The final publication is available at

https://doi.org/10.1007/978-3-030-32033-1_16

Additional Information

Forecasting Building Electric Consumption Patterns through Statistical Methods

Xavier Serrano-Guerrero[1], Luis-Fernando Siavichay[1], Jean-Michel Clairand[2], and Guillermo Escrivá-Escrivá[3]

[1] Universidad Politécnica Salesiana, Grupo de Investigación en Energías, 010105, Cuenca, Ecuador
jserranog@ups.edu.ec,
lsiavichay@est.ups.edu.ec,
[2] Universidad de las Américas, Facultad de Ingenería y Ciencias Agropecuarias, 170122, Quito, Ecuador
jean.clairand@udla.edu.ec,
[3] Universitat Politècnica de València, Institute for Energy Engineering, 46022, Valencia, Spain
guieses@die.upv.es

(Preprinted Version. This is not for mass dissemination.)

Abstract. The electricity sector presents new challenges in the operation and planning of power systems, such as the forecast of power demand. This paper proposes a comprehensive approach for evaluating statistical methods and techniques of electric demand forecast. The proposed approach is based on smoothing methods, simple and multiple regressions, and ARIMA models, applied to two real university buildings from Ecuador and Spain. The results are analyzed by statistical metrics to assess their predictive capacity, and they indicate that the Holt-Winter and ARIMA methods have the best performance to forecast the electricity demand (ED).

Keywords: ARIMA, Electric demand, Forecast, Load Uncertainties, Statistical methods, Winter

1 Introduction

The electricity is an essential element for the development of the current world, it promotes the progress of societies and raises the standard of living of people. However, the increasing population and new generation solutions, such as renewable energy, create new challenges for the operation, planning and energy management of power systems [10,19]. Thus, it is essential to have proper tools to forecast electricity demand (ED), considering its significant uncertainties, which have widely been studied for several years [8].

There are three kinds of load forecasting depending on the time scale: short-term, medium-term, and long-term. The short-term carries out a forecast of the ED many hours or days ahead. The medium-term carries out a forecast of various weeks and months ahead, and it is of particular interest for negotiating energy

contracts. The long-term forecasting corresponds for a study of several years ahead [7]. Short-term forecasting allows managing generation, distribution, and transmission operation efficiently [18, 20].

Researchers in this area have developed several techniques. For example, fuzzy procedures have been performed for short-term [9], and medium-term forecasting [6]. In [3], Echo State Networks and Principal Component Analysis were used for a day horizon. Another technique significantly used are neural networks. For example, the authors of [5] studied a neural-network-based model for the short-term load forecast of the distribution grid. In [17], a method based on backpropagation neural networks and election of important variables as inputs were studied. The authors of [15] proposed an ensemble forecast framework, combining three neural network predictors.

In [14], advanced metering infrastructure was used to improve the load forecasting, based on clustering techniques of a group of customers.

Most of these techniques suffer from a lack of an efficient feature selection technique. To address this, the authors of [1] studied the load and price forecast based on new feature selection techniques.

Fewer works have investigated probabilistic load forecasting techniques. For example, in [11] the load forecasting problem includes quantile regression averaging on a set of sister point forecasts.

Although these works and others have studied the forecast of demand in distribution systems, fewer works have explored the demand forecast in buildings, which could allow managing efficiently the energy. For example, in [16] a statistical methodology to assess changes in the electric demand of buildings was proposed. The authors of [2] compared various techniques to forecast electricity consumption of buildings. The aim of this paper is to assess probabilistic forecasting methods in university buildings.

The rest of the paper is organized as follows: Section II describes an overview of various probabilistic forecasting techniques. Section III presents the Case Study and the Overall procedure. The results are discussed in Section IV. Finally, Section V highlights the main conclusions.

2   Background

This section describes probabilistic forecasting techniques.

2.1   Simple Linear Regression (SLR)

The SLR models the ED from an independent variable. The information of this independent variable is used for performing a forecast of the ED, based on the following mathematical expression:

$$\widehat{y} = \beta_0 + \beta_1.x + e \qquad (1)$$

where $\widehat{y}$ is the forecast value of the ED, β_0 the intercept, β_1 the slope, x the independent variable, and e the error.

