Electric Power Systems Research 154 (2018) 401–412

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

New online load forecasting system for the Spanish Transport System Operator

Miguel López [a,*], Sergio Valero [a], Ana Rodriguez [b], Iago Veiras [c], Carolina Senabre [a]

[a] Universidad Miguel Hernández de Elche, Spain
[b] Red Eléctrica de España (REE), Spain[1]
[c] ALTEN, Spain[2]

ARTICLE INFO

Article history:
Received 12 December 2016
Received in revised form 16 August 2017
Accepted 4 September 2017
Available online 20 September 2017

Keywords:
Load forecasting
Power demand
Neural Network application

ABSTRACT

This paper presents the implementation of a new online real-time hybrid load-forecasting model based on an autoregressive model and neural networks. This new system is currently running at the Spanish Transport System Operator (REE) and provides an hourly forecast for the current day and the next nine days timely every hour for the national system as well as 18 regions of Spain. These requirements impose a heavy computational burden that needs to be considered during the design phase. The system is developed to improve forecasting accuracy specifically on difficult days like hot, cold and special days. In order to achieve this goal, a deep analysis of the temperature series from 59 stations is made for each region and the relevant series are included individually in the model. Special days are also analyzed and a thorough classification of days is proposed for the Spanish national and regional system. The model is designed and tested with data from 2005 to 2015. The results provided for the period from December 2014 to October 2015 show how the addition of the proposed model to the TSO's ensemble causes a 5% RMSE overall error reduction and a 15% reduction on the 59 difficult days considered in the testing period.

1. Introduction

Short-term load forecasting (STLF) has been an active research topic for a long time. The changing characteristics of the consumers (air conditioning universalization or appearance of electric vehicle) and lately also of the producers (different renewables and distributed generation) keep forcing the industry to obtain more and more accurate forecasts every day. Short-term forecasting includes lead times from 1 hour to several days and it provides relevant information to system operators to ensure reliability of the system and to producers for determining schedules and utilization. Another application of STLF is the optimization of market bidding for both sides of the market. The deregulation of the Spanish markets in the past decade has put a lot of pressure on forecasts to improve trading profits.

As it was previously stated, STLF has received a lot of attention in the last decades [1–6]. Forecasting models have evolved from statistical models to more complex models based on different sorts of artificial intelligence. Statistical models include multiple linear regression models [7–9], exponential smoothing techniques [10] and time-series [11–14]. These methods offer accurate results and their research is currently active. Artificial intelligence in STLF comprises several techniques like Artificial Neural Networks (ANN) [15–19], fuzzy logic [16,18,20–23], Support Vector Machines (SVM) [24] or Evolutionary Algorithms [18,20,25–27]. The aforementioned categories refer to the mathematical entity that processes the data, the forecasting engine. Many of the referred techniques are combined in hybrid models that produce forecasts in several steps.

However, the forecasting engine is not the only key aspect of a forecasting models and other processes like data normalizing, filtering of outliers, clustering of data or decomposition by data transform [26–29] are also relevant. Specifically, this paper will focus on temperature and special day data treatment. The characteristics of the load (influence of meteorology, type of day or social events among others) need to be taken into account in order to develop an accurate model for the specific data base [3,4], therefore it is not possible to determine a single technique that outperforms the rest.

Moreover, those systems working under real conditions or tested as real world applications [30,31] are of special relevance [4]. The significance of created knowledge that is validated by con-

* Corresponding author.
E-mail address: m.lopezg@umh.es (M. López).
[1] www.ree.es.
[2] www.alten.es.

http://dx.doi.org/10.1016/j.epsr.2017.09.003
0378-7796/

Nomenclature

Indices

d	Day
d^*	Similar day
h	Hour
m	Forecasting model
s	Weather station
t	Instant

Variables

$e(d, h)$	Forecasting error on day d at hour h
$f(m, d, h)$	Forecast from model m for day d at hour h
$L(T_{sd}, h)$	Load as a function of temperature at station s, on day d at hour h
$Lr(d, h)$	Actual load on day d at hour h
$p(d, h)$	Combined forecast for day d at hour h
T_{sd}	Temperature at station s, on day d
$y(t)$	Output of the model

Parameters and settings

$\alpha_{(1,3)}$	Slope of linearized relation between cold (1) and hot (3) temperatures and load
$\beta_{(1,3)}$	Intercept of linearized relation between cold (1) and hot (3) temperatures and load
p	Number of lags of the AR model
q	Number of lags of the MA model
ϕ	Coefficients for the AR part of the model
θ	Coefficients for the exogenous part of the model
n_y	Number of lags of the Neural Network model
n_u	Number of exogenous inputs of the Neural Network model
k	Intercept for the linear combination of forecasts
Λ	Weight for each model in the combination
n_s	Number of similar days in the optimization period
w	Weight of the similar days in the optimization function
THH_s,THH_c	Temperature thresholds for hot (s) and cold (c) days on station s

tinuous use by the industry instead of self-reported results under lab conditions is especially important for the advancement of the field.

Red Eléctrica de España (REE), the Spanish Transport System Operator (TSO) is a key agent for the efficiency of the electric system and it periodically seeks improvements on its STLF system. Specifically, their previous research work [32], states the relevance of understanding how the load reacts to special days and weather variables. Lack of accuracy in STLF for special days is the main loss derived from their forecasting system and, therefore, it is the focus of this research. Special days show a deviation from the expected load profile due to several reasons: national or regional holidays, Daylight Savings Time (DST), special periods like Christmas or Easter and even extreme temperatures. Even though the effects that these factors may have is generally known, in order to achieve the desired level of accuracy, it is required that each type of day identified for each electric system and even subsystems.

This paper presents a complete, functional forecasting system that is currently running at the REE headquarters providing timely forecasts every hour. The system's ability to perform accurately in a real environment is one of the key aspects of this research, but its main accomplishment is how the local weather variables and the calendar data is treated in order to build an information system in which a forecasting engine may be able to detect patterns and forecast future loads.

Considering all of the above, this paper presents a new hybrid model based on neural networks (NN) and autoregressive (AR) techniques that includes all the processing steps to be executed online (in a real-time environment) and provides a timely forecast not only for the national aggregate but also for each of the 18 regions that the Spanish TSO considers. The information treatment for both temperature data and calendar information conforms the key innovation presented in this paper:

- Section 2.2.2 describes a methodology which is valid to select the best locations from the available temperature data series as well as to obtain the key parameters for the treatment of the selected series.
- Section 2.2.3 describes the classification system for each type of day employing two levels of variables (exclusive and modifiers) to assign a proper classification to each day of the year. This system requires a deep understanding of the behavior of the system but its results prove that a more simplistic classification is insufficient to achieve the necessary accuracy.

The described procedures are original methods that outperforms the initial system described in Ref. [32].

The use of both NN and AR techniques have been widely described and it is not a key aspect of this model: Both methods have been employed as they produce low correlated error from similar input information. Each technique provides a separate forecast, and the final output is based on a linear combination of both techniques optimized over a period of time. The national output is obtained by a similar linear combination of both national forecasts and both aggregations of the regional forecasts. The actual forecast consists on a 24-h curve for each day from the current day to nine days ahead, considering the most relevant information for each horizon of prediction. Although, these forecasts are made hourly the results shown in this paper refer to the forecast made at 9 a.m. for the next day. The results of the proposed methods allow the TSO to improve the general accuracy of its whole system, especially on the difficult days that motivated this work.

This paper is organized as follows: Section 2 summarizes the characteristics of the databases used, the different techniques that have been evaluated and the definition of the definitive model used. On Section 3, the results of the model are shown along with a detailed analysis regarding each sub-model, meteorology, type of day and other variables affecting accuracy. Section 4 contains the conclusions and recommendations of use considering the results exposed on the previous section.

