Available online at www.sciencedirect.com

ScienceDirect

Applied Mathematics and Computation

Applied Mathematics and Computation 00 (2019) 1–22

Periodically Correlated Models for Short-Term Electricity Load Forecasting

Eduardo Caro[1a], Jesús Juan[a], F. Javier Cara[a]

[a] *Universidad Politécnica de Madrid, Madrid, Spain*

Abstract

During the last two decades, the model developed by Cancelo and Espasa (1991) [1] has been used for predicting the Spanish electricity demand with good results. This paper proposes a new approach for estimating multiequation models that extends the previous work in different and important ways. Primarily, 24-hour equations are assembled to form a periodic autoregressive-moving-average model, which significantly improves the short-term predictions. To reduce the computational problem, the full model is estimated in two steps, and a meticulous model of the nonlinear temperature effect is included using regression spline techniques. The method is currently being used by the Spanish Transmission System Operator (*Red Eléctrica de España*, REE) to make hourly forecasts of electricity demand from one to ten days ahead.

Keywords: Reg-ARIMA models, Time series, Forecasting practice, Hourly and daily models, Energy forecasting

1. Introduction

Supplying energy to homes and businesses across any country involves three key elements: generation, transportation and distribution. In most countries where the electricity market is liberalized, the management of the national transmission network falls under an independent operator known as the Transmission System Operator (TSO). The TSO is responsible for managing the transmission of electrical power from generation plants over the electrical grid to regional or local electricity distribution operators. The TSO is also required to maintain a continuous (second-by-second) balance between electricity supply from power stations and demand from consumers, which is achieved by determining the optimal combination of generating stations and reserve providers for each market trading period,

[1] Corresponding Author. E-mail address: eduardo.caro@upm.es, C/ José Gutiérrez Abascal, 2, 28006 Madrid (Spain).

instructing generators regarding when and how much electricity to generate, and managing any contingent events that disrupt the balance between supply and demand. In addition to its roles regarding the real-time dispatch of generation and security management, one of the main tasks assigned to the TSO is the daily scheduling of energy production plants, adding ancillary service markets to the results of the spot electricity market. Each day, generation schedules are drawn up for the next day based on hourly load forecasts, which means that hourly predictions are needed for a time horizon 1 to 48 hours in advance. Often (e.g., when the period contains holidays), it is necessary to prolong the prediction horizon by several days. This work describes a time-series model currently used by the Spanish TSO to make hourly forecasts of electricity demand one to ten days ahead.

Electricity load forecasting has significant economic repercussions. It is important to provide accurate estimates for operating the power system as a basis for energy transactions and decision making in energy markets. Thus, electricity load forecasting has attracted the attention of leading statisticians all over the world during the last three decades, to increase accuracy. From the mid 1980s, numerous articles have been dedicated to methods and models for hourly load forecasting. The main approaches are based on autoregressive-integrated-moving-average (ARIMA) models, multiple regression models, exponential smoothing and structural models, or a mixture of types of models. The collection of papers in [2] gives an indication of the methods that were used.

Univariate methods such as those based on ARIMA models or exponential smoothing can be found in [3]–[5]. These models focus on prediction up to one day ahead, but the authors mention their specific interest in predictions for lead times less than six hours ahead. The results show a huge difference between the prediction errors in the first six hours and those up to 24 hours. Additionally, it is worth noting that the prediction errors depend on the hour of the day at which the predictions are made, and some hours are much more difficult to predict than others. Univariate models are simple, robust and have advantages for very short-term predictions, but their errors are higher than those of other models for lead times more than 12 hours ahead [6].

The desire to improve forecasting accuracy has led to extra incentives in the form of forecasting competitions with numerous participants and a huge variety of proposals, some conventional and others innovative. In the first competition organized by the Puget Sound Power and Light Company in 1990, the winner was an outstanding regression model [7]. The approach based on multiple regression models with separate equations for each hour of the day, has inspired one of the main lines used at present [8]–[10].

Cancelo et al. [1] propose a reg-ARIMA model to predict the daily demand for electricity in Spain. The model includes an extensive set of dummy variables to capture the changes in demand for holidays, and the equation considers the nonlinear relationship between demand and temperature including several piecewise linear regressors. The total number of parameters in the daily model is 185, of which 10 correspond to the moving average part (there are no AR parameters), 32 belong to climatological variables and the remaining 143 correspond to dummy variables related to holidays and other calendar effects. The same model was estimated for each of the 24 hours and used for computing hourly predictions for horizons up to three days at REE. A complete description of the model can be found in [9].

Dordonnat et al. [10] present a linear multivariate periodic state-space model for hourly electricity loads. The model includes stochastic trend components together with fixed and time-varying regression effects. Each equation is associated with a specific hour and has various coefficients and time-varying processes, which are possibly correlated through the disturbances that drive them. The model provides evidence that temperature has a significant effect on the load and that this effect is subject to yearly nonlinear behavior. Cottet and Smith [8] present a Bayesian approach for estimating multiequation regression models coupled with estimation using MCMC. The results show that the weekly, seasonal, meteorological, and dynamic effects differ substantially at different times of day, confirming the basic precept of the multiequation model that is used in this work.

The effect of temperature on the use of electric power has been considered in multiple ways in electricity demand prediction models. The relationship between temperature and demand is nonlinear because the consumption of electric power increases at both low and high temperatures [11]. A simple and effective way to consider this nonlinearity is to include the quadratic term of temperature in the regression equation ([2], [12]). This solution has been improved with many alternatives. Cancelo and Espasa ([1], [9]) propose a piecewise linear regression model: temperature range is divided into four or five sections, thereby allowing a different relationship in each section. This approach has been used by several authors, (e.g., [6]). Dordonnat et al. [10] simplifies this idea and considers only the heating effect within a state space model. Generalized additive models (GAMs) were studied in [13] and [14], where the semiparametric approaches were shown to be well-adapted to the nonlinear behaviors of the electricity load signal. Gaillard et al. [15] introduce a new procedure for performing quantile regression using GAMs. In our article, an additive model based on splines is introduced, and it includes nonlinear effects within the linear regression framework. To review the literature on the effects of temperature and weather conditions on demand, it is necessary to consider that the relationship can change enormously depending on the country in question. An acceptable solution for one electrical system may not be valid for another. In France ([10]), the temperature effect is crucial because winter heating mainly uses electricity. In Spain, the temperature effect is important in winter (for heating), but it is even more important during the summer months (for cooling).

