Comparison of Short-Term Load Forecasting Performance by Neural Network and Autoregressive Based Models

M. López, S. Valero, C. Sans, C. Senabre
Dept. of Mechanic Engineering and Energy
Universidad Miguel Hernández
Elche, Spain
m.lopezg@umh.es, svalero@umh.es

A. Gabaldón
Dept. Electric Engineering
Universidad Politécnica de Cartagena
Cartagena, Spain

***Abstract*—In the past decade, many techniques ranging from statistical methods to complex artificial intelligence systems have been proposed by implementing their application to an electric system and highlighting its performance; usually providing a measure of accuracy like RMSE over a definite period. However, there is little research in which a fair comparison among methods is demonstrated, and it is difficult to determine which method would be better suited to a particular electric system or data set. This paper analysis one of the forecasting models running at the National Transport Operator of the Spanish system (REE), which is based on both autoregressive and neural network techniques. The results of this paper help to determine under which circumstances each of the models shows a better performance, which periods are more accurately forecasted by each model and provide valid criteria to choose one or the other depending on the characteristics of the application.**

***Index Terms*— Autoregressive processes, demand forecasting, neural networks, power demand.**

I. Introduction

The development of Short-Term Load Forecasting (STLF) tools has been a common topic in the late years[1]–[3]. Different techniques have been proposed as forecasting engines ranging from statistical methods [4]–[6] to complex artificial intelligence application [7]–[11]. However, the forecasting process involves more stages than selecting a mathematical model. In many cases, the selection of the input variables, the period of time used for training or the treatment of variables (normalization, linearization, filtering…), are more relevant to the final performance of the model than selecting a specific engine [12].

As the referred reviews show, many published papers on the subject describe a particular forecasting model by defining its input, the forecasting engine characteristics and its method of configuration and its results when applied to a particular database. This has provided a wide variety of models for the scientific community to choose from but very little information regarding how to compare the methods against each other, as the characteristics of the database are usually not analyzed.

This problem is analyzed in [13], [14], by proposing a certain methodology to adopt different techniques according to the forecasting problem at hand. In addition, [15] has approached the study of the predictability of load series and how it is possible to characterize a load data series in order to determine which type of forecaster would work best. The importance of standardization and how it would improve the process of developing techniques to solve specific forecasting problems: effects of temperature, long-term trends, special days, etc is also addressed in [12].

Consequently, there is a consensus that a global solution that fits all cases does not exist. Nevertheless, the objective of this paper is to provide a fair comparison of two of the most common forecasting engines working under the same conditions and determine the specific situations in which each of them performs more accurately.

Section 2 of this paper describes the parameters that were analyzed, the specifics of both models used and the methodology used to obtain the results. Section 3 includes all the results for each of the realized tests. The results are provided as an aggregate for a 365-day period but they are also categorized by weather and type of day. The conclusions are exposed in section 4.

II. Material And Methods

The starting point of this paper is a forecasting system designed for the Spanish national system operator, Red Eléctrica de España (REE) [16]. This forecasting system is based on the combination of the forecasts of two independent models, each based on a different forecasting engine. The input information for each model is identical; however, the results from each model are sufficiently uncorrelated so that their combination produces a more accurate forecast.

978-1-5386-1488-4/18/$31.00 ©2018 IEEE

This section presents the data used as input, the characteristics of each of the two models and the methodology used to analyze each model's sensitivity to the available data, to the data pretreatment and the type of day.

A. *Data*

The forecasting system uses temperature, load and calendar data as input. The load data available is an hourly series from 2007 to 2017 for the entire inland system of Spain. Electric load in Spain is affected by many different factors at different scales and with different periodicity. Long-term trend is determined by socio-economic factors: economy growth, energy efficiency investments or air conditioning availability. Temperature is a relevant factor, which non-linearly increases demand on days with extreme values both high and low. In addition, daily load is mostly determined by the type of day. As it is thoroughly explained in [16], there is a complex set of relations between different holidays, adjacent days and other factors that determine how the calendar affects the load.