2. Materials and methods

The work described in this paper starts as a project to enhance the STLF system for the Spanish TSO (REE). The project started from scratch from the data acquisition system to the forecast data formatting and alarms system. The input of the system is the historic database along with a series of text files from REE and AEMET (National Weather Agency) with both new load and weather information. The output is a text file with a current-day and nine following days hourly forecast which is casted every hour. One of these files is created for each of the 18 zones of the peninsula plus one for the whole system. These files are then included in the TSO's forecasting ensemble.

During the development of the final models, several techniques were evaluated from which some were finally used and others were discarded. The criteria used to include each technique were based on forecasting results over testing periods of at least two years.

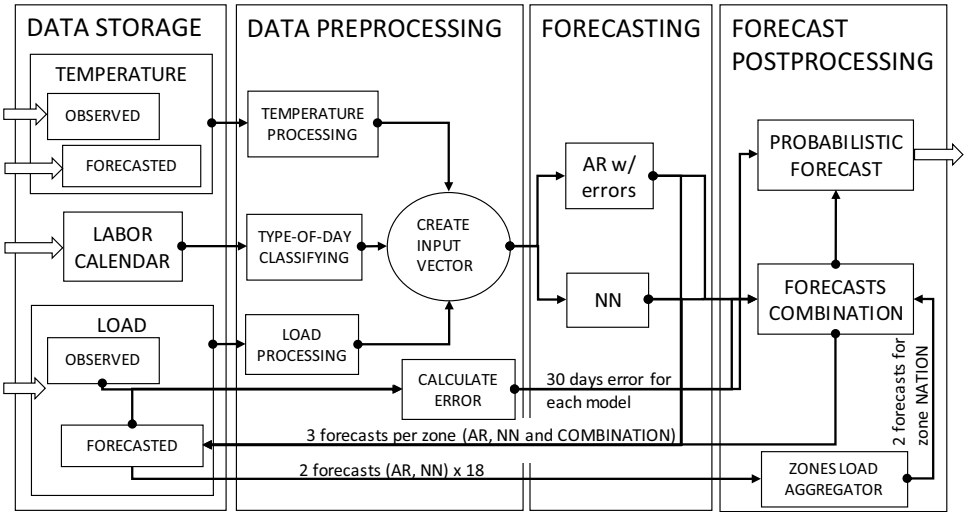

Fig 1. (a) Data storage: temperature, calendar and load data are read from their sources and stored for later use. (b) Data preprocessing: data is filtered for abnormalities and missing points and the input vector is composed according to current time and which hour is being forecasted. (c) Forecasting: input vector is presented to both forecasting engines to obtain 2 values for the same hour. (d) Forecast postprocessing: forecasts are combined and a probabilistic forecast is obtained considering previous errors.

Specifically, the use of wavelets transform was analyzed but, as it is mentioned in Ref. [33], the problem of border distortion forced us to use a previous forecast to pad the series of data at the forecasting end but the output of the system did not provide a forecast more accurate than the initial one. In addition, the forecasting engine described using Self-Organizing Maps (SOM) described in Ref. [19] was tested. However, it was discarded because even though the lab results were valid, the implementation in a real-time application was too computationally consuming. As a conclusion, even though forecasting engines and their training methods receive much of the research attention. The processes that yielded larger improvements and, therefore, are the key aspect of this paper are the information treatment of temperature and calendar data.

This section describes the details of the forecasting system as it was finally deployed and the data flow through each forecasting process.

2.1. Forecasting Sytem and Data Flow

The forecasting system provides hourly load curves for the current day and the next nine days. However, this is constructed by using one model for each of the forecasted hours. In addition, forecasts are done at different times of the day with different information available so we use a different model depending on the time of the forecast. The forecasting system selects the correct models (Neural Network and Auto-regressive) for each forecast. The forecast made at 9 a.m. for the next day is the reference used for the results section: in order to obtain the current-and-next-nine-days curve at this time a total of $15+9\times24=231$ models are used. At 10 a.m., current day and days 1, 2 and 3 are forecasted again using $14+3\times24=86$ different models. The structure of the models to forecast day 1 at 9 a.m. and 10 a.m. is identical but the trained models are different as the relevance of each variable shifts. For instance, the weights of temperature related variables increase as the forecasted day approaches the current time as the weather forecast becomes more reliable. The whole data flow to forecast one hour can be observed in Fig. 1. The process is repeated for each hour of the forecasted interval.

2.2. Data analysis

Even though forecasting engines and training methods receive most of the attention on forecasting models, a proper processing

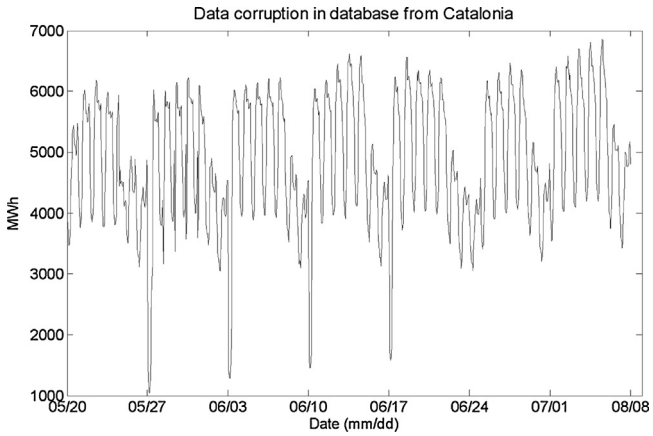

Fig. 2. Example of data corruption at a regional level.

of the right data is probably even more important. The type of data normally used are temperature, calendar and previous load. In this study, humidity and pressure data were also tested but yielded no significant improvement over two-year periods: not in overall accuracy nor in any specific type of day.

This subsection will present the data types used, their main characteristics and the processing stages applied to them in order to produce an input vector that properly suits the capabilities of the forecasting engines.

2.2.1. Load data

Load data is available from 2005 to 2015. Load database contains actual load from the entire inland system and regional data. For each region, the load data used are actually the energy flows at the frontier of each area, which causes that aggregation of regional data differs from inland data in about 0.1%.

In addition, regional data may contain misleading information if energy flow is affected by operations on the grid and, therefore, a proper filtering system should be used to discard any faulty data. Fig. 2 shows data incoherence on four consecutive Mondays starting May, 27th.

Such filtering system is included in the system and it identifies abnormalities by comparing them to an expected range based on forecasts and previous real data. Permanent storage of these cor-

Fig. 3. Long term trend of Spanish national demand.

rected data requires human authorization but if the data is needed and human validation has not arrived then the corrected value is used and an alarm is issued.

Long-term trend in the load series is affected by economic and demographic factors. On previous works on Spanish load forecasting [19], long-term trend is modeled as a linear variable (Fig. 3). However, the impact that the economic recession has had on national electric demand not only requires non-linear terms, but also reduces greatly the predictability of the trends. This has caused that, in order to avoid great long-term discrepancies between the model and reality, it needs to include a linear and a quadratic term and be trained at least annually using only the most recent data.

During our validation process, the frequency of training was tested at 3, 6, 12, and 24 months while using data from the last 3, 5 and 7 years. The results showed that using more than 3 years of data resulted in a loss of accuracy due to the long term trends of the model. Retraining more frequently than once a year did not result in any further improvement.

Seasonal and shorter trends are mostly due to weather and socioeconomic factors like holidays, vacation periods and DST. This type of behavior is modeled through the weather and calendar variables but in order to capture the most recent evolution of load, the last hourly load known at the time of the forecast is included also as a variable.