Holidays have a great impact on electricity demand, and they are usually the hardest days to predict; therefore, holidays are one of the main concerns of the system operator. The usual way to consider the effect of holidays on demand is through the inclusion of dummy variables ([1], [9], [16]) in the regression term. Modeling is very complex for many reasons: the demand profile of a holiday is different throughout the year and in turn changes depending on the day on which it is celebrated; in addition, a holiday in the middle of the week alters the demand of the adjoining days. Ziel [16] presents state-of-the-art techniques to deal with public holidays and provides a large load forecasting study for Germany. The number of holidays and their distribution throughout the year differs by country; thus, their relevance for proper modeling may differ in each case, which explains why different approaches to the problem have been used in the scientific literature. Some authors simply renounce predicting holidays by considering them atypical days and removing them from the series [4], [17]. Smith [18] treats them as if they were Sundays. In some studies, the effects of special days are considered random effects [10]. The most common treatment is to include them in the dynamic model using dummy variables. The number of these variables can vary; some authors solve the problem with tens of parameters, while others use several hundreds of parameters [1]. In this work, the characteristics of each special day have been analyzed using dummy variables by considering the information collected over several years. The procedure detects identifiable patterns of behavior that can be used to predict changes in demand for future holidays, and it requires a large number of parameters to consider all the features described in the preceding paragraphs. The method has been validated exhaustively: with several years of data, a model is estimated and its performance is checked by predicting the entire following year. This procedure has been repeated several times, and the prediction results obtained for holidays are highly accurate. This behavior agrees with the findings of Ziel [16] in Germany.

Until now, some statistical models related to the procedure described in this article have been introduced, but a broader revision should include techniques associated with the area of artificial intelligence such as artificial neural networks ([19] and [20]), support vector machines and model hybrids ([21]–[23]). Neural networks have received a great deal of attention in load forecasting literature, and many of the papers applying neural networks present relatively small prediction errors that are comparable to those of time series approaches. Other standard procedures include expert systems and fuzzy logic. The proposed algorithm in this paper is focused exclusively on time-series models.

This work presents three contributions to the problem of predicting demand:

1. The periodic ARIMA model: This model includes a term that connects the 24 equations for each hour. The model obtained is very competitive when compared with other models, for broad time horizons ranging from 1 to 48 hours. The effect of this term on prediction accuracy is meticulously analyzed and explained in Section 5.5.
2. Temperature effect: The nonlinear effect of temperature on demand is considered with a spline regression model. This approach is very useful for adapting the temperature effect in each hour, and as a consequence, it improves prediction accuracy.
3. Special day effect: A critical element in demand prediction is holiday modeling, especially in Spain, where

there are many holidays that have very different effects on demand. This factor is addressed with a systematic procedure that uses dummy variables in the regression model. This procedure could be easily adapted to other electrical systems.

The rest of this paper is organized as follows: In Section 2, the notation is presented and a reg-ARIMA model is implemented for computing electricity demand predictions. Sections 3 and 4 provide the mathematical models for including the effects of temperature and special days, respectively. In Section 5, the developed algorithm is tested using real data from the Spanish electricity system. Finally, conclusions are presented in Section 6.

2. Periodic reg-ARIMA model

Let $\mathbf{Y_d}$ be a 24-dimensional vector that contains the (log) hourly energy demand for day d. This vector series is analyzed through 24 independent univariate models. Denoting $Y_{d,h}$ the component h of the vector $\mathbf{Y_d}$, each univariate model explains the evolution of the series $Y_{d,h}$, where the hour h is fixed and the time index is d. These 24 models are the starting points of the analysis, although the joint analysis requires working with $y_{t(d,h)}$, i.e., the complete univariate hourly series. The time index in $y_{t(d,h)}$ is t, and the cumulative hourly index, $t(d,h) = 24d + h$. This article uses either $Y_{d,h}$ or $y_{t(d,h)}$ to denote the logarithm of the demand for hour h of day d.

The electricity load series $\{y_{t(h,d)}\}$ exhibits strong daily, weekly and yearly seasonal cycles. The mean, variance and correlation structure of the log of the electricity load depend on the hour of the day. The standard models based on the assumption that the mean and autocovariance function are time invariant are clearly inappropriate. In such circumstances a convenient framework is the periodic autoregressive-moving-average (PARMA) model ([24]–[26]), which is an extension of the commonly used ARMA models that allows seasonally dependent parameters. As the seasonal variation is mainly due to the daily pattern, a periodic model with $s = 24$ periods is used.

The main seasonal effect corresponds to the period of the model; in this case, $s_1 = s = 24$. Other seasonal effects can be included, such as weekly and yearly effects, whose cycle lengths are $s_2 = 7 \times s$ and $s_3 = 7 \times 52 \times s$, respectively. All cycle lengths are multiples of the periodic order $s = 24$, which is a necessary condition for our approach. For simplicity, the explanation is restricted to the case of double seasonality.

For each hour h, the periodic model is described using the following two equations:

$$y_{t(h,d)} = \beta_h^T \mathbf{x}_d + z_{t(h,d)} \tag{1a}$$

$$\phi_h(B)\nabla_{s_1}^{D_1}\nabla_{s_2}^{D_2} z_{t(h,d)} = \theta_h(B^{s_1})\Theta_h(B^{s_2}) w_{t(h,d)}. \tag{1b}$$

where (1a) represents a multiple linear regression model with non-stationary and correlated disturbances $z_{t(h,d)}$. Note that the vector of explanatory variables $\mathbf{x}_d$ is constant for all hours of day d and the vector of parameters β_h corresponding to hour h is varies by hour. The disturbances $z_{t(h,d)}$ follow a periodic autoregressive-integrated-moving-average process, as modeled in (1b): B is the backshift operator such that $B^k z_{t(h,d)} = z_{t(h,d)-k}$; ∇_{s_1} and ∇_{s_2} are seasonal difference operators (e.g., $\nabla_{s_1} = (1 - B^{s_1})$); D_1, and D_2 are the orders of differencing for both seasonalities (there is no regular difference in the model); $\phi_h(B), \theta_h(B^{s_1}), \Theta_h(B^{s_2})$ are polynomials in B, B^{s_1} and B^{s_2}, of orders p, Q_1 and Q_2, respectively; and $w_{t(h,d)}$ are independent random variables with mean zero and variance σ_h^2 that can be different for each hour h.

The structure of Eq. (1b) is justified from empirical, theoretical, and computational points of view. First, from an empirical perspective, the standard identification methodology, which is based on the analysis of the simple and partial autocorrelation functions of the 24 daily series, suggests the two moving-average components of the right-hand side and the autoregressive polynomial and the differences that appear on the left-hand side. Second, the model theoretically includes the three components required for explaining the main characteristics of electricity load dynamics: the autoregressive component considers short-term effects, while the moving-average components collect the daily and weekly cycles[2]. Finally, from the computational point of view, a model with such a structure can be estimated very efficiently.