2.2   Multiple Linear Regression (MLR)

The MLR is an extension regression of the SLR. The forecast is obtained based on the information of two or several k independent variables:

$$\hat{y} = \beta_0 + \beta_1.x_1 + ... + \beta_k.x_k + e \tag{2}$$

To determine the best relation between the independent variables, the least-squares approach has to be implemented, which minimizes the sum of squared residuals.

2.3   Time Series

Time series models consist of data recorded in an orderly manner over time. They can be stationary or non-stationary. The time series is stationary if the mean and variance are constant. The trend, seasonal variation, cyclical and irregular variation, are factors that have time series. The trend shows a uniform behavior that grows or decreases within time. Seasonal variation exists when the data series shows a pattern that varies in a similar way year after year. The cyclic variation is described as the fluctuation of the time series data in defined periods. Finally, the random variation presents fortuitous changes in the time series.

2.4   Exponential smoothing (ES)

Exponential smoothing (ES) methods involve standard procedures for continuously revising a forecast in light of more actual information corresponding to the estimated data. In brief terms, the methods relegate exponentially decreasing weights as the observation gets older [12]. Various approaches of exponential smoothing exists, of which the Holt and Winter methods are detailed:

Holt Method (HM) The HM appropriates for series including a linear time trend and additive seasonal variation. The series are expressed as follows:

$$A_t = \alpha.Y_t + (1 - \alpha).(A_{t-1} + T_{t-1}) \tag{3}$$

The trend estimation is defined by:

$$T_t = \beta.(A_t - A_{t-1}) + (1 + \beta).T_{t-1} \tag{4}$$

The forecast is computed by:

$$\hat{Y}_{t+p} = A_t + p.T_t \tag{5}$$

where $\hat{Y}$ is the new series value.

Winter Method (WM) The WM includes a linear trend and multiplicative seasonal variation. The smoothed series are defined:

$$\hat{Y} = (A_t + p.T_t).S_{t-L+p} + e_t \tag{6}$$

where S is the seasonal factor, the coefficients are defied by the recursions:

$$A_t = \alpha.\frac{Y_t}{S_{t-L}} + (1 - \alpha).(A_{t-1} + T_{t-1}) \tag{7}$$

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta).T_{t-1} \tag{8}$$

$$S_t = \gamma.\frac{Y_t}{A_t} + (1 - \gamma).S_{t-L} \tag{9}$$

2.5   ARIMA Models

The ARIMA Models (p,d,q) are a class of stochastic processes used to analyze time series [4]. They include autoregressive processes, which are univariable models whose value in time depends on its data in a previous time series and on a random term. The time series data are used for searching and prediction [13]. The obtained forecasts are aggregated for each bootstrapped time series to generate the final output Y_t [12]. It is defined as follows:

$$Y_t = \mu + \beta_0.u_t + \beta_1.u_{t-1} + \beta_2.u_{t-2} + ... + \beta_q.u_{t-q} \tag{10}$$

where μ is a constant term, β_q are the means , and u_t the static errors.

3   Case Study and Overall Procedure

3.1   Case Study

The ED data corresponds to two university buildings: the Edificio Cornelio Merchán (ECM) of the Universidad Politécnica Salesiana (UPS), Cuenca, Ecuador, and the 8E building (8EB) of the Universitat Politècnica de València (UPV), Spain. The ED data was adjusted in order to be hourly. Besides, the analysis of variance (ANOVA) allowed classifying the ED according to the similarity of the consumption, as well as relating them with working and non-working days. The data of the ED from the ECM corresponds from March 14th 2017 to December 8th 2017. The data of the ED from the 8EB corresponds from July 1st 2014 to November 27th 2016.

3.2   Evaluation Measures

To evaluate the effectiveness of each method, various evaluation measures are used:

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i| \tag{11}$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = 100.\frac{1}{n} \sum_{i=1}^{n} |\frac{y_i - \widehat{y}_i}{y_i}| \tag{12}$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2} \tag{13}$$

3.3   Overall Procedure

The studied ARIMA model is estimated based on the methodology of Box-Jenkins. Firstly, the Autocorrelation Function (ACF) is obtained to identify the possible parameters: p, d, and q. The procedure used includes four steps: identification, estimation, validation, and evaluation.