2.2.2. Meteorological data

The meteorological data analyzed consisted on daily maximum and minimum temperatures, humidity and atmospheric pressure from 59 stations scattered across Spain.

Meteorological data have a definite influence on electricity consumers. Seasonal trends occur causing load peaks during summer and winter time due to extreme high and low temperatures. In order to capture the dynamics of this influence, it is important to determine not only which locations are relevant for each region, and to which extent, but also how the variable lags affect over time. For each location considered, a number of lagged variables is also analyzed. In this manner, we can evaluate the effect of each variable current and recent values. The effect of temperature on load has been widely studied; in this research, the concepts of cold and hot degree days (HDD and CDD as described in Ref. [32]) are used to preprocess the data in order to address the non-linear relationship of load and temperature. However, while in Ref. [32] the temperature input to the system is calculated as a weighted average from 10 locations, our approach is to select for each region the most relevant locations and include all of them as input to the model. In addition, we employ a lagged variable for each location which provides a definite improvement on sudden temperature changes.

The selection of the stations and the parameters to obtain HDD and CDD variables becomes an optimization problem in which we first define a function of the load in terms of the temperature, and the hour (1).

$$L(T_s^d, h) = \begin{cases} \alpha_1 \cdot T_s^d + \beta_1, & T_s^d > THH_s \\ \beta_2 & THC_s < T_s^d < THH_s \\ \alpha_3 \cdot T_s^d + \beta_3, & T_s^d < THC_s \end{cases} \quad (1)$$

where h is the time of the day, T_s^d is the temperature on day d at station s, THH_s and THC_s are the thresholds for cold and hot days at station s. The coefficients α and β are obtained by linear regression for each station and threshold, but in order to select the optimal of said station and thresholds combination, it is also necessary to solve the following optimization problem (2), which defines the minimization of root mean squared error (RMSE) over the training period:

$$\min \frac{\sum_{d=1}^{n} \sqrt{\frac{\sum_{h=1}^{24} \frac{(L(T_s^d,h) - Lr_h^d)^2}{Lr_h^d}}{24}}}{n} \quad (2)$$

The stations taken into consideration are the ones that are inside or near the considered region. This reduces greatly the complexity of the problem as for each region we only evaluate about 8 stations. For the national region, all stations are considered. Fig. 4 shows the scattered plot of the load of Madrid at 18 h vs temperature at 3 different locations and the functions calculated for each case. The RMSE of the fitting shows that temperature data from Madrid has a higher correlation and, therefore should be used over Alicante's and Alicante's over Bilbao's. Fig. 5 shows the result for optimizing the parameters from Eq. (1). The thresholds are selected to minimize RMSE between actual load and $L(T_s^d, h)$. Humidity and atmospheric pressure data have been analyzed as both normalized and raw data but provided no significant improvement and were, therefore, discarded.

Once the stations and thresholds are selected for each region, it is also necessary to determine how many past days to take into account. This is done by splitting the data set into a training set (2005–2012) and a testing set (2013 divided itself into 3 sub-sets). The training set is used to train the models while the testing set is used to determine how many past days allow an improvement in forecasting accuracy. The significance condition is achieved if the variable (or lag) improves the mean average percentage error (MAPE) on all 3 testing sub-sets.

The temperature variables for each model are one HDD and CDD pair of variables for each station and for each day lag considered. Therefore, for a region with 3 stations and a 2-day lag the number of variables is $2 \times 3 \times 3 = 18$.

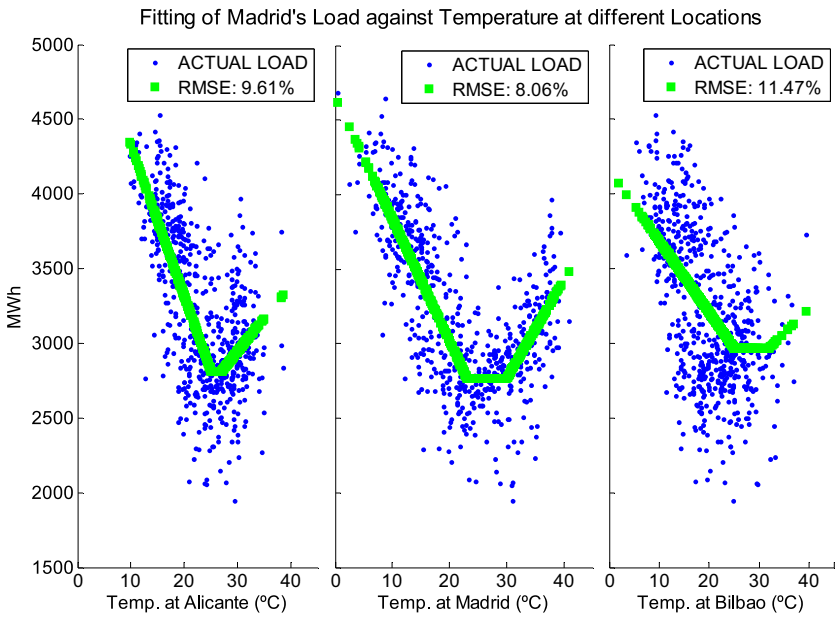

Fig. 4. Actual load in Madrid against temperature at different locations.

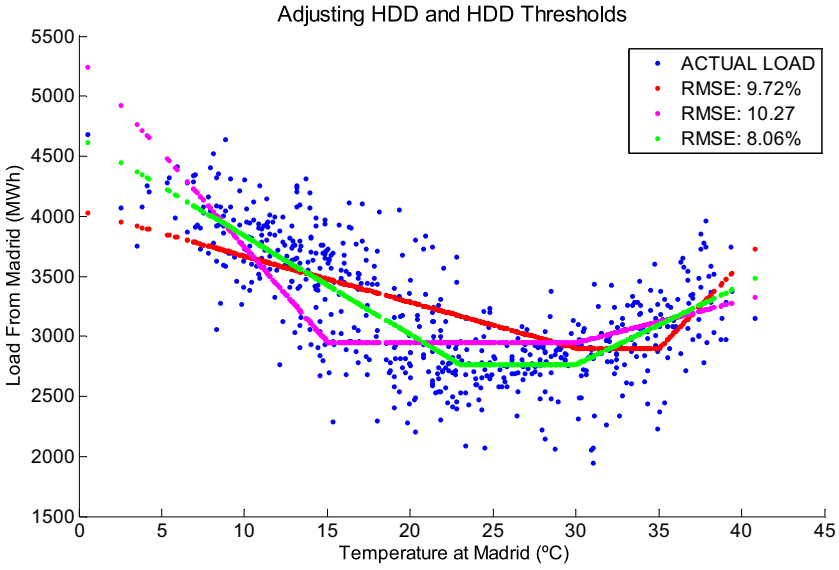

Fig. 5. Actual load against temperature: selection of thresholds.

2.2.3. Calendar data

The effect of calendar on electric load includes single special days like national or regional holidays but also periods of time in which the consumers behavior is shifted, like vacation periods or days after DST. This section will describe the classification of day-types obtained in this research that yielded the best result over the testing period and have proven effective over the actual experiment.

The information about holidays is obtained from the official gazette of Spain (BOE) and from each of the regions. In order to extract all relevant information from this source it is important to understand that special day load profiles are very different from that of a normal day, but also that there can be significant differences among the special days. Therefore, it is necessary to determine how to group properly the special days to facilitate the learning procedure of the forecasting engine.