This approach can be seen as a generalization of other multiple regression models with autoregressive errors used in the literature (e.g., [7] and [9]). The main novelty of the proposed model is the autoregressive polynomial $\phi_h(B)$ that connects the 24 equations. It is important to realize that as s_1 and s_2 are multiples of the periodicity of the model s, all terms that appear on the right-hand side for the hour h correspond to loads of the same hour but for different days and, if $\phi_h(B) = 1$ for all h, then the 24 regression models would be decoupled and could be estimated independently. Function $\phi_h(B)$ is precisely the term that relates the demand of an hour with the demands of the immediately preceding hours, which may correspond to hours on the same day or to hours of the previous day. As discussed in Section 5, this factor is key factor and substantially improves the accuracy of short-term predictions. However, the inclusion of this factor in the model has the disadvantage that every hour is nested with previous ones, which greatly complicates estimation. To address this problem, an approximation to estimate the full model in two steps is proposed in this study.

If all the zeros of $\phi_h(B)$ lay outside the unit circle (notice that there are no first-order differences in the model), it is defined as $v_{t(h,d)} = \phi_h^{-1}(B) w_{t(h,d)}$ and (1b) can be written as:

$$\nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} z_{t(h,d)} = \theta_h(B^{s_1}) \Theta_h(B^{s_2}) v_{t(h,d)}. \tag{2}$$

[2]Any ARMA model can be written using the proposed model, possibly with a polynomial with infinite terms.

Using the 24-dimensional vectors $\mathbf{Z}_d$ and $\mathbf{V}_d$ with elements $Z_{h,d} = z_{t(h,d)}$ and $V_{h,d} = v_{t(h,d)}$, respectively, the above univariate hourly process can be written as a daily vector process:

$$\mathbf{Y}_d = \mathbf{G}\mathbf{X}_d + \mathbf{Z}_d \tag{3a}$$

$$\nabla^{D_1}\nabla_m^{D_2}\mathbf{Z}_d = \theta_h(\widetilde{B})\mathbf{\Theta}_h(\widetilde{B}^m)\mathbf{V}_d \tag{3b}$$

where $\mathbf{G}$ is a matrix whose rows contain the regression parameters β_h^T; operator $\widetilde{B}$ is the new backshift operator such that $\widetilde{B}\,\mathbf{Y}_d = \mathbf{Y}_{d-1}$; the difference operators $\nabla^{D_1}, \nabla_m^{D_2}$ have been adapted in a vectoral form, and $\theta_h(\widetilde{B})$ and $\mathbf{\Theta}_h(\widetilde{B}^m)$ are matrix polynomials that contain the moving-average parameters of the model, where m is an integer such that $s_2 = m \times s$ and $s_1 = s$. The vector process $\nabla^{D_1}\nabla_m^{D_2}\mathbf{Z}_d$ is covariance stationary [24]. The important aspect of the new vector model is that matrices in polynomials $\theta_h(\widetilde{B})$ and $\mathbf{\Theta}_h(\widetilde{B}^m)$ are diagonal and under some circumstances each equation can be estimated separately. Note that, although vectors $\mathbf{V}_d$ and $\mathbf{V}_{d-1}$ are correlated, the autocorrelation coefficients for the daily series corresponding to hour h, $\{V_{h,d}\}$ can be zero; in that particular case, it is possible to consistently estimate the parameters associated with an hour h using only the data corresponding to this hour h.

2.1. Estimation algorithm

In short, the proposed algorithm estimates the above model in two steps:

Step 1) The first step consists of estimating 24 univariate models, one for each hour of the day:

$$Y_{h,d} = \beta_h^T\mathbf{X}_d + Z_{h,d} \tag{4a}$$

$$\nabla^{D_1}\nabla_m^{D_2}Z_{h,d} = \theta_h(\widetilde{B})\Theta_h(\widetilde{B}^m)V_{h,d}. \tag{4b}$$

Step 2) The errors of each model in (4) allow the multivariate error $\mathbf{V}_d$ or the univariate time series $v_{t(h,d)}$ of errors to be built. The hourly series $v_{t(h,d)}$ is a periodic autoregression (PAR) of period $s = 24$,

$$\phi_h(B)v_{t(h,d)} = w_{t(h,d)}, \quad h = 1, 2, ..., 24. \tag{5}$$

The PAR model may be expressed in terms of an equivalent multivariate VAR model. One way to compute maximum likelihood estimates of the PAR model parameters would be converting (5) to its equivalent multivariate AR representation [25]–[26]. The computational difficulties associated with such an approach are excessive. Pagano [25] proves that the minimum mean squared estimate of each hour produces consistent estimators of the parameters. A noteworthy result by [25] is that, asymptotically the estimates obtained are independent.

The idea of the aforementioned global procedure is first prewhitened with a reg-ARIMA filter for each hour, and then a periodic autoregression is performed on the residuals.

2.2. Computation of predictions

In this section, the method of computing the predictions by employing the estimated two-step model described in Section 2.1 is explained. The following reasons for using this method are of a practical nature: (i) it takes advantage of the predictions provided by reg-ARIMA models; (ii) the 24-hour daily models are very complex and demanding from a computational point of view; and (iii) it can be very beneficial to use some of the methods available in software packages such as R and MATLAB.

Every time a new observation is obtained corresponding to hour h_0 of day d_0, predictions for the next days corresponding exclusively to the h_0 series are updated using (4), and then, with the newly obtained residual, the prediction of the next hours are updated using (5). This correction significantly improves the predictions up to a 48-hour horizon (see Section 5).

In the hourly cumulative index, the last observation is obtained at $t_0 = 24d_0 + h_0$. To simplify the explanation, only the cumulative index $t = 24d + h$ is used, and when needed, the hour $h_k = h(t_k)$ and the day $d_k = d(t_k)$ corresponding to t_k are used. The residual v_{t_0} is obtained by taking the observed load at t_0, y_{t_0} and the one-step prediction $\widetilde{y}_{t_0}$ obtained from the daily model of hour h_0, as follows:

$$y_{t_0} = \widetilde{y}_{t_0} + v_{t_0}. \tag{6}$$

The same equation can be written for any of the next 24 hours $t_0 + k$, $1 \leq k \leq 24$, but v_{t_0+k} and y_{t_0+k} are unknowns:

$$y_{t_0+k} = \widetilde{y}_{t_0+k} + v_{t_0+k}, \tag{7}$$

where $\widetilde{y}_{t_0+k}$ is the one-step prediction of the corresponding load using the daily model of hour $h_k = h(t_0 + k)$.