The temperature data available comprises realdata series from 59 stations across the country reporting actual and forecasted daily maximum and minimum temperatures. The national forecast uses only five selected locations from the 59 available. This selection is based on empirical results and it represents the most demand-intensive areas of the country. In order to capture the dynamics of the behavior of consumers regarding temperature, the temperature lags for up to 4 days is included in the model. Finally, to address the non-linear relation between temperature and load, a Heating and Cooling Degree Day approach is used as described in [16], [17]. The thresholds for each location are obtained by empirical experimentation.

The type of day is determined by a classification system that takes the national holidays from the Official Gazette [18] and assigns each day to one of 41 categories. In addition to the specific category of the day, the month is also included as 11 binary variables. The long-term trends are accounted for by a variable calculated as a 52-weeks moving average of the load.

B. *Models*

The forecasting system includes two models: one that uses an autoregressive structure with errors (AR) and one that uses an autoregressive neural network with exogenous variables (NARX) as a forecasting engine. Both models forecast each hour separately and, therefore, include 24 sub-models. Furthermore, both of them take the same data input previously described. Both models are described in detail in [16] but can be represented by (1) and fig. 1.

$$y_t = \sum_{i=1}^{p} \varphi_i \cdot e_{t-i} + X_t \cdot \theta + \varepsilon_t \qquad (1)$$

The main difference between both models is that while the autoregressive order of the AR model is seven, the order of the NARX model is 14. Both orders were optimized empirically.

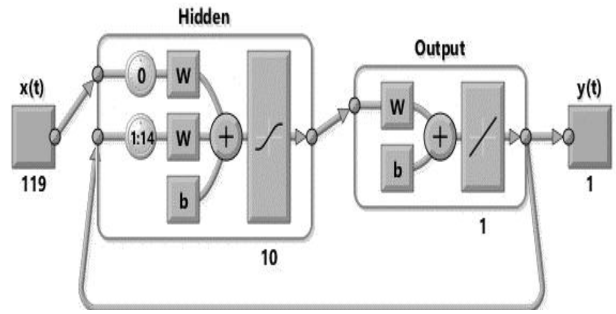

Fig 1. Matlab© representation of a NARX model.

C. *Methodology*

In order to compare the performance of both techniques under different situations, it is necessary to obtain results for a number of days sufficiently large to reflect most real conditions. The average accuracy results reported in this paper refer to a one-year period so that all seasons, types of weather, holidays and special days are included. Both models are tested under real-time conditions in which forecasted temperature values are used at forecasting times when real values are not known yet. This implies that both models include an intrinsic error due to temperature forecast deviations.

The conditions tested on this research work are the amount of historical data available, the amount of temperature locations available, the treatment of temperature variables and, for the NARX model, the number of neurons in the hidden layer.

1) *Historical data available*

Many times, under lab conditions, the databases are deep and there are no restrictions regarding the quality or quantity of available data. However, sometimes, and especially in industry application where data acquisition and storing systems have only recently been set up, the depth of the historical database is shallow and only data from few months or years is available. How far in the past should we include data in our model is a valid question. If we include only recent data, it is possible that our database did not include behaviors that, even though have not occurred recently, they may happen in the forecast horizon. In addition, data from long ago in the past could contain obsolete behaviors that may not repeat under the same conditions in the future, thus lowering the quality of the forecast.

In our research, both models have been trained with datasets from the 3, 5 and 7 previous years. The objective is to determine how each technique reacts to the change of data availability.

2) *Temperature locations*

The forecast of small regions in which climate is homogenous requires only one series of data to capture weather related behaviors. However, when larger regions with higher weather variability are considered, it is important to determine which locations represent behaviors from each weather region better. In addition, not all regions will have the same relevance in overall electricity consumption, as some regions have lower or higher electric capacity than others.

Spain is a large country with high weather and power demand variability across regions. The working model from REE includes data locations on Fig. 2. They represent all weather regions with relevant economic weight: North (Bilbao), East (Barcelona), South (Sevilla), upper center (Zaragoza), lower center (Madrid).