Special and anomalous days have attracted less attention than normal days from an academic point of view. However, there have been several attempts to tackle this problem [21–23,32,34,35]. The proposal on Ref. [21] introduces the concept of fuzzy logic in the classification of type of days. However, the number of fuzzy categories used is limited to weekdays, Saturdays and Sundays and holidays which is insufficient for the accuracy required in our problem. The problem is similar in Ref. [23], where again four categories are used. The concept that the weekday in which a holiday occur is relevant has been taking into account but not with a sufficiently differentiation among holidays themselves. Fuzzy logic is employed differently in Ref. [22], but also the method proposed requires specific neural networks for each category making it increasingly difficult if the categories increase in number. The use of unsupervised neural networks like SOM is presented on Ref. [34]. This tool is used to classify the load profiles into different categories and analyze their membership in terms of variable known beforehand. In

the case presented in Ref. [34], only 8 categories were employed, but the reason why this technique was discarded in our study is that once the classification is done in as many categories as necessary, the rules to assess beforehand the membership to each category were inconsistent and difficult to articulate. The rule system proposed on Ref. [35] addresses most of the problems encountered in forecasting special days: holidays are different among themselves, weekday affects the holiday profile and days adjacent to holidays are affected. The definition proposed in this model is similar to that of rule 3 on Ref. [35] but it increases the number of considered holidays and it has been design in order to ensure that there are at least 3 similar days to the forecasted day in any 3-year training period. Finally, the previous work described in Ref. [32] set the grounds to determine different types of special days. The specified classification, however, is restricted to six types of days, which is found insufficient, and it is, therefore, thoroughly expanded in this work.

The proposed method consists on using a set of binary variables to express membership of a given day to each category. The categories are split into two groups. The first one includes mutually exclusive categories while the second group contains categories conceived as modifiers to the first category. Each category must be defined carefully in order to avoid a low number of members within the training period that could damage the validity of the sample. Table 1 describes the mutually exclusive variables while Table 2 shows the modifiers:

The bottom line on Tables 1 and 2 show the number of variables used for each type category. For instance, it takes 11 variables for the national special holidays plus 13 for each Easter day as each day is different to any other. However, it only takes one variable to categorize national holidays as all the days listed are considered to have similar profile. Therefore, each day must fall under one and only one of the mutually exclusive variables.

There are also 18 modifying variables that may be simultaneously active along with one of the mutually exclusive ones. Modifiers are mutually exclusive among themselves by their definition as they belong to different times of the year. The only exception to this being the regional modifier which takes into account the percentage of the GNP that the provinces on holiday represent.

These variables are used as input to the corresponding model in order to forecast the national as well as the regional loads. However, the regional forecast uses a set of extra variables for special regional days and modifies the "*before & after holiday*" category to include days adjacent to regional holidays. In addition, the regional modifier is adjusted to represent the portion of the regional GNP that the provinces on holiday within the region represent.

2.3. Forecasting engines

Forecasting engines are the mathematical entities that provide a forecast based on a given input after a proper training that would have enable them to learn the behavior of the load profile. In order to obtain a reliable output, three different forecasting engines were designed employing autoregressive techniques (AR) and nonlinear autoregressive networks with exogenous inputs (NARX). Self-organizing maps (SOM) as described in Ref. [19] were tested but incurred in high computational burden which rendered them useless for online use.

The objective is to obtain as many accurate forecasts as possible as long as their error is not highly correlated. In a later stage, the different forecasts are combined to improve the overall performance. All models were developed using Matlab toolboxes (Econometric and Neural Network developed by Mathworks) and the SOM toolbox 2.0 developed by the Helsinki University of Technology [36]. Both models received the same input variables, as described previously. However, the results achieved by each of them and their configuration are different. The following points provide a brief

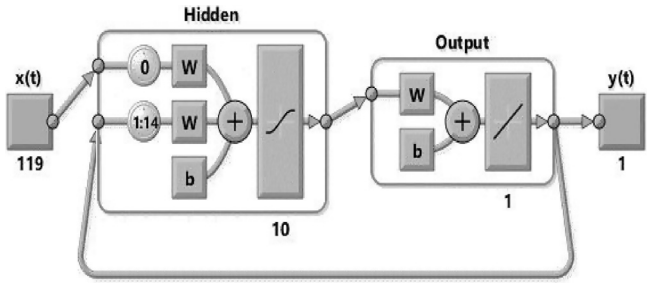

Fig. 6. Schematic view of the NARX system as shown on a Matlab Mathworks visualization.

description of the theoretical foundations of each model and how they have been employed for this application:

Autoregressive Model: Autoregressive models, described in (3), are the most fundamental time series models. In a model AR(p), the current value of the process is expressed as a linear combination of p previous values of the process and a random shock.

$$y_t = \sum_{i=1}^{p} \varphi_i \cdot y_{t-i} + \varepsilon_t \tag{3}$$

If a number of exogenous variables or predictors known at time t, X_t, are considered, then the model becomes (4). However, the model adopted in this research is autoregressive with errors and includes exogenous input, as described in (5).

$$y_t = \sum_{i=1}^{p} \varphi_i \cdot y_{t-i} + X_t \cdot \theta + \varepsilon_t \tag{4}$$

$$y_t = \sum_{i=1}^{p} \varphi_i \cdot e_{t-i} + X_t \cdot \theta + \varepsilon_t \tag{5}$$

The first term of the right side of (5) is very similar to that of a moving average model (MA), shown in (4). While in a MA(q) model the current value of the output is expressed as a linear combination of q previous random shocks, in an AR(p) model with errors, the p previous values considered are the actual known previous errors of the model and, therefore, are not considered random.

$$y_t = \sum_{i=1}^{q} \varphi_i \cdot \varepsilon_{t-i} + \varepsilon_t \tag{6}$$

The parameter p of the model from (5) is obtained empirically and the coefficients φ_i, and $_\beta$ are calculated from the training data by a maximum likelihood method.

NARX Model: NARX neural networks are a nonlinear autoregressive system with exogenous input and its mathematical expression is:

$$y_t = f(y_{t-1}, \ldots, y_{t-n_y}, u_{t-1}, \ldots, u_{t-n_u}) \tag{7}$$

The output value is expressed as a function of n_y previous outputs and n_u inputs known at the time of forecasting. The function implemented for our model is a feedforward neural network. The training for this type of network is realized offline with an open loop in which the output is back fed from the data base as if it was another of the inputs. Once the training is complete, the loop can be closed and the model is able to provide multistep forecasts. The parameters for this model are the same of a regular feedforward network, such as number of hidden layers, number of neurons or training length, plus the number of previous outputs (lags) fed back to the model. Fig. 6 shows a visualization of the model working in closed loop.

Table 1
Definition of exclusive variables used to describe the type of day.

Exclusive variables							
National special days			National holidays		Before & after holiday		Day of the week
Special days with their own specific profile.			Special days with a similar profile		Days which profile is affected by an adjacent holiday		Regular days
All	Easter	Only Monday–Friday	Rest of national holidays marked at B.O.E.		Before holiday	After holiday	
January 1st	Monday before Good Friday to Saturday after	January 2nd	Typically	August 15th	Only Monday–Friday	Only Monday–Thursday	Monday
January 6th	Easter Monday	January 5th		October 12th			Tuesday
May 1st		December 7th		November 1st		Only Friday	Wednesday
December 24th		December 26th		December 6th			Thursday
December 25th		December 30th		December 8th			Friday
December 31st							Saturday
6	13	5	1		1	2	6

Table 2
Definition of modifying variables used to describe the type of day.