The values v_{t_0+k} can be partially predicted by $\widehat{v}_{t_0+k}$ for $k = 1, 2, ..., 24$ through iteration using the autoregressive residual model:

$$\widehat{v}_{t_o+k} = \sum_{i=1}^{p_k} \phi_{h_k,i} \widehat{v}_{t_o+k-i}, \tag{8}$$

where $\widehat{v}_{t_o+k-i} = v_{t_o+k-i}, \quad i \geq k$, and p_k is the order of the autoregressive model of hour h_k. The updated prediction for y_{t_0+k} using the residual information is

$$\widehat{y}_{t_0+k} = \widetilde{y}_{t_0+k} + \sum_{i=1}^{p_k} \phi_{h_k,i} \widehat{v}_{t_o+k-i}. \tag{9}$$

When the time horizon is longer than 24 hours, multiple step predictions of the daily models are used. For example, if k is between 25 and 48 hours, then

$$y_{t_0+k} = \widetilde{y}_{t_0+k} + v_{t_0+k} + \Psi_{h_k,1} v_{t_0+k-24} \tag{10}$$

where $\Psi_{h_k,1}$ is the coefficient of the unit impulse response function of the h_k-daily model. Applying (8), the predictions for $t_0 + k$, with $25 \leq k \leq 48$, are obtained by

$$\widehat{y}_{t_0+k} = \widetilde{y}_{t_0+k} + \sum_{i=1}^{p_k} \phi_{h_k,i} \widehat{v}_{t_o+k-i} + \sum_{i=1}^{p_k} \Psi_{h_k,1} \phi_{h_k,i} \widehat{v}_{t_o+k-24-i} \tag{11}$$

and in general, for $t_0 + k$, with $24(r-1) + 1 \leq k \leq 24r$ and $r > 1$

$$\widehat{y}_{t_0+k} = \widetilde{y}_{t_0+k} + \sum_{j=0}^{r} \sum_{i=1}^{p_k} \Psi_{h_k,j} \phi_{h_k,i} \widehat{v}_{t_o+k-24j-i} \tag{12}$$

where $\Psi_{h_k,0} = 1$ for any k.

Obviously, the effect of the correction is less useful when applied to large horizons. However, it has been found to be highly effective for the first two days of prediction (as discussed in Section 5).

3. Influence of temperature

It is well-known that weather conditions have an impact on energy demand. However, there is not an established method of modelling the relationship. The complexity of the problem is due to the following:

- The effect of temperature in energy demand changes throughout the year: in cold seasons, the demand increases as temperature drops; in hot seasons, the demand increases when temperature increases; and in seasons with milder weather conditions, temperature changes do not significantly alter the energy demand.
- The effect of temperature on demand is not instantaneous. For example, a temperature drop in a single day in winter produces an increase in demand for several consecutive days.
- In large systems, the weather usually varies considerably from one location to another in the grid, indicating that it is necessary to consider the spatial nature of the problem.

In this work, regression-spline techniques [27] are proposed to model the nonlinear relationship between demand and temperature, which basically involves dividing the temperature range into sections, as defined by a sequence of nodes, and fitting each section to a polynomial. The employed procedure ensures that the polynomials are joined smoothly in the nodes and that the whole function and its first and second derivatives are continuous. The positions of the nodes are chosen depending on the shape of the curve to be adjusted, and the number of nodes is chosen using cross-validation with out-of-sample predictions (in general, two or three sections are enough).

Given the nodes, denoted by $\left\{x_i^* : i = 1, 2, ..., r\right\}$, there are many equivalent alternatives when choosing the spline base functions; the following function has been chosen [28]–[29]:

$$b_0(x) = 1 \quad , \quad b_1(x) = x \quad , \quad b_{i+1}(x) = R(x, x_i^*) \qquad i = 1, 2, ..., r \tag{13}$$

$$R(x, x^*) = \left[\left(x^* - \frac{1}{2}\right)^2 - \frac{1}{12}\right]\left[\left(x - \frac{1}{2}\right)^2 - \frac{1}{12}\right]\frac{1}{4} - \left[\left(|x - x^*| - \frac{1}{2}\right)^4 - \frac{1}{2}\left(|x - x^*| - \frac{1}{2}\right)^2 + \frac{7}{240}\right]\frac{1}{24}, \tag{14}$$

where x is a value between 0 and 1. The regression-spline model for variable x is built from (13) and (14).

As might be expected, if temperature observations are available from different locations, it is necessary to select those locations that are considered most representative from the point of view of electricity demand. After analyzing multiple options, the mean of the maximum daily temperature for the 10 most representative cities has been chosen as the most convenient explanatory variable for weather. Using the minimum and maximum values from the historical series (T_{min} and T_{max}, respectively), x is defined as $x = (T - T_{min})/(T_{max} - T_{min})$, where T is the maximum observed temperature of the day.

As mentioned above, the demand of day d depends not only on the temperature of that day but also on the temperature of the previous days, $d-1, d-2, ..., d-K$; thus, the linear term of the model for hour h corresponding to the temperature effect $g_{h,d} = \alpha_h^T X_d$ can be written as:

$$g_{h,d} = \alpha_{h,0} + \sum_{i=1}^{r+1} \alpha_{h,i}^0 b_i(x_d) + \sum_{i=1}^{r+1} \alpha_{h,i}^1 b_i(x_{d-1}) + \cdots + \sum_{i=1}^{r+1} \alpha_{h,i}^K b_i(x_{d-K}). \tag{15}$$

The total number of parameters in this component of the model is $(r + 1)(K + 1) + 1$, where r is the number of nodes and K is the number of lags.

4. Influence of special days

Public holidays have been classified as m types or groups, named $G_1, G_2, ..., G_m$, and every day d is identified using the *dummy* variable Z_{di}. This variable is defined as follows: if $d \in G_i$, then Z_{di} is set to one ; otherwise, $Z_{di} = 0$. The variable $I_{d,j}$ is defined as an indicator of the weekday, as follows: if $mod(d - j, 7)$ equals zero, then $I_{d,j} = 1$; otherwise, $I_{d,j} = 0$.

The days before a holiday are considered with the dummy variables U_{di}^k, where i is the type of day and k the number of days before the holiday. Their effect is modeled as: if $d + k \in G_i$, then U^k_{di} is set to one; otherwise, $U^k_{di} = 0$ $\forall k \in [1, \cdots, k_U]$. Similarly, variables V^k_{di} are employed for the days following a holiday, modeled as follows: if $d - k \in G_i$, then V^k_{di} is set to one; otherwise, $V^k_{di} = 0$ $\forall k \in [1, \cdots, k_V]$. Finally, for regional and local holidays, let P_d be a number from 0 to 1 that represents the percentage of the total population affected by the regional holiday for day d. For special days that affect the total population, P_d equals 1.