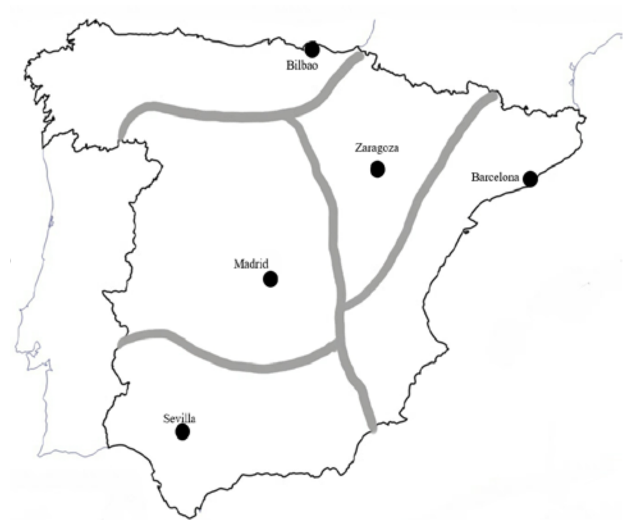

Fig 2. Location of temperature series considered in the models.

In order to test the availability of temperature data, we have tested both models by limiting the temperature input to fewer locations. Our aim is to determine how each technique is affected by having more or fewer data series available.

3) Temperature treatment

The relationship between temperature and load is highly nonlinear and, therefore, it may require some pre-treatment to facilitate the modelling of its effects by the forecasting engines. Fig. 3 shows the scatter plot of average load of regular days at 18h against average temperature for the day. It can be inferred from the graph that there is an increment in load when temperature drops below a lower threshold and whenever it rises above a higher one. This type of behavior is usually modeled using a technique known as Heating Degree days and Cooling Degree Days (HDD and CDD), which splits a temperature series into two series, one responsible for the increase on hot days and the other for the increase on cool days. In order to do this, it is necessary to determine both thresholds for each series.

In our research, the thresholds have already been optimized for each location. However, in order to test the robustness of each model against variation of the optimal thresholds, we have tested both models by introducing variations of up to 12 degrees on each threshold.

4) Number of neurons

In real time applications, the response time of the model may be a critical feature. If the computational burden of the model is too high, then the forecasts may not be produced on time,

This work is supported by the Ministerio de Economía, Industria y Competitividad, Agencia Estatal de Investigación (Spanish Government) under research projects ENE-2016-78509-C3-2-P, and EU FEDER funds and a collaboration with National TSO Red Eléctrica de España.

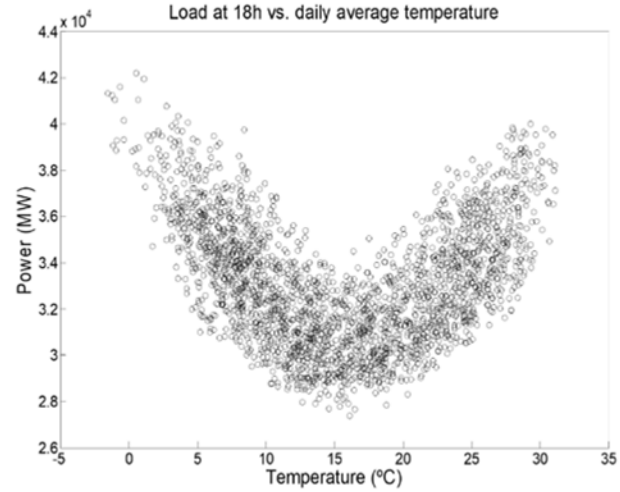

Fig 3. Non-linear relation between temperature at Madrid and National load.

rendering the effort useless. In order to compare both models, we have tested execution time for both models by introducing variations in the topology of the neural network modifying the number of neurons on the hidden layer. This modification affects the ability of the network to model complex behaviors but at a heavy computational cost. Therefore, if the model's accuracy is not compromised, then the number of neurons should be minimized. In addition, it is possible that too many neurons in the hidden layers may cause the model to over fit the training data.