Modifiers						
Christmas			Regional	August	Daylight saving time	
Days affected by Christmas period			Days affected by regional holidays	Days affected by holiday periods in August	Days affected by Daylight Savings Time	
Only Monday–Saturday	Only Monday–Friday	Only Sunday	Decimal number from 0 to 1	Traditionally, vacation periods are assigned	Spring	Fall
December 20th	December 27th–29th	December 26th–30th	Expresses the fraction of the GNP that the	splitting August in 4 periods	Sunday	Sunday
December 21st	Jan, 2nd–5th		provinces on holiday	Period 1	Monday	Monday
December 22nd			represent	Period 2	Tuesday	Tuesday
December 23rd				Period 3		
				Period 4		
4	2	1	1	4	3	3

The training of neural networks is based on random initialization and, therefore, its output has a random component. This component is minimized by employing more than one network for redundancy. Several tests using from 3 to 20 networks were carried out to determine the optimal number of networks. Accuracy over two-year periods did not improve when using more than 10 networks while computational burden increased linearly. The 10 different networks are trained and the final output is calculated as the average output once maximum and minimum are discarded.

The settings for the autoregressive and neural networks models are fundamentally different. The only common parameter is the number of lags, which was found to be optimum at a value of 7 for the AR model and 14 for the neural one. A possible explanation is that a linear system like the autoregressive model may not be able to find relevant patterns that the nonlinear system actually finds. The neural network is defined to have 1 hidden layer with 15 neurons, is trained by the Levemberg–Marquadt algorithm. Both the number of lags for each model and the number of hidden neurons are obtained through the same validation process as the number of redundant neural network. Number of lags tested ranged from 3 to 21 and the number of hidden neurons from 5 to 25.

2.4. Forecast combination and probabilistic forecast

The previous point described the two models used to obtain forecasts. The forecasts made by each model should be similar but each one presents different characteristics on their error series. As it was previously stated, if the error series of the different models is poorly correlated, a linear combination of them would provide a new forecast. The information available to determine the optimal combination are the results of the previous days. The type of day is taken into account by giving a heavier weight to the errors occurred on days similar to the forecasted day. This optimization problem is then stated as the determination of the coefficients (0–1) assigned to each model that minimize the average error of the last 30 days. As each models forecast one hour, the minimized error is the error only at the forecasted hour. The case of the national aggregate forecast is different as it uses four forecasts: two from forecasting as a single region plus two from aggregating the results from all regions. The optimization problem, however, is analogous and it is described as follows:

Let ${f^m}_{d,h}$ be the output of model m for the national load on day d and hour h, Λ_m the coefficient from 0 to 1 assigned to model m and k a constant term, then the combined forecast $p_{d,h}$ is defined in (8):

$$p_{d,h} = k + \sum_{m=1}^{4} \Lambda_m \cdot f_{d,h}^m \tag{8}$$

The term k and the coefficients Λ_m are then calculated as a minimization problem of the error, $e_{d,h}$ over a period of time. Let $l_{d,h}$ be the actual load of day d at hour h:

$$e_{d,h} = p_{d,h} - l_{d,h} \tag{9}$$

In order to assign a heavier weight to days similar to the forecasting day, we will assume n_s similar days and a total number of

30 days. We will denote the error on similar days as e^*. Then, the minimization problem is defined as follows:

$$\min\left(\sum_{d=1}^{30-ns} e_{d,h} + w\sum_{d^*=1}^{ns} e^*_{d^*,h}\right) \quad (10)$$

The coefficients are restricted to be in the (0,1) range and the parameter w is experimentally obtained and equal to 5.

Also a probabilistic forecast was provided. While this kind of forecast can be analytically obtained from a linear forecaster, it is different for a system that includes non-linearities like the use of a neural network or a linear combination based on results. In these cases, it is possible to use an empirical approach based on previous results. The first step is to obtain the cumulative distribution function (CDF) from the previous errors as shown in Fig. 7. The probabilistic forecast required to include percentiles 10, 15, 50, 85 and 90. These values are obtained by adding the correspondent offset from the CDF to the actual forecast. Fig. 8 shows a box plot including the actual load and percentiles 10, 15, 50, 85 and 90 for an example date.

2.5. Online system, computational burden and maintenance

The results of the model are needed to take appropriate actions on an hourly basis. Therefore, the tool must provide accurate forecasts on a timely manner. This means that it should include the latest information available and process it sufficiently fast. A ten days forecast for 18 regions of Spain plus a national aggregate was provided. Therefore, to obtain the full national forecast, it is necessary to forecast all regions first.

The execution time for each region may vary depending on the number of variables that its input matrix includes (ranges from 2 to 3 s per region per forecasted day). The 10-day forecast takes about 12 min to complete. In order to reduce this time, the forecast from day 4 to 10 ahead is done only once every day when meteorological forecasts arrive. The updated information available every hour is not significant enough to improve a potential forecast for these longer terms. Then an hourly forecast takes about 4 to 5 min.

The simulations were run on a PC with an i3 processor running at 2 GHz with 4 Gb RAM. The code had not been yet compiled into an executable program so execution in real time of the final application should never exceed the time limits expressed above.

The tool is also designed to be retrained with new data in order to capture new long-term trends like the recent economic recession or its incipient recovery. It is not encouraged to retrain the model more often than once a year since the behavioral change expected is a longer term feature.

3. Results

The forecasting tool presented has been tested from December, 1st 2014 to October, 30th 2015, for a total of 315 days. All data used to test the model was obtained in real time conditions. We will only include results for the whole country, but the contribution from the regional models is reflected in the aggregated models. In this section we will discuss the results from each model that contributes to the forecasting tool and the effectiveness of the combining method proposed. The results will be shown in terms of Mean Average Percentage Error (MAPE) and also Root Mean Squared Error (RMSE) as they are the two metrics used in the industry and will be categorized in order to understand the strong aspects of the model.

Table 3
Forecasting Errors of Proposed Models and Weights in Optimized Reference.

	RMSE	MAPE	Reference weights
Autoregressive	3,66%	3,33%	0,13
Neural Network	2,95%	2,50%	0,22
AR(aggregate)	2,29%	1,97%	0,58
NN(aggregate)	3,09%	2,69%	0,07
Actual forecast	1,83%	1,56%	DNA[a]
Optimized reference	1,95%	1,65%	DNA[a]

[a] Does Not Apply: only the four models are assigned a weight in the reference.

3.1. Forecasting accuracy of the general system and its different sub-models

There are 4 sub-models that provide a national forecast:

- National autoregressive.
- National neural network.
- Cumulative regional autoregressive.
- Cumulative regional neural network.

The difference among the sub-models stem from the type of forecasting engine used (autoregressive model vs. neural network) and whether it is an actual forecast of the national load or the aggregate of all regional forecasts.

The final forecast provided by the system is a linear combination of all four models, whose coefficients are obtained by an optimization of the results from the last 30 days. Table 3 shows the results from all four methods, plus the final combined forecast during the testing period. In addition, in order to provide a benchmark for the combination, the result of the best constant combination of the four models calculated a posteriori is also included. The effects of the combination process will be further analyzed in Section 3.6 but for now the results allow us to assess that the best model is the one based on neural networks (for the national load), but also that the aggregated models provide the best forecast overall. In addition, the combination method is validated as it is more accurate than the benchmark combination. This is possible because the method proposes a dynamic combination that can adapt to the results of the models while the reference is constant for the whole interval. The weight assigned to each model varies as their performance fluctuates.

3.2. Contextualization of the results: comparison with existing forecasters

In order to contextualize the previous results, the Spanish TSO provided reference forecasts from the currently working forecasters at their facilities. These reference models are based on autoregressive engines (Forecasters #1 and #3) and a combination of autoregressive and neural networks (Forecaster #2). The raw input for all three references is the same as the one used for the proposed model except for Forecaster #1, which also uses forecasted cloudiness.