Using the above variables, the influence of the special days on demand can be written using the following linear function:

$$f_{h,d} = \sum_{i=1}^{m}\sum_{j=1}^{7}\beta_{hij}Z_{di}I_{dj}P_d + \sum_{i=1}^{m}\sum_{j=1}^{7}\sum_{k=1}^{k_U}\gamma^k_{hij}U^k_{di}I_{dj}P_d + \sum_{i=1}^{m}\sum_{j=1}^{7}\sum_{k=1}^{k_V}\delta^k_{hij}V^k_{di}I_{dj}P_d \tag{16}$$

This formulation contains three sets of parameters: β_{hij}, which measures the increase or decrease in the demand of the hour h of a special day of group G_i that falls on the day of the week j; γ^k_{hij}, which measures the change in the demand of the hour h, k days before a holiday group G_i that falls on the day j; and in the same way, δ^k_{hij} for postholiday days. The values k_U and k_V indicate the number of affected days before and after the holiday, respectively. This model requires $7 \times m \times (k_U + k_V + 1)$ parameters for each hour.

The values k_U and k_V may depend on the type of holiday and especially on the day of the week; therefore, they can be written as k_{Uij} and k_{Vij}, respectively. We have used values equal to 0, 1, 2 or 3 depending mainly on the number of previous/posterior days affected. In total, the number of parameters for this component, depending on the number of groups, may be between 250 and 350.

5. Case study

5.1. Case study settings

In this section, the results for the Spanish mainland electric system are described. This subsection details the settings employed for the case study.

5.1.1. Estimation settings

The time period considered for estimating the parameters is the 12-year period from 2003 to 2014. Such a large estimation period is required to allow for estimating the effects of all special day. Note that to properly estimate the model in (16), each national holiday must fall on each day of the week. Using 12 years ensures that all national days fall on all days of the week.

The demand data have been provided by the Spanish TSO (REE). This information is open-source and can be publicly accessed through `https://demanda.ree.es/demanda.html`.

Concerning the temperature model, the weather information (both forecasted and observed) corresponds to the average maximum daily temperature of the ten most representative locations in Spain. The model detailed in Section 3 uses three equidistant nodes ($r = 3$ and $x_i^* \in \{11.6\,^oC, 21.3\,^oC, 31.0\,^oC\}$) and three delays ($K = 3$). Therefore, for each hourly model, the total number of coefficients to be estimated in Eq. (15) is 16 (the intercept is not considered). The weather information is modeled using 16 regressors.

Concerning the special days model, for each hourly model, 181 parameters are employed to model the eight groups of national holidays using (16) with $P_d = 1$, 13 parameters are employed to model the regional or local holidays using (16) with $0 < P_d < 1$, 14 parameters are employed to model the Easter holidays, 25 parameters are employed to model August, 17 parameters are employed to model daylight saving time clock changes, and 18 parameters are employed to model other atypical days (e.g., strikes). Therefore, 268 regressors are employed to model special day effects.

5.1.2. Forecasting settings

To evaluate the forecasting performance of the model, the proposed procedure is employed to compute one day-ahead forecasts from January 1st, 2015 to December 31st, 2015. Thus, the performance of the method is tested using the 365 days of the year 2015.

The model has been estimated using data until December 2014. Note that all information of national and regional holidays is published by public administrations before the end of December, and no re-estimation has been computed in the forecasting period (2015). The computation of forecasts for 2015 is completely automatic and does not require any human intervention.

5.2. Estimation results: ARIMA component

Box and Jenkins [30] developed a practical approach for building ARIMA models with the best fit to a given time series. Their methodology has had enormous successes in the field of time series analysis and forecasting. The Box-Jenkins methodology uses a three-step iterative approach that include the following: (1) model identification, (2) parameter estimation and (3) diagnostic checking to determine the best parsimonious model from a general class of ARIMA models. This three-step process is repeated several times until a satisfactory model is selected. The identification procedure is based on autocorrelation and partial autocorrelation functions (ACF and PACF). These functions of the 24 series indicate a lack of stationarity that is corrected with a nonseasonal difference ($D_1 = 1$) and a weekly difference ($D_2 = 1$) for all series. The patterns of the 24 autocorrelation functions of the doubly differentiated series are very similar and suggest pure moving average models with polynomials of degrees 3 and 2 for the nonseasonal and seasonal parts, respectively.

As shown in Section 2.1, model estimation (1) is done in two phases. In the first phase, the 24 reg-ARIMA models corresponding to each of the hours of the day are estimated. This gives the polynomials θ_h and Θ_h of model (1). This phase is the most demanding from a computational point of view since it requires the joint estimation of the parameters of the regression equation and the ARIMA component for each hour. Each model requires the estimation of hundreds of regression parameters and a few more (less than ten) corresponding to the polynomials of the ARIMA model. Joint estimation of the parameters is performed using the maximum likelihood method by assuming a Gaussian distribution of the white noise process. The solution is obtained with the subroutines included in the MATLAB Econometrics Toolbox. The nonlinear optimization problem is solved by sequential quadratic programming (SQP).

The last step of the Box-Jenkins methodology, i.e., diagnostic checking, examines the goodness-of-fit of the estimated model. In this case, the residues obtained from the 24 estimated models do not show significant autocorrelation. During this process, several alternative solutions can be obtained that are evaluated through goodness-of-fit measures such as the Akaike Information Criterion (AIC) [31] or the Bayesian information criterion (BIC) of Schwarz [32]. When a sufficiently long time series is available, as in the case at hand, a usual way to check the validity of the model is to evaluate the accuracy of its predictions with an out-of-sample procedure. Table 1 provides for each hour the three parameters included in the moving average (MA) polynomial θ_h and the two parameters corresponding to the seasonal polynomial (SMA) Θ_h. The root mean squared error (the estimated standard deviation of $V_{h,t}$) for each model is given in the column labeled “RMSE(1)”.

In the second phase of model estimation, the residuals of the 24-hour reg-ARIMA models are combined by means of Eq. (5) to estimate ϕ_h, which completes the estimation of the periodic model (1). The estimation of these parameters is performed by least squares, using a linear regression equation that relates the residuals of each hour with the residuals of the immediately preceding hours. Three coefficients have been used in each regression equation. The results of the estimation are shown in columns AR(k) of Table 1. Although some coefficients are not significant, they have been maintained in the models to keep the same structure for each hour. In the last column of Table 1, the root mean squared error of each model in (5) is provided (RMSE (2)). Table 1 shows how the prediction error is reduced from RMSE(1) to RMSE(2) when the information from the previous hours is included. The explanation is simple and logical: the first column corresponds to errors made in a 24-hours-in-advance prediction, and the second column corresponds to errors made in a prediction for the next hour.