Identification Due to the behavior of the time series, an integration has been applied. Fig. 1 shows the ACF and the partial autocorrelation function (PACF). The results of the functions shown in the figures already have a differentiation to guarantee seasonality in the data. The first ACF presents a decrease toward zero but is not erased, approaching the behavior of first-order autoregressive (AR) model. While, when reviewing the PACF in Fig. 1 (b), it can be observed that the behavior is similar to a moving average (MA) process.

Estimation The model is approximated to the theoretical patterns of the ACFs, which are shown in Table 1. First, order parameters are estimated for the ECM model of the UPS, that is, an AR (1), I (1) and MA (1). Due to the initial integration, the final model is an ARIMA.

Validation Various considerations are taken to select the best alternative. The ARIMA model has to reflect the most effective results in RMSE, Bayesian information criteria (BIC), and having an $R^2 > 0.8$, with a p>0.05. Table 1 and 2 show the ARIMA model parameters for the ECM, in working and in a non-working day respectively. Based on Table 1, the ARIMA (100) (011) is the most suitable and is used for the forecast of the ED in working days. Based on Table 2, the ARIMA is the ideal for non-working days.
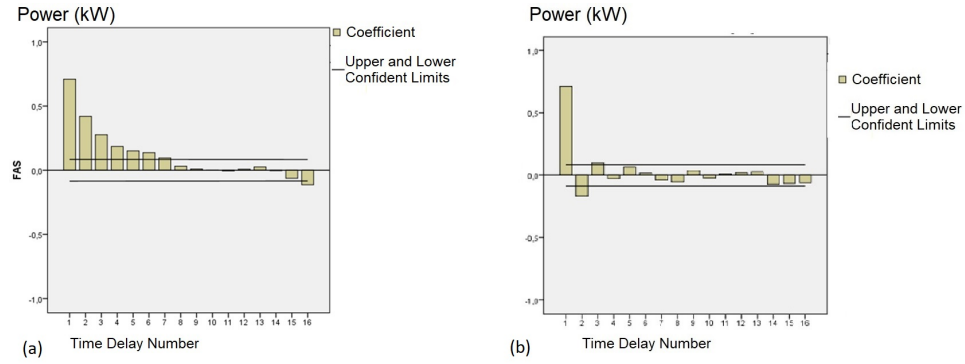
Fig. 1. (a) Autocorrelation Function; (b)Partial Autocorrelation Function.

Table 1. Arima Model Parameters for the ECM in a working day

Model	Param.	Param.	RMSE	R^2	p	CIB
(100)(111)	AR	SARMA	1,227	0,974	0	0,437
(100)(011)	AR	SMA	1,226	0,974	0	0,43
(001)(110)	MA	SAR	1,344	0,968	0	0,613
(101)(101)	ARMA	SARMA	1,33	0,969	0	0,606
(102)(111)	ARMA	SARMA	1,235	0,970	0.001	0,433

Table 2. Arima Model Parameters for the ECM in a non-working day

Model	Param.	Param.	RMSE	R^2	p	CIB
(100)(011)	AR	SMA	0,771	0,796	0,0414	-0,456
(101)(111)	ARMA	SARMA	0,764	0,808	0,056	-0,432
(101)(011)	ARMA	SMA	0,767	0,799	0,102	-0,446
(001)(111)	MA	SAR	0,954	0,688	0	-0,31
(001)(011)	AR	SARMA	0,769	0,798	0,02	-0,441

Evaluation To evaluate the ARIMA model, several predictions were performed in the SPSS software, for working and non-working days. The ARIMA model (100)(011) with R^2=0.974 and RMSE=1.266 kW is used, and its result is depicted in Fig. 2 (a). For the prediction of non-working days, the ARIMA (101)(111) is performed, with a RMSE=0.764 kW and R^2=0.808, and the forecast is illustrated in Fig. 2 (b).