Table 4 shows the result of the three forecasters plus the proposed one. In terms of lowest error, the proposed model ranks 3rd out of the four models. However, while accuracy is the most important factor for a model acting alone, when it comes to combine groups of forecasters it is important to study their correlation and their contribution to the combination. Essentially, two models with low correlation between their errors are likely to yield a more accurate combination. To illustrate this phenomenon, three references have been included: reference A combines the three forecasters from the TSO, reference B adds the proposed model to the combination and reference C combines the best 3 of the four models (Forecasters #1, #2 and this proposal). These references have been

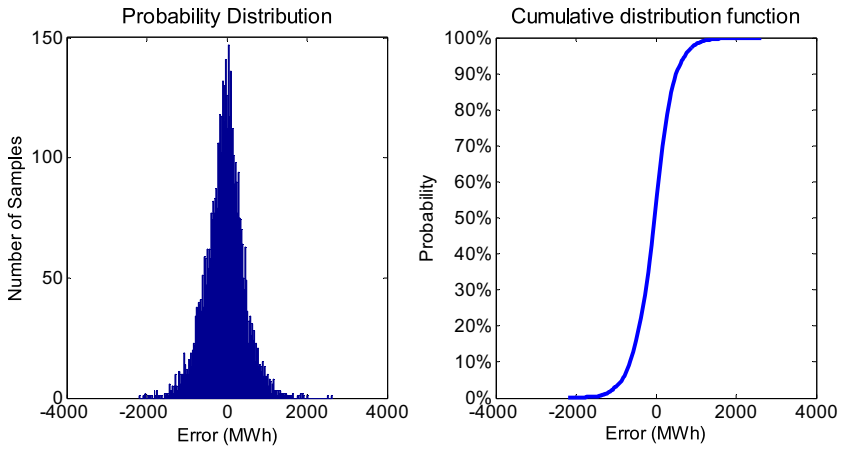

Fig. 7. Probability distribution of the error and cumulative distribution function used for probabilistic forecasting.

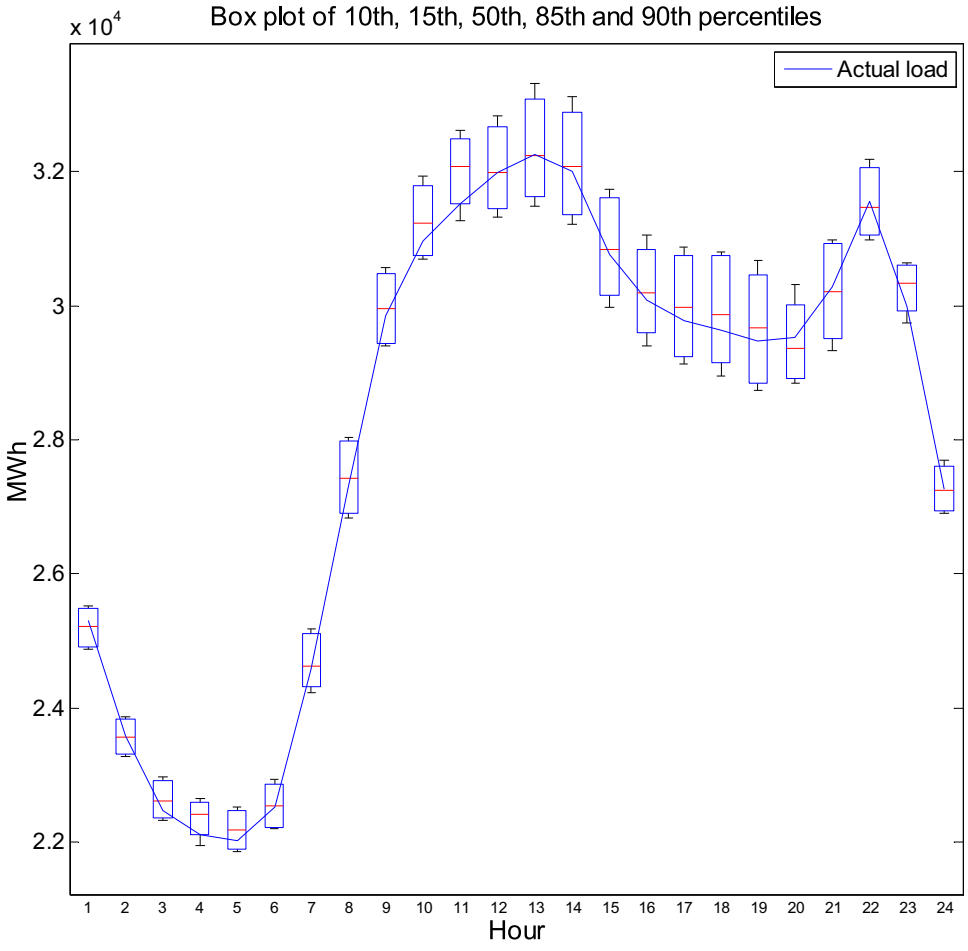

Fig. 8. Probabilistic forecast of day May, 14th 2015.

Table 4
Forecasting Results for Current and Proposed Models and Weights in References .

	RMSE	MAPE	Weights (A)	Weights (B)	Weights (C)
Forecaster #1	1,68%	1,43%	0,27	0,17	0,21
Forecaster #2	2,06%	1,77%	0,2	0,1	DNA[a]
Forecaster #3	1,59%	1,35%	0,53	0,45	0,47
Proposed model	1,83%	1,56%	DNA[a]	0,28	0,32
Optimized ref (A)	1,46%	1,24%	DNA[b]		
Optimized. ref (B)	1,39%	1,18%			
Optimized. ref (C)	1,40%	1,19%			

[a] Does Not Apply: The forecaster does not participate in the reference.
[b] Does Not Apply: Only the forecasters may be assigned a weight in a reference.

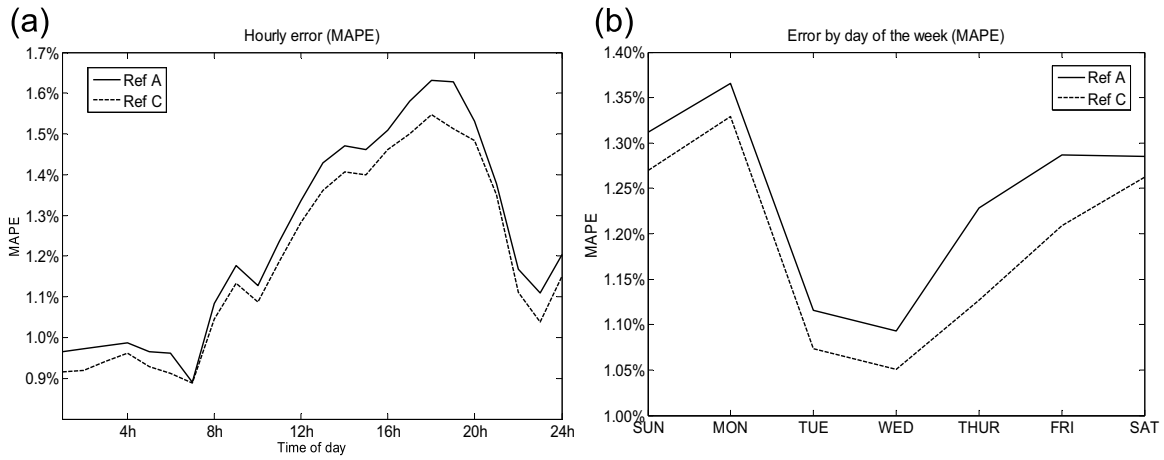

Fig. 9. (a) Forecasting error as a function of the time of day; (b) forecasting error as a function of the day of the week.

obtained as the optimized combination of the forecasters along the testing period considering the results a posteriori.

From the coefficients from reference C it can be observed that while the proposed model is less accurate than forecaster #1, it has a heavier weight in the optimal combination due to its lower correlation between their errors. Therefore, it is safe to say that the proposed model is the second most relevant out of the four forecasters evaluated.