hour	Intercept	MA(1)	MA(2)	MA(3)	SMA(7)	SMA(14)	AR(1)	AR(2)	AR(3)	RMSE(1)	RMSE(2)
1	3,328	-0,288	-0,117	-0,054	-0,838	-0,048	0,672	-0,101	0,125	0,01432	0,01081
2	3,234	-0,309	-0,115	-0,043	-0,848	-0,063	0,954	-0,095	0,081	0,01464	0,00582
3	3,159	-0,298	-0,126	-0,035	-0,892	-0,043	0,960	-0,062	**0,020**	0,01472	0,00608
4	3,116	-0,290	-0,132	-0,038	-0,908	-0,019	0,911	0,053	**-0,020**	0,01493	0,00548
5	3,091	-0,294	-0,128	**-0,049**	-0,915	-0,008	0,955	**-0,015**	**-0,003**	0,01489	0,00501
6	3,077	-0,282	-0,150	-0,049	-0,917	**-0,017**	0,944	-0,054	**0,013**	0,01434	0,00491
7	3,077	-0,282	-0,166	-0,044	-0,893	-0,044	1,076	-0,124	-0,070	0,01426	0,00625
8	3,078	-0,245	-0,178	**-0,023**	-0,833	-0,110	1,190	-0,157	-0,089	0,01562	0,00699
9	3,098	-0,282	-0,201	-0,051	-0,796	-0,120	1,047	-0,112	**-0,014**	0,01671	0,00777
10	3,177	-0,346	-0,185	-0,060	-0,849	-0,053	0,980	-0,280	0,148	0,01587	0,00710
11	3,277	-0,394	-0,177	-0,061	-0,854	**-0,031**	1,200	-0,366	0,096	0,01609	0,00609
12	3,327	-0,409	-0,186	-0,066	-0,832	**-0,021**	1,233	-0,334	0,067	0,01672	0,00509
13	3,330	-0,436	-0,180	-0,072	-0,823	**-0,020**	1,202	-0,191	-0,048	0,01731	0,00492
14	3,337	-0,442	-0,183	-0,073	-0,835	**-0,026**	1,154	-0,101	-0,111	0,01749	0,00471
15	3,321	-0,446	-0,170	-0,098	-0,867	-0,049	1,218	-0,214	-0,064	0,01760	0,00514
16	3,268	-0,420	-0,165	-0,120	-0,871	-0,061	1,162	-0,178	**0,002**	0,01835	0,00539
17	3,240	-0,410	-0,184	-0,111	-0,857	-0,066	1,191	-0,240	**0,028**	0,01890	0,00472
18	3,237	-0,401	-0,178	-0,102	-0,826	-0,073	1,236	-0,230	-0,053	0,01920	0,00515
19	3,275	-0,355	-0,170	-0,064	-0,757	-0,055	1,345	-0,361	-0,099	0,01906	0,00686
20	3,346	-0,328	-0,197	-0,079	-0,725	**-0,035**	1,029	-0,257	0,060	0,01755	0,00704
21	3,432	-0,290	-0,185	-0,066	-0,713	**-0,029**	1,108	-0,306	0,040	0,01617	0,00633
22	3,509	-0,330	-0,193	-0,062	-0,771	-0,067	0,822	-0,179	0,131	0,01403	0,00654
23	3,466	-0,324	-0,187	-0,080	-0,889	-0,080	0,974	-0,297	0,098	0,01304	0,00708
24	3,391	-0,322	-0,163	-0,052	-0,842	-0,116	1,058	-0,260	0,117	0,01382	0,00666

Table 1. Two-stage estimates of the model using the Spanish mainland demand data.

5.3. *Estimation results: the influence of temperature*

The weather explanatory variable used in the model is the average value of the maximum temperatures recorded every day in ten representative peninsula locations. Weather models predict these values with great accuracy from one day to another.

The individual interpretation of each coefficient is not very informative. However, the joint interpretation of the estimated functions is meaningful. Fig. 1 provides the estimated influence of temperature on demand for four different hours. The effect of the day's temperature and the two previous days are represented for each hour. For brevity, the 24 graphs are not included in the document.

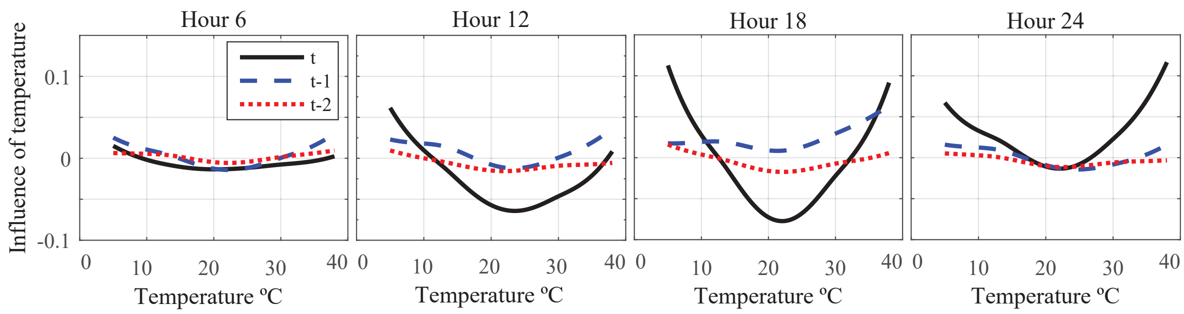

Figure 1. Influence of temperature on Spanish electricity demand: day's temperature (solid line), the previous day $d-1$ (dashed line) and the day $d-2$ (dotted line).

Comparing the four graphs, it is observed that the effect of temperature on demand differs greatly by the time of day. The solid line in the four graphs shows the parabolic shape of the effect of the temperature of the current day on electricity load. Temperature affects demand less at night than during the rest of the day. During the early morning hours, the temperature of the previous day has the most significant effect. After examining the 24 graphs, it is observed that the estimated coefficients vary by hour in a rather systematic fashion.

5.4. *Estimation results: special days*

A large number of dummy variables have been included in the model to study the effect of holidays and special days using (16). The effect of a holiday with a fixed calendar date (e.g., December 25th) is difficult to measure because it varies according to the day of the week on which it falls, and its estimate is usually based on only one or two observations.

The model in (16) has been estimated by considering nine sets of holidays ($m = 9$), eight of which were determined by the eight national holidays, and the ninth of which includes all regional holidays. The estimated parameters allow for the interpretation of how electric demand is affected by public holidays.

