109/IECON.2018.8591799.

Benmouiza, K. and Cheknane, A. (2015). Small-scale solar radiation forecasting using arma and nonlinear autoregressive neural network models. *Theoretical and Applied Climatology*, 124. doi:10.1007/s00704-015-1469-z.

Brenna, M., Foiadelli, F., Longo, M., and Zaninelli, D. (2017). Solar radiation and load power consumption forecasting using neural network. 726–731. doi:10.1109/ICCEP.2017.8004771.

Brito, N.S.D., Souza, B.A., Melo, A.C.M., Lima, R.B., and Pamplona, F.M.P. (2006). Aplicação de redes neurais artificiais para previsão de carga em sistemas de distribuição. 1–6.

Chow, T.W.S. and Leung, C.T. (1996). Neural network based short-term load forecasting using weather compensation. *IEEE Transactions on Power Systems*, 11(4), 1736–1742. doi:10.1109/59.544636.

Dube, M., Awodele, K., Olayiwola, O., and Akpeji, K. (2017). Short term load forecasting using arima, ann and hybrid ann-dwt. In *Procedia Computer Science*, volume 48.

EPE, E.D.P.E. (2017). Projeção da demanda de energia elétrica para os próximos 10 anos (2017-2026).

Fiorotti, R. (2015). *Metodologia para Determinar a Potência Firme das Unidades de Geração Distribuída e sua Aplicação no Processo de Previsão de Demanda das Redes de Distribuição de Energia Elétrica.* Master's thesis.

Flores, J.H.F., Engel, P.M., and Pinto, R.C. (2012). Autocorrelation and partial autocorrelation functions to improve neural networks models on univariate time series forecasting. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 1–8. doi:10.1109/IJCNN.2012.6252470.

Haykin, S. (2007). *Redes Neurais: Princípios e Prática.* Artmed.

Jin, L., Fengand, Y., and Jilai, Y. (2006). Peak load forecasting using hierarchical clustering and rprop neural network. 1535 – 1540. doi:10.1109/PSCE.2006.296528.

Kagan, N., de Oliveira, C.C.B., and Robba, E.J. (2005). *Introdução aos sistemas de distribuição de energia elétrica.* Edgard Blucher.

Kandananond, K. (2011). Forecasting electricity demand in thailand with an artificial neural network approach. *Energies*, 4, 1246–1257. doi:10.3390/en4081246.

Maxwell, C.O. (2014). *Prediction of River Discharge Using Neural Networks.* Master's thesis, University of Nairobi, Nairobi.

Mendenhall, W. and Sincich, T. (1989). *A second course in business statistics: regression analysis.* Dellen Publishing Company.

Morettin, P.A. and Toloi, C.M.C. (1986). *Series temporais.* Atual.

Pan, X., Lee, B., and Zhang, C. (2013). A comparison of neural network backpropagation algorithms for electricity load forecasting. In *2013 IEEE International Workshop on Inteligent Energy Systems (IWIES)*, 22–27. doi:10.1109/IWIES.2013.6698556.

Park, D.C., El-Sharkawi, M.A., Marks, R.J., Atlas, L.E., and Damborg, M.J. (1991). Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, 6(2), 442–449. doi:10.1109/59.76685.

Riedmiller, M.A. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning: the rprop algorithm. *IEEE International Conference on Neural Networks*, 586–591 vol.1.

Sfetsos, A. and Coonick, A. (2000). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, 68(2), 169 – 178.

Singh, N.K., Singh, A.K., and Paliwal, N. (2016). Neural network based short-term electricity demand forecast for australian states. In *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, 1–4. doi:10.1109/ICPEICES.2016.7853603.

Singh, S., Hussain, S., and Bazaz, M.A. (2017). Short term load forecasting using artificial neural network. In *2017 Fourth International Conference on Image Information Processing (ICIIP)*, 1–5. doi:10.1109/ICIIP.2017.8313703.

Siwek, K., Osowski, S., and Swiderski, B. (2011). Trend elimination of time series of 24-hour load demand in the power system and its application in power forecasting. *Electrical Review*, 3, 249–253. doi:10.13140/2.1.3900.8962.

Turai, T., Ballard, I., and Rob, R. (2017). Short-term electrical load demand forecasting using artificial neural networks for off-grid distributed generation applications. In *2017 Saudi Arabia Smart Grid (SASG)*, 1–7. doi:10.1109/SASG.2017.8356472.