3.3. Forecast of special and difficult days

The previous section showed how the proposed method allows an improvement on the forecasting accuracy over the whole testing period. This section will provide results for a better understanding of what type of days this improvement is focused on. Table 5 shows the results categorized by the 10 coldest and hottest days, national holidays, Mondays before a holiday and Fridays after a holiday, 14 days from Christmas Eve to Jan 6th, and 15 days on Easter.

The results show that the proposed model obtains the most accurate results in all categories except for hot days and Easter, in which it ranks second. Most importantly, it offers the best results over the 59 difficult days considered. Also, the reference forecasts A and C show that by substituting the proposed model instead of forecaster 2, the error is reduced in all categories, except for hot days (forecaster 2 shows a remarkable good behavior in that category) and the error for all difficult days is reduced from 2.01% to 1.70% (RMSE). It also achieves a small reduction on the error for the rest of the days from 1.36% to 1.34% (RMSE).

3.4. Forecast of different weekdays and time of day

Load forecasting error depends highly on the time of day and the day of the week. Generally, it is desired that forecasting error is evenly distributed, but actually certain days or hours prove to be more difficult than others. In this sense, it is important that the general improvement in accuracy of a new model does not overshadow a more unevenly distributed error. On Fig. 9, panels (a) and (b) show how the overall improvement in forecast accuracy of reference C in respect of reference A does not cause any larger errors at any time or weekday.

3.5. Forecast error depending on the forecasting horizon

One of the requirements for the system was to provide a forecast not only for the current and next day but also for the next 9 days. In fact, several forecasts are made throughout the day in order to obtain more refined forecasts for the next and current days. Fig. 10 shows the results in terms of MAPE for the model depending on how much time ahead the forecast is emitted. The marker shows the time at which the forecast analyzed in the previous sections is casted. The accuracy of the next day forecast continues to improve throughout the day from a 1.56% (MAPE) at 8 a.m. to a 1.32% at 11 p.m.

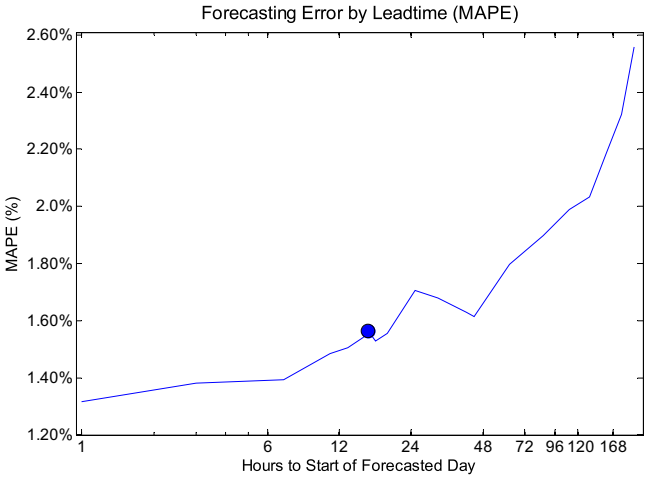

Fig. 10. Forecasting error as the forecast horizon increases.

In addition, the error reaches a maximum of 2.56% (MAPE) when the forecast is made 9 days ahead. The improvement in accuracy of the model from 4 to 9 days ahead is only due to the improvement of the accuracy of meteorological forecasts.

3.6. Combination process

The combination process described in Section 2.4 shows that an independent term is calculated along with the weight for each technique. This means that the combination of all four forecasts is able to reduce the error due to long term trends not captured by the model. This type of errors causes the models to continuously either under or overestimate the load. The optimization over a recent period helps identifying this bias and minimizing it. In this aspect is particularly helpful that the optimization assigns heavier weight to the days similar to the forecasted day than to the rest of days within the optimization period. In this way, the system is able to distinguish a different effect of long term trends on work days than on weekends or special days.

Table 5
Forecasting Results for Current and Proposed Models in Difficult Days.

		Colder days	Hotter days	National holidays	Long weekends	First & last week of year	Easter weeks	Total difficult days	Rest	# Days w/RMSE >4%
# Of days		10	10	7	3	14	15	59	256	N/A
Forecaster 1	RMSE	2,13%	2,22%	3,59%	2,55%	3,29%	1,62%	2,30%	1,57%	8
	MAPE	1,75%	1,87%	2,99%	2,19%	2,76%	1,33%	1,91%	1,34%	
Forecaster 2	RMSE	2,13%	1,68%	4,48%	6,11%	4,70%	2,68%	3,08%	1,88%	24
	MAPE	1,81%	1,42%	3,75%	5,67%	4,27%	2,20%	2,67%	1,61%	
Forecaster 3	RMSE	1,86%	2,14%	3,53%	1,92%	2,79%	1,65%	2,18%	1,48%	8
	MAPE	1,57%	1,79%	2,96%	1,56%	2,40%	1,35%	1,83%	1,26%	
Proposed	RMSE	1,71%	2,03%	2,62%	1,68%	2,61%	2,01%	2,16%	1,75%	7
	MAPE	1,43%	1,69%	2,27%	1,40%	2,22%	1,57%	1,78%	1,49%	
Optimized ref. (A)	RMSE	1,80%	1,67%	3,29%	2,57%	2,78%	1,40%	2,01%	1,36%	5
	MAPE	1,51%	1,36%	2,73%	2,14%	2,38%	1,16%	1,69%	1,16%	
Optimized ref. (C)	RMSE	1,57%	1,69%	2,37%	1,54%	2,09%	1,24%	1,70%	1,34%	4
	MAPE	1,28%	1,40%	2,01%	1,35%	1,84%	1,01%	1,43%	1,15%	

4. Conclusions

This paper presents the last addition to the STLF system of REE and proves how the inclusion of the proposed techniques provides a more accurate forecast especially on the days that historically incurred on largest errors.

4.1. Implementation under real time conditions

The results in Table 4 show that the proposed system is a good candidate to become the third model in the TSO's ensemble as it outperforms Forecaster #2. In addition, Table 4 shows how the best combination of 3 systems not only includes the proposed model but it assigns it a weight of 32% only second to Forecaster #3 and with a significant improvement in overall accuracy of 5%.

In addition, Table 5 shows how the proposed method is actually the most accurate over the 59 total difficult days, which is currently the largest loss of the system and the aspect targeted by the key aspects of the proposed model: temperature and special day treatments.

The model provides forecasts (hourly profiles) for the current and next nine days every hour, as it is needed by the TSO. In order to achieve this in a timely manner, some techniques have been dropped due to their poor performance in relation to their computational burden (SOMs). In addition, the input selection and the neural network topology and redundancy have been designed in order to meet the timing requirements. The proposed system has been widely tested in working conditions.

4.2. Results on hot and cold days

One of the main targets of this research was to define a proper way to treat temperature information for large region forecasts. The approach described in this paper shows a more detailed and tailored solution than the treatment found in Ref. [32]. Its implementation requires an analysis of the relation between the load and each temperature series. However, the results shown in Table 5 show that a one-fits-all solution to the temperature problem is not valid to reach maximum accuracy and that such detailed analysis and customized solution is needed.

The proposed method obtains the best result for the 20-day sample of hottest and coldest days.

4.3. Results on special days

Another target of this research was to improve accuracy of the forecasting system on special days. For the testing period, a total of 59 days has been considered special under the categories named on Table 5. This table shows the results for these specific days, which account for 18% of the total of days considered. The proposed method performs more accurately than any of the individual REE forecasters in 4 out of 6 categories and overall. The comparison of two optimizations, both including and not including the proposed model in the mix, shows that the error on special days is reduced by 15% while the error on the rest of the days is scarcely reduced a 0.8%. These results show how the proposed model achieves both the general and specific goals set. The scarcity of special days is the reason that the overall result is a mere 5% reduction.