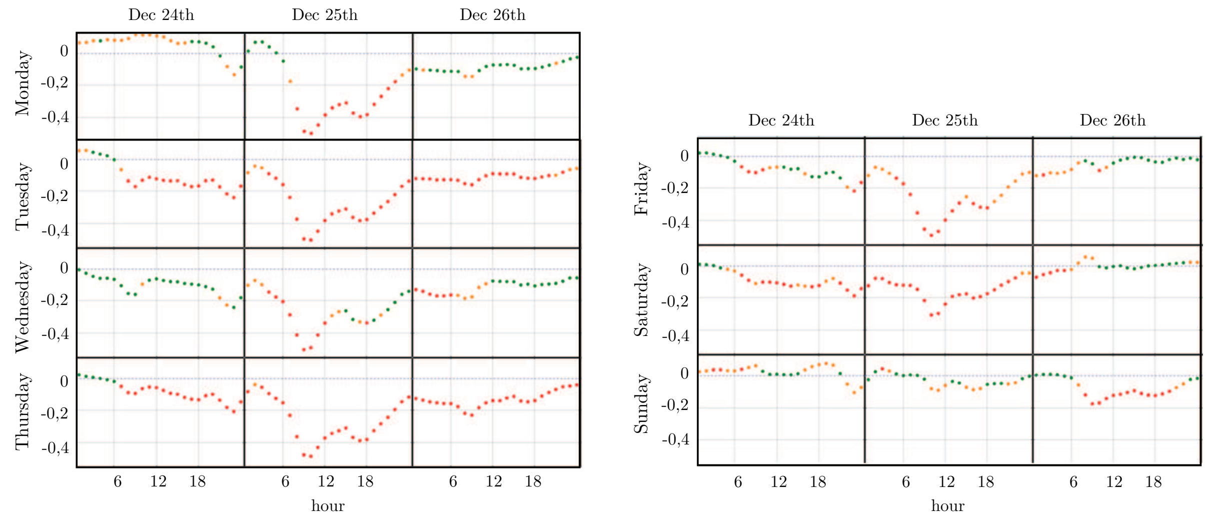

Figure 2. Influence of the public holiday on December 25th

For illustrative purposes, Fig. 2 shows the effect of the public holiday Christmas Day and the two adjacent days (in columns) depending on the day of the week of Dec. 25th (in rows). Each point corresponds to a parameter, the abscissa indicates the time of day, and the ordinate value represents the estimated value. Each value measures the load reduction of the holiday compared with a normal day. When Christmas Day is celebrated on a weekday (Monday to Friday), a significant decrease in demand occurs relative to what would be expected for a working day. For example, the maximum reduction is observed at 9 am and 10 am, when demand may decrease by fifty percent. The variation is much smaller if Dec 25th is celebrated on Saturday or Sunday. The reduction is somewhat smaller on Dec 24th, when the biggest decline is approximately 20% and occurs at 10 pm and 11 pm from Monday to Friday. Dec 26th is also affected but more modestly; the largest change takes place on Monday (Dec 25th is a Sunday).

This detailed analysis has been performed for each group of holidays. The parameter estimates are very different for different groups. Behavioral differences are also observed depending on the day of the week, which means that each hour modeled requires approximately 280 parameters. This particular parametrization is also recommended in the original work of [1] and has provided excellent results in predictions of demand for holidays [9].

5.5. *Forecasting results*

The objective of this work is to obtain accurate forecasts for prediction horizons ranging from one hour to several days. In this section, the results for horizons from 1 to 48 hours are analyzed. The prediction time interval corresponds to the year 2015. A complete series of 365 days has been considered.

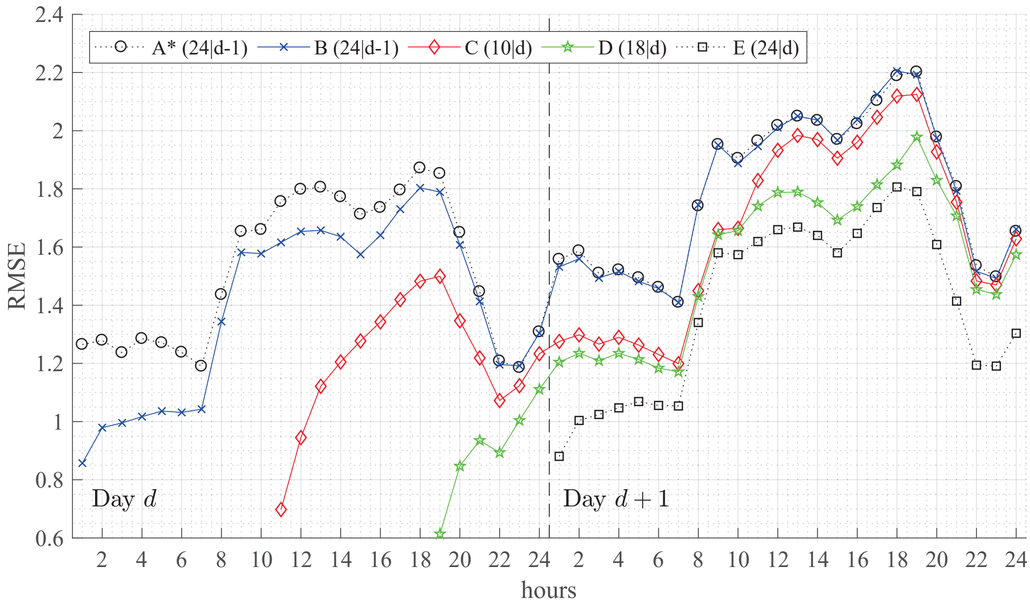

Figure 3. Hourly RMSE for the prediction period (January 1st, 2015 until December 31st, 2015).

Fig. 3 shows the accuracy for the predictions computed at different instants for day d and day $d+1$. The forecasting performance is assessed by means of the RMSE, as suggested in [33].

1. The top line $A*$ represents "uncorrected" predictions $\widetilde{y}_{t_0+k}$ (see Section 2.1) made with the information up to midnight of the previous day $d-1$. The predictions are made using the 24 daily models in (4) without considering

the correction provided by the autoregressive component in (5). Day d errors (left part of the graph) correspond to one-step predictions, and $d + 1$ day errors (right part of the graph) correspond to two-step predictions.

2. Line B (marked with x) is similar to line $A*$ (predictions from midnight of $d - 1$) but is updated with autoregressive terms. A significant reduction in error can be observed in the early hours, but the effect gradually disappears until 5 pm on day d, where the two lines overlap.

3. The prediction errors made on day d at 10 am are shown in line C (marked with diamonds). Logically, the observed demand up to 10 am is very informative for predicting what will happen at 11 am and the hours that follow. Thus, compared with B, a huge improvement can be seen in the predictions for the same day d. However, from a practical point of view, the reduction in errors that can be seen in the predictions of the day $d + 1$ is far more crucial. While comparing B and C in day $d + 1$, two different factors should be considered as follows: (a) from 1 am to 10 am, the differences are mainly because line C corresponds to one-step predictions and line B corresponds to two-step predictions; and (b) from 11 am to midnight, both lines B and C correspond to two-step predictions (the reduction in this part is a direct consequence of the autoregressive component in (5)).