5) Classification of days

In addition to modifying the aforementioned forecasting conditions, it is important to analyze the results in terms of the different types of days. Changes in temperature treatment or historical availability may not produce big changes in overall accuracy of the model but their effect may concentrate on a small number of days for which the right configuration of the model is critical. Therefore, each result is reported for each category regarding type of day and temperature.

III. RESULTS

The results from the tests described in the previous section are presented here. Each result is analyzed globally, but also considering the categories of days that are more closely related to the condition at study. In these tables, special days include days which fall under any of the 41 categories described in [16]. Hot and cold days include those on the top 20 and bottom 20 on the average temperature ranking. Regular days are those that do not fall under any of these categories.

A. Historical data availability

The results in Table 1 show how each model improves its accuracy when more training data becomes available. On these results, the data used in training shifts from 3 to 7 years while other conditions are fixed.The AR model performs better than the NN across the three tests. However, the difference becomes smaller when the database is more complete. This behavior shows that AR model perform better when less historic data is available. Table 1 also shows the results categorized by type of day. These data point out that non-regular types of day are

TABLE I. FORECASTING ERROR WITH INCREASING TRAINING PERIODS

Type of Day	3-yrs		5-yrs		7-yrs	
	AR	NN	AR	NN	AR	NN
OVERALL	1,50%	2,17%	1,52%	1,72%	1,49%	1,59%
REGULAR	1,44%	1,96%	1,47%	1,57%	1,44%	1,47%
SPECIAL	1,91%	3,62%	1,81%	2,71%	1,80%	2,43%
HOT	1,63%	2,65%	1,53%	2,08%	1,66%	1,89%
COLD	1,61%	2,79%	1,73%	1,81%	1,72%	1,48%

affected the most by the lack of availability of larger historical databases. However, it is interesting to notice that neural networks predict cold days better than the AR model when 7 years of data are available.

B. *Temperature locations*

The results of the progressive addition of series of temperature data from new location is included in Table 2. The accuracy of the AR model improves with every new location and the AR model shows a better performance when multiple locations are available. However, the results prove that when only one location is available, the NN outperforms the AR model.

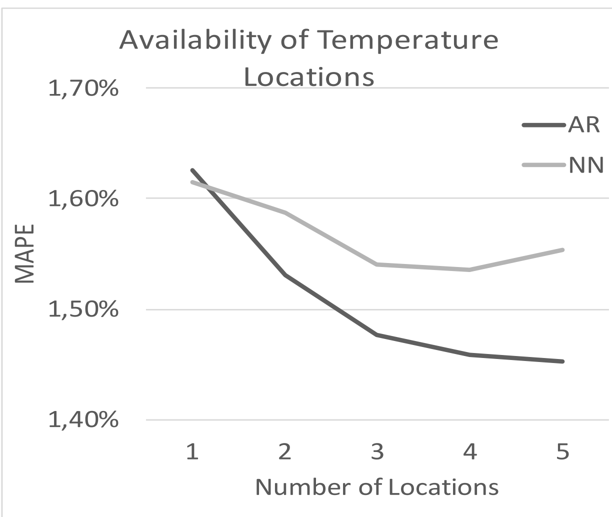

Fig 4. Forecasting error with increasing number of temperature series included in the model.

C. *Temperature treatment*

The key to the temperature treatment is the selection of the thresholds. Fig. 5 shows the forecasting error of both models (AR and NN) as the CDD thresholds are shifted from 13ºC to