4.4. Further work

As an extended collaboration between the research team at UMH and REE, this work will be expanded in the following months to include forecasts of extra-peninsular systems: Balearic and Canary Islands and the two autonomous cities of Ceuta and Melilla, thirteen new zones.

Acknowledgment

This study has been funded by Red Eléctrica de España (REE) as part of its Research and Development activities.

References

[1] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: a review and evaluation, IEEE Trans. Power Syst. 16 (February (1)) (2001) 44–55.
[2] L. Hernandez, et al., A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings, IEEE Commun. Surv. Tutorials 16 (3) (2014) 1460–1495, Third.
[3] M.L. García, S. Valero, C. Senabre, A.G. Marín, Short-term predictability of load series: characterization of load data bases, IEEE Trans. Power Syst. 28 (August (3)) (2013) 2466–2474.
[4] T. Hong, S. Fan, Probabilistic electric load forecasting: a tutorial review, Int. J. Forecasting 32 (3) (2016) 914–938.
[5] A.K. Srivastava, A.S. Pandey, D. Singh, Short-term load forecasting methods: a review, 2016 International Conference on Emerging Trends in Electrical Electronics Sustainable Energy Systems (ICETEESES) (2016) 130–138.
[6] A.R. Khan, A. Mahmood, A. Safdar, Z.A. Khan, N.A. Khan, Load forecasting, dynamic pricing and DSM in smart grid: a review, Renew. Sustain. Energy Rev. 54 (2016) 1311–1322.
[7] A.D. Papalexopoulos, T.C. Hesterberg, A regression-based approach to short-term system load forecasting, IEEE Trans. Power Syst. 5 (November (4)) (1990) 1535–1547.
[8] N. Charlton, C. Singleton, A refined parametric model for short term load forecasting, Int. J. Forecasting 30 (2) (2014) 364–368.
[9] P. Wang, B. Liu, T. Hong, Electric load forecasting with recency effect: a big data approach, Int. J. Forecasting 32 (3) (2016) 585–597.
[10] J.W. Taylor, Short-term load forecasting with exponentially weighted methods, IEEE Trans. Power Syst. 27 (February (1)) (2012) 458–464.
[11] M.T. Hagan, S.M. Behr, The time series approach to short term load forecasting, IEEE Trans. Power Syst. 2 (August (3)) (1987) 785–791.

[12] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Trans. Power Syst. 16 (August (3)) (2001) 498–505.
[13] M.H. Amini, A. Kargarian, O. Karabasoglu, ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation, Electr. Power Syst. Res. 140 (2016) 378–390.
[14] K.G. Boroojeni, M.H. Amini, S. Bahrami, S.S. Iyengar, A.I. Sarwat, O. Karabasoglu, A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon, Electr. Power Syst. Res. 142 (2017) 58–73.
[15] P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi, A neural network based several-hour-ahead electric load forecasting using similar days approach, Int. J. Electr. Power Energy Syst. 28 (6) (2006) 367–373.
[16] Z. Yun, Z. Quan, S. Caixin, L. Shaolan, L. Yuming, S. Yang, RBF neural network and ANFIS-based short-term load forecasting approach in real-time price environment, IEEE Trans. Power Syst. 23 (August (3)) (2008) 853–858.
[17] K. Kalaitzakis, G.S. Stavrakakis, E.M. Anagnostakis, Short-term load forecasting based on artificial neural networks parallel implementation, Electr. Power Syst. Res. 63 (3) (2002) 185–196.
[18] G.-C. Liao, T.-P. Tsao, Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting, IEEE Trans. Evol. Comput. 10 (3) (2006) 330–340.
[19] M. López, S. Valero, C. Senabre, J. Aparicio, A. Gabaldon, Application of SOM neural networks to short-term load forecasting: the Spanish electricity market case study, Electr. Power Syst. Res. 91 (2012) 18–27.
[20] V.H. Hinojosa, A. Hoese, Short-term load forecasting using fuzzy inductive reasoning and evolutionary algorithms, IEEE Trans. Power Syst. 25 (February (1)) (2010) 565–574.
[21] D. Srinivasan, C.S. Chang, A.C. Liew, Demand forecasting using fuzzy neural computation, with special emphasis on weekend and public holiday forecasting, IEEE Trans. Power Syst. 10 (November (4)) (1995) 1897–1903.
[22] K.-H. Kim, H.-S. Youn, Y.-C. Kang, Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method, IEEE Trans. Power Syst. 15 (May (2)) (2000) 559–565.
[23] K.-B. Song, Y.-S. Baek, D.H. Hong, G. Jang, Short-term load forecasting for the holidays using fuzzy linear regression method, IEEE Trans. Power Syst. 20 (February (1)) (2005) 96–101.
[24] Y. Chen, Y. Yang, C. Liu, C. Li, L. Li, A hybrid application algorithm based on the support vector machine and artificial intelligence: an example of electric load forecasting, Appl. Math. Model. 39 (9) (2015) 2617–2632.
[25] J. Wang, S. Jin, S. Qin, H. Jiang, Swarm intelligence-based hybrid models for short-term power load prediction, Math. Probl. Eng. 2014 (2014), p. 17.
[26] Z.A. Bashir, M.E. El-Hawary, Applying wavelets to short-term load forecasting using PSO-based neural networks, IEEE Trans. Power Syst. 24 (February (1)) (2009) 20–27.
[27] N. Amjady, F. Keynia, Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm, Energy 34 (1) (2009) 46–57.
[28] C. Kim, I. Yu, Y.H. Song, Kohonen neural network and wavelet transform based approach to short-term load forecasting, Electr. Power Syst. Res. 63 (3) (2002) 169–176.
[29] Y. Chen, P.B. Luh, S.J. Rourke, Short-term load forecasting: similar day-based wavelet neural networks, 2008 7th World Congress on Intelligent Control and Automation (2008) 3353–3358.
[30] A. Khotanzad, R. Afkhami-Rohani, D. Maratukulam, ANNSTLF-Artificial Neural Network Short-Term Load Forecaster generation three, IEEE Trans. Power Syst. 13 (November (4)) (1998) 1413–1422.
[31] S. Fan, K. Methaprayoon, W.-J. Lee, Multiregion load forecasting for system with large geographical area, IEEE Trans. Ind. Appl. 45 (4) (2009) 1452–1459.
[32] J.R. Cancelo, A. Espasa, R. Grafe, Forecasting the electricity load from one day to one week ahead for the Spanish system operator, Int. J. Forecasting 24 (4) (2008) 588–602.
[33] M. Rana, I. Koprinska, Wavelet Neural Networks for Electricity Load Forecasting — Dealing with Border Distortion and Shift Invariance, in: Artificial Neural Networks and Machine Learning — ICANN 2013: 23rd International Conference on Artificial Neural Networks Sofia, Bulgaria, September 10-13, 2013. Proceedings, V., Mladenov, P. Koprinkova-Hristova, G., Palm, A. E. P. Villa, B. Appollini, and N. Kasabov, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, (2013), pp. 571–578.
[34] R. Lamedica, A. Prudenzi, M. Sforna, M. Caciotta, V.O. Cencellli, A neural network based technique for short-term forecasting of anomalous load periods, IEEE Trans. Power Syst. 11 (November (4)) (1996) 1749–1756.
[35] S. Arora, J.W. Taylor, Short-term forecasting of anomalous load using rule-based triple seasonal methods, IEEE Trans. Power Syst. 28 (August (3)) (2013) 3235–3242.
[36] J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas, SOM toolbox for Matlab 5, 2000.