Concerning the 8EB building, the ACF shows a similar behavior than the ECM. The ACF defines an AR. As a result, the ARIMA model (101) (111) is selected for predicting both working and non-working days. In Fig. 3 the
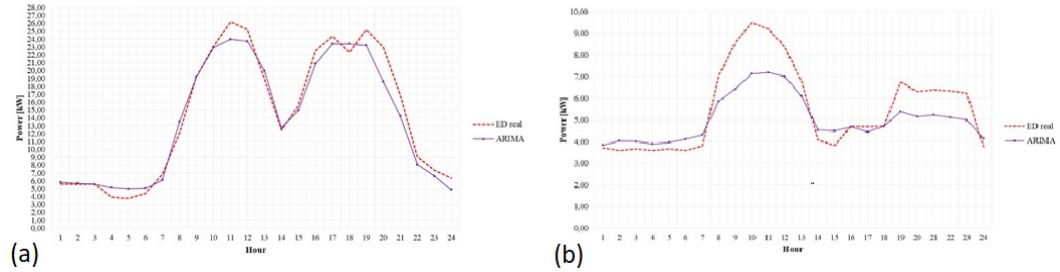
Fig. 2. ARIMA forecast of ECM: (a) in a working day; (b) in a non-working day.
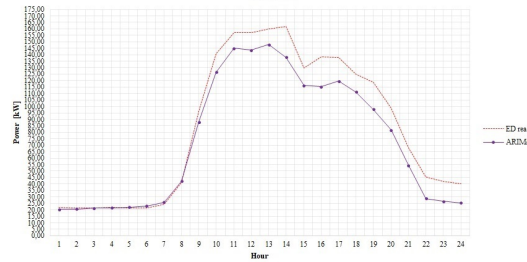

Fig. 3. ARIMA forecast of 8EB Building in a non-working day.

prediction of a working day with the chosen model is shown, the coefficient of determination is R^2=0.985, and RMSE=6.686 kW.

4   Results and Discussion

4.1   Results Comparison

To assess the different forecasting methods, their results are compared. The methods used are SLR, MLR, MHW, and ARIMA. The ES and MH were not considered since its forecasting capacity is weak and the errors are high enough. Fig. 6 illustrates the forecasting patterns of the different methods of a working day of the two university buildings studied. The forecast errors of the ECM and the 8EB are shown in Table 3.

The results indicate that the ARIMA method and the HW have better forecast capacity. Although SLR and MLR have a low prediction error, these models have difficulty following the shape of the electrical source profile.

Then, to assess the forecast capacity of the ARIMA and HW methods, a non-working and a working day were also simulated. Fig. 5 depicts the ED forecast of a non-working day in the two studied university buildings for the corresponding methods. Table 4 shows the forecast errors of ARIMA and HW of a non-working
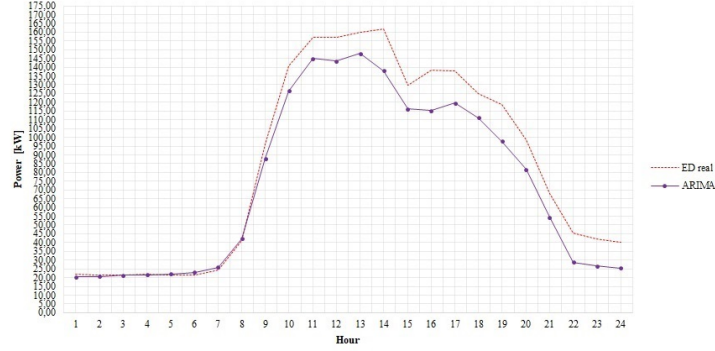
Fig. 4. ED Forecast of a working day: (a) In the ECM; (b) In the 8EB.

Table 3. Forecast Errors of a working day in the ECM

	Holt-Winter	SLR	MLR	ARIMA
MAE (ECM) [kW]	11.77	3.6493	6.47	0.9001
MAPE (ECM) [%]	78.79	24.1601	65.19	7.2902
RMSE (ECM) [kW]	14.10	5.3338	7.71	1.1763
MAE (8EB) [kW]	35.21	30.58	46.8818	8.3135
MAPE (8EB) [%]	40.27	58.49	75.0664	12.8791
RMSE (8EB) [kW]	44.38	38.18	51.7378	10.2593

Table 4. Forecast Errors of ARIMA and HW of a non-working day in the ECM

	Holt-Winter	ARIMA
MAE (ECM) [kW]	13.1100	0.8469
MAPE (ECM) [%]	239.2056	13.6818
RMSE (ECM) [kW]	14.6200	1.0692
MAE (8EB) [kW]	13.7052	1.1036
MAPE (8EB) [%]	63.6774	5.1225
RMSE (8EB) [kW]	15.4199	1.2538

day in the ECM. It is observed that the HW does not present proper results and the ARIMA is significantly more accurate comparing to the real demand.