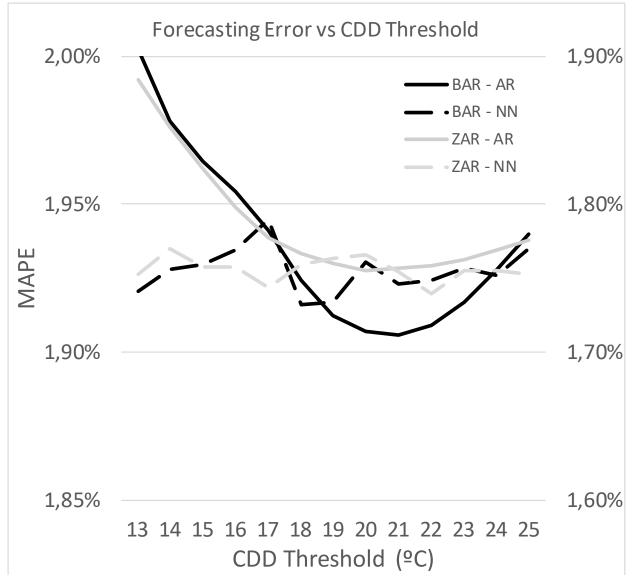

Fig 5. Forecasting Error with Different CDD Thresholds at Zaragoza and Barcelona.

25ºC. The accuracy of both models is affected by the selection of the CDD threshold. It is shown that the AR model may achieve a more accurate performance if the optimal threshold is selected. However, it also shows that the accuracy of the NN model is more robust and has very little variation across the range of the tested thresholds. The optimal threshold can be selected by analyzing the database, but several factors may change it overtime (number of consumers with A/C machines, sensitivity of the consumers to heat and electricity price,…). Therefore, the NN model may have an advantage in applications in which the threshold may be difficult to obtain.

D. *Number of neurons*

The number of neurons on the hidden layer tested varies from 3 to 20. The computational burden of the NN model is lower to the AR model when the number of neurons is lower than 15. The results shown in table 3 show that the optimal number of neurons is 4 and, therefore, the computational burden is not an issue. The fact that the number of recommended neurons is as few as four can be explained if we consider the pretreatment of the data before it is fed to the forecasting engine. The type of day is classified in a binary matrix with one variable for each class, the temperature series have been linearized and the lag of the temperature has been included. Therefore, there are few phenomena to justify the need for a large number of neurons in the hidden layer.

TABLE II. FORECASTING ERROR WITH INCREASING LOCATIONS AVAILABLE

Type of Day	MAD		MAD,BAR		MAD,BAR,VIZ		MAD,BAR,VIZ,SEV		MAD,BAR,VIZ,SEV,ZAR	
	AR	NN	AR	NN	AR	NN	AR	NN	AR	NN
OVERALL	1,63%	1,61%	1,53%	1,59%	1,48%	1,54%	1,46%	1,54%	1,45%	1,55%
REGULAR	1,59%	1,53%	1,48%	1,50%	1,43%	1,45%	1,41%	1,44%	1,40%	1,44%
SPECIAL	1,84%	2,22%	1,86%	2,21%	1,81%	2,15%	1,80%	2,17%	1,81%	2,31%
HOT	1,83%	2,02%	1,63%	1,91%	1,52%	1,84%	1,55%	1,94%	1,55%	1,93%
COLD	2,00%	1,61%	1,81%	1,49%	1,83%	1,47%	1,76%	1,48%	1,75%	1,50%

TABLE III. FORECASTING ERROR WITH INCREASING NUMBER OF NEURONS

Type of Day	Number of neurons 3	4	5	10	15	20
OVERALL	1,56%	1,55%	1,56%	1,59%	1,58%	1,62%
REGULAR	1,49%	1,46%	1,45%	1,47%	1,46%	1,50%
SPECIAL	2,00%	2,10%	2,28%	2,43%	2,36%	2,46%
HOT	2,00%	1,93%	1,95%	1,89%	2,00%	2,04%
COLD	1,55%	1,45%	1,51%	1,48%	1,51%	1,58%

E. Overall results by type of day

The overall performance of both models under optimal conditions (7 years, 4 neurons in the hidden layer and optimal thresholds) is quite similar: 1.45% AR vs 1.55% NN. However, the NN performs significantly better on cold days while it is less accurate on hot days. This could imply that consumer behavior on cold days is not as linear as it is on hot days. Moreover, the accuracy of NN is especially lower on special days. Special days are defined by having very few occurrences each year and, therefore, the data from which to infer consumer behavior is scarce. NN seem to have more problems capturing such behavior than the AR model.

The inclusion of temperature from multiple location reduces forecasting error in all categories but special days, as it was expected. Both hot and cold days improve their results from both models similarly.

IV. CONCLUSIONS

The main objective of this paper is to provide objective proof of which characteristics of a load-forecasting problem favor each of the two techniques analyzed: autoregressive with errors and neural network (NARX). The test conditions of the forecasting problem were availability of historical data, availability of temperature from several locations and size of the NN (as an assessment of computational burden).

The tests that were carried out show that the performance of both models under optimal conditions is very similar. However, the NN model shows a better performance when only one temperature location is available. As a counter fact, the NN model requires at least 7 years of available data to match the performance of the AR model. The computational burden of both models is similar because the size of the optimal number of neurons in the hidden layer of the NN is small.

Depending on the type of day, the NN shows a better performance on cold days, while the AR is more accurate on special days. These conclusions could be used to obtain a weighted combination of both forecast that would take into consideration the type of the forecasted day.

REFERENCES

[1] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, 2016.

[2] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Feb. 2001.

[3] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustain. Cities Soc.*, vol. 35, pp. 257–270, Nov. 2017.

[4] M. T. Hagan and S. M. Behr, "The Time Series Approach to Short Term Load Forecasting," *IEEE Trans. Power Syst.*, vol. 2, no. 3, pp. 785–791, Aug. 1987.

[5] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1535–1547, Nov. 1990.

[6] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 498–505, Aug. 2001.

[7] Y. Chen, Y. Yang, C. Liu, C. Li, and L. Li, "A hybrid application algorithm based on the support vector machine and artificial intelligence: An example of electric load forecasting," *Appl. Math. Model.*, vol. 39, no. 9, pp. 2617–2632, 2015.

[8] J. Wang, S. Jin, S. Qin, and H. Jiang, "Swarm Intelligence-Based Hybrid Models for Short-Term Power Load Prediction," *Math. Probl. Eng.*, vol. 2014, p. 17, 2014.

[9] M. López, S. Valero, C. Senabre, J. Aparicio, and A. Gabaldon, "Application of SOM neural networks to short-term load forecasting: The Spanish electricity market case study," *Electr. Power Syst. Res.*, vol. 91, pp. 18–27, 2012.

[10] V. H. Hinojosa and A. Hoese, "Short-Term Load Forecasting Using Fuzzy Inductive Reasoning and Evolutionary Algorithms," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 565–574, Feb. 2010.

[11] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," *Energy*, vol. 34, no. 1, pp. 46–57, 2009.

[12] M. López, S. Valero, C. Senabre, J. Aparicio, and A. Gabaldon, "Standardization of short-term load forecasting models," in *2012 9th International Conference on the European Energy Market*, 2012, pp. 1–7.

[13] J. Jimenez Mares, K. Donado Mercado, and C. G. Quintero M., "A Methodology for Short-Term Load Forecasting," *IEEE Lat. Am. Trans.*, vol. 15, no. 3, pp. 400–407, Mar. 2017.

[14] E. Almeshaiei and H. Soltan, "A methodology for Electric Power Load Forecasting," *Alex. Eng. J.*, vol. 50, no. 2, pp. 137–144, Jun. 2011.

[15] M. L. García, S. Valero, C. Senabre, and A. G. Marín, "Short-Term Predictability of Load Series: Characterization of Load Data Bases," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2466–2474, Aug. 2013.

[16] M. López, S. Valero, A. Rodriguez, I. Veiras, and C. Senabre, "New online load forecasting system for the Spanish Transport System Operator," *Electr. Power Syst. Res.*, vol. 154, pp. 401–412, Jan. 2018.

[17] J. R. Cancelo, A. Espasa, and R. Grafe, "Forecasting the electricity load from one day to one week ahead for the Spanish system operator," *Int. J. Forecast.*, vol. 24, no. 4, pp. 588–602, 2008.

[18] Gobierno de España, "Boletín Oficial del Estado," *www.boe.es*.