The ED forecast of ARIMA and HW of a working day is depicted in Fig. 6, in the two university buildings. The forecast errors are shown in Table 5. Note also that the ARIMA model presents the most accurate results, and the HW is not proper to forecast the ED of the university buildings.
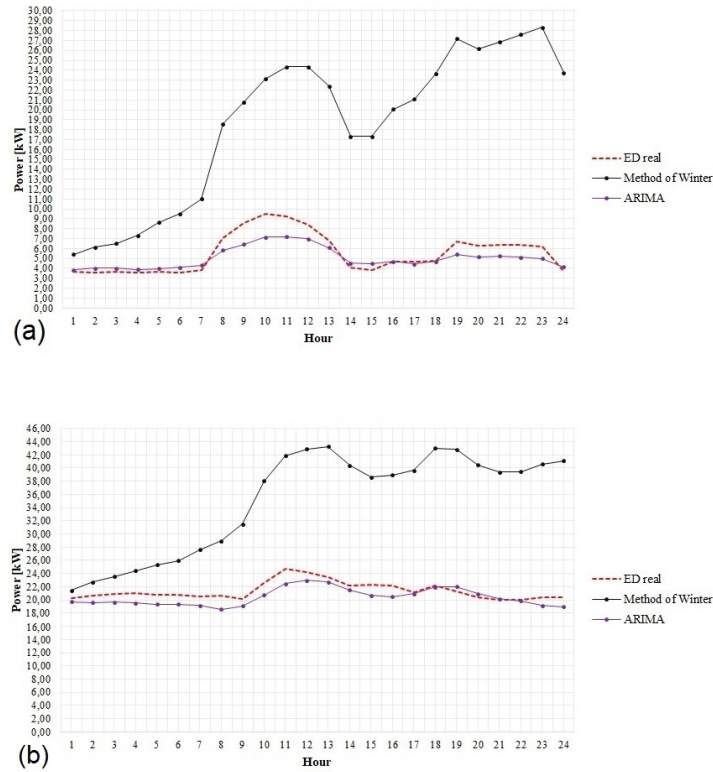
Fig. 5. ED Forecast of ARIMA and HW a non-working day: (a) In the ECM; (b) In the 8EB.

Table 5. Forecast Errors of ARIMA and HW of a working day in the ECM

	Holt-Winter	ARIMA
MAE (ECM) [kW]	17.8018	1.1365
MAPE (ECM) [%]	107.9935	9.9044
RMSE (ECM) [kW]	22.6148	1.4875
MAE (8EB) [kW]	20.8968	8.6825
MAPE (8EB) [%]	61.1155	12.0744
RMSE (8EB) [kW]	24.2468	13.2142

4.2   Discussion

The ARIMA model obtains the best short-term prediction results. The average value of the MAPE for five working days (one week) is 9.90% and 12.07% for the
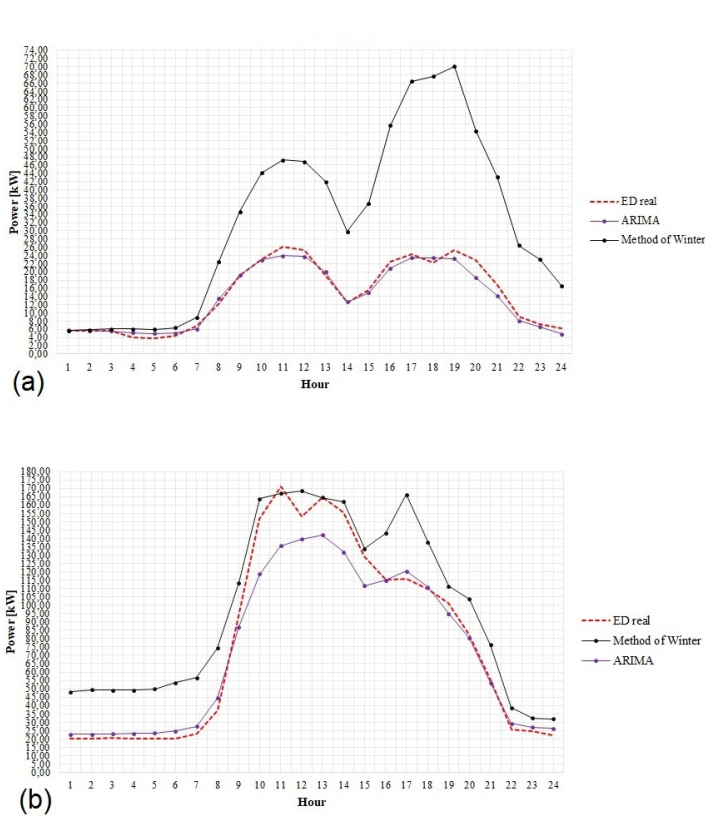
Fig. 6. ED Forecast of ARIMA and HW of a working day: (a) In the ECM; (b) In the 8EB.

ECM and 8EB respectively while the HW method obtains a MAPE of 107.99% and 61.12% for the same case.

Regarding the predictions for non-working days, the HW method evidences a MAPE of 239% and 65% for the ECM and the 8EB respectively. The ARIMA model for this same case presents MAPE values of 13.68% and 5.12% respectively.

The results obtained indicate that the simple exponential smoothing methods and the Holt method cannot predict electrical source profiles. The models based on SLR and MLR did not achieve either good results because the dependent variables have a coefficient of determination (R^2) less than 0.6 concerning the output of the model. On the other hand, the HW method obtains better results, but the prediction errors increase with time.

The application of the Box-Jenkins methodology allows estimating suitable ARIMA models for the prediction of electrical sources for any ED. An adequate

theoretical approximation of the ACF allows estimating the order of the parameters of the model. In comparison with other statistical models, the ARIMA models are the ones that best adjust to the shape of the real electrical consumption, obtaining low prediction errors.

5   Conclusions

In this paper, various statistical forecasting methods have been studied for two university buildings, to assess the effectiveness of these methods. This study was performed in working and non-working days in the Universidad Politécnica Salesiana and Universitat Politècnica de Valencia, which are universities from Ecuador and Spain, and present different load patterns.

Forecasting methods have presented strengths and weakness. The forecasts from Winter methods have strong similarities with the real electricity load pattern, but the power values differ significantly. Linear Regression methods have presented severe disadvantages because the external variables do not have a significant relationship with electricity consumption.

Finally, the ARIMA models are the most suitable, comparing to the other studied models. However, the forecast requires an organized experimental process to select the best model.

The prediction of the ARIMA models could be improved by the incorporation of advanced data clustering techniques, seasonality analysis of time series data and artificial intelligence techniques.

References

1. Abedinia, O., Amjady, N., Zareipour, H.: A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems. IEEE Trans. Power Syst. **32**(1), 62–74 (2017). DOI 10.1109/TPWRS.2016.2556620
2. Amber, K.P., Ahmad, R., Aslam, M.W., Kousar, A., Usman, M., Khan, M.S.: Intelligent techniques for forecasting electricity consumption of buildings. Energy **157**, 886–893 (2018). DOI 10.1016/j.energy.2018.05.155. URL https://doi.org/10.1016/j.energy.2018.05.155
3. Bianchi, F.M., Santis, E.D.E., Rizzi, A., Sadeghian, A.: Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition. IEEE Access **3**, 1931–1943 (2015). DOI 10.1109/ACCESS.2015.2485943
4. Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J.: ARIMA Models to Predict Next-Day Electricity Prices. IEEE Trans. Power Syst. **22**(9), 57–57 (2002). DOI 10.1109/MPER.2002.4312577. URL http://ieeexplore.ieee.org/document/4312577/
5. Ding, N., Benoit, C., Foggia, G., Bésanger, Y., Wurtz, F.: Neural Network-Based Model Design for Short-Term Load Forecast in Distribution Systems. IEEE Trans. Power Syst. **31**(1), 72–81 (2016). DOI 10.1109/TPWRS.2015.2390132
6. Elias, C.N., Hatziargyriou, N.D.: An Annual Midterm Energy Forecasting Model Using Fuzzy Logic. IEEE Trans. Power Syst. **24**(1), 469–478 (2009)
7. González-Romera, E., Jaramillo-Morán, M.Á., Carmona-Fernández, D.: Monthly Electric Energy Demand Forecasting Based. IEEE Trans. Power Syst. **21**(4), 1946–1953 (2006)

8. Hippert, H.S., Pedreira, C.E., Souza, R.C.: Neural networks for short-term load forecasting : a review and evaluation Full Text as PDF Full Text in HTML. IEEE Trans. Power Syst. **16**(1), 4333 (2014)

9. Khosravi, A., Nahavandi, S.: Load Forecasting Using Interval Type-2 Fuzzy Logic Systems : Optimal Type Reduction. IEEE Trans. Ind. Informatics **10**(2), 1055–1063 (2014). DOI 10.1109/TII.2013.2285650

10. Kroposki, B., Johnson, B., Zhang, Y., Gevorgian, V., Denholm, P., Hodge, B.M., Hannegan, B.: Achieving a 100% Renewable Grid: Operating Electric Power Systems with Extremely High Levels of Variable Renewable Energy. IEEE Power Energy Mag. **15**(2), 61–73 (2017). DOI 10.1109/MPE.2016.2637122

11. Liu, B., Nowotarski, J., Hong, T., Weron, R.: Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts. IEEE Trans. Smart Grid **8**(2), 730–737 (2017). DOI 10.1109/TSG.2015.2437877

12. Meira, E., Oliveira, D., Luiz, F., Oliveira, C.: Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. Energy **144**, 776–788 (2018). DOI 10.1016/j.energy.2017.12.049. URL https://doi.org/10.1016/j.energy.2017.12.049

13. Park, S., Han, S.: Demand Power Forecasting with Data Mining Method in Smart Grid. In: 2017 IEEE Innov. Smart Grid Technol. - Asia (2017). DOI 10.1109/ISGT-Asia.2017.8378423

14. Quilumba, F.L., Lee, W.J., Huang, H., Wang, D.Y., Szabados, R.L.: Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. IEEE Trans. Smart Grid **6**(2), 911–918 (2015). DOI 10.1109/TSG.2014.2364233. URL http://ieeexplore.ieee.org/document/6945384/

15. Raza, M.Q., Mithulananthan, N., Li, J., Lee, K.Y.: Multivariate Ensemble Forecast Framework for Demand Prediction of Anomalous Days. IEEE Trans. Sustain. Energy **3029**(c), 1–9 (2018). DOI 10.1109/TSTE.2018.2883393

16. Serrano-Guerrero, X., Escrivá-Escrivá, G., Roldán-Blay, C.: Statistical methodology to assess changes in the electrical consumption profile of buildings. Energy Build. **164**, 99–108 (2018). DOI 10.1016/j.enbuild.2017.12.059. URL https://doi.org/10.1016/j.enbuild.2017.12.059

17. Serrano-Guerrero, X., Prieto-Galarza, R., Huilcatanda, E., Cabrera-Zeas, J., Escriva-Escriva, G.: Election of variables and short-term forecasting of electricity demand based on backpropagation artificial neural networks. 2017 IEEE Int. Autumn Meet. Power, Electron. Comput. ROPEC 2017 **2018-Janua**(Ropec), 1–5 (2018). DOI 10.1109/ROPEC.2017.8261630

18. Tao, S., Li, Y., Xiao, X., Yao, L.: Load Forecasting Based on Short-term Correlation Clustering. In: 2017 IEEE Innov. Smart Grid Technol. - Asia, pp. 1–7. IEEE (2017). DOI 10.1109/ISGT-Asia.2017.8378416

19. Vartanian, C., Bauer, R., Casey, L., Loutan, C., Narang, D., Patel, V.: Ensuring System Reliability: Distributed Energy Resources and Bulk Power System Considerations. IEEE Power Energy Mag. **16**(6), 52–63 (2018). DOI 10.1109/MPE.2018.2863059

20. Ye, X.Z., Ji, T.Y., Li, M.S., Wu, Q.H.: A Morphological Filter-based Local Prediction Method with Multi-variable Inputs for Short-Term Load Forecast. In: 2017 IEEE Innov. Smart Grid Technol. - Asia (2017). DOI 10.1109/ISGT-Asia.2017.8378323