4. Line D corresponds to the errors for predictions made at 6 pm of day d (solid circles). It is interesting to compare curves C and D. The new observations from 11 am to 6 pm do not greatly improve the predictions from 1 am to 10 am. These observations significantly improve the predictions from 11 am to 6 pm (notice that C represents two-step predictions and D represents one-step predictions). Finally, predictions at 6 pm are improved from 7 pm to midnight; however, in this case, both are two-step predictions.

5. Line E has been included as a reference. Line E in $d + 1$ is exactly the same as line B in day d. A large gap between one-step (line E) and two-step predictions (line B) for the 24 hours can be observed. During day d, the more updated information that is included, the it gets to curve E.

Concerning the computational specifications, these case studies have been implemented in MATLAB using a 64-bit eight-core i7 processor (3.6 GHz max.) with 16 GB of RAM. Model estimation requires about two hours of computing CPU time, and the computing of one-day-ahead predictions requires less than 20 seconds. Note that model estimation should be performed once a year, whereas forecasting computation should be performed hourly.

5.6. Benchmarking and performance comparison

To evaluate the accuracy improvement of the model for temperature and holiday effects, the performance of the proposed method is compared with the following modifications:

- Proposed approach without the temperature effect, where the thermal information is not included in the model.

- Proposed approach with a basic model for holiday effect, which considers only the influence of holidays on demand depending on the day of the week from Monday to Saturday. Thus, six groups of 24 regressors have been employed.

- Proposed approach using a piecewise linear temperature model ([34]).

Fig. 4 and Table 2 provide the hourly RMSE for the previous methods, using the same forecast settings described in Section 5.1.2. It can be observed that:

- The model that does not consider the temperature effect provides a higher-error forecast, especially during sun hours. The one-day-ahead prediction accuracy of this method is 41% worse at computing the forecasts at 10:00 am.

- The proposed modeling of holidays provides a 36% improvement compared with the forecaster with a basic model for holidays.

- For temperature modeling, the proposed method based on spline-regression techniques provides an improvement of 11% compared with a piecewise linear model ([34]).

Hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	TOTAL
Proposed	1.3	1.3	1.3	1.3	1.3	1.2	1.2	1.5	1.7	1.7	1.8	1.9	2.0	2.0	1.9	2.0	2.1	2.1	2.1	1.9	1.8	1.5	1.5	1.6	1.68
No temp. Effect	1.7	1.6	1.5	1.5	1.5	1.4	1.3	1.6	1.9	2.0	2.3	2.5	2.6	2.7	2.8	3.0	3.2	3.3	3.3	3.1	2.8	2.5	2.5	2.6	2.38
No holid. Effect	1.6	1.6	1.6	1.6	1.6	1.6	1.7	2.3	2.6	2.4	2.5	2.5	2.5	2.4	2.3	2.5	2.6	2.8	2.8	2.7	2.7	2.5	2.4	2.2	2.29
Temp. Linear	1.6	1.6	1.5	1.5	1.5	1.4	1.4	1.6	1.9	1.9	2.0	2.1	2.1	2.1	2.0	2.1	2.2	2.3	2.3	2.1	1.9	1.7	1.7	1.8	1.86

Table 2. Hourly RMSE for the prediction period, compared with alternative models.

5.6.1. Comparison with Cancelo et al. (2008)

To check the performance of the proposed method, it has been compared with the work by [9], which provides the real results of the forecasting system implemented by the Spanish Transmission System Operator on that date. All settings of the case study have been replicated: the methodology proposed in this paper has been applied to obtain a one-day-ahead forecast for the Spanish electrical energy consumption, computed at 10 am the day before. The study period is from January 1st until December 31th of 2016. The obtained results are provided in Table 3: the MAPE of the method described in Section 2 is compared with the MAPE reported in [9] (indicated within parenthesis).

As in [9], the results detailed in Table 3 and Fig. 5 refer to 365 consecutive days, without exceptions or corrections for vacation periods, public holidays, sudden weather changes, or unexpected events. From Table 3 and Fig. 5, it can

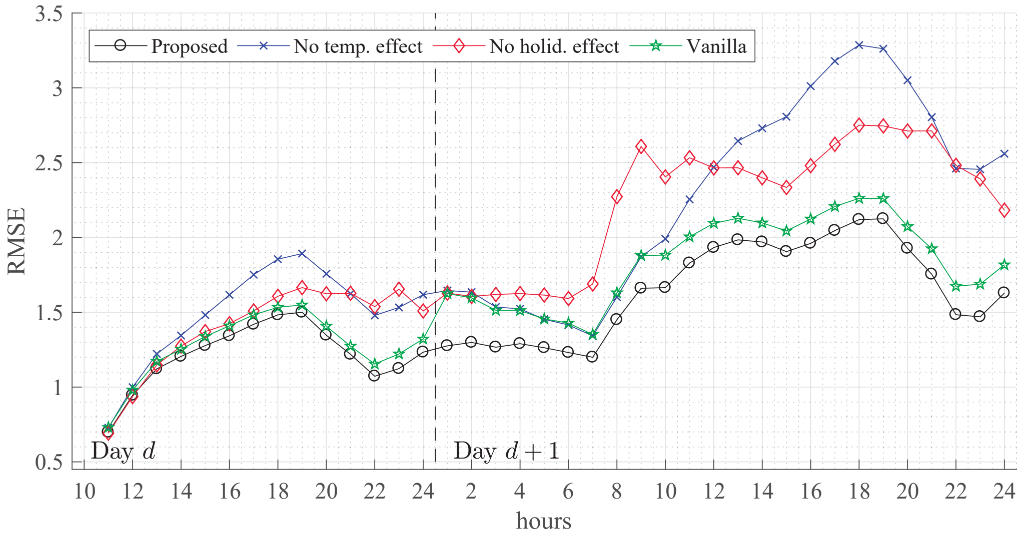

Figure 4. Hourly RMSE for the prediction period, compared with alternative models.

be observed that the proposed method clearly outperforms the traditional one: for all days of the week and for all the hours of the day.

6. Conclusions

Power system management needs short-term forecasts with horizons of one hour to one week to ensure system stability and optimal dispatching; however, load forecasting is an essential part of day-to-day trading.

This article describes the model used by the Spanish TSO to make hourly forecasts up to 10 days ahead. The model considers the hourly series as a periodically correlated process. It incorporates temperature as an explanatory variable and considers the effect of holidays on demand. Direct estimation of the model by maximum likelihood is complicated because it contains a large number of parameters. The problem is solved with the following estimation procedure: first, an independent model is obtained for each hour, which is the usual strategy in most of the models proposed in the literature, and second, the parameters are jointly estimated to relate each hour to the previous hours.

This article has applied the methodology of regression splines to the problem of estimating the functional relationship between weather and electricity demand. A nonlinear relationship is observed graphically. The results agree with intuitive expectations, and the graphs clearly show how temperature influences demand changes throughout the day. The dynamic effect of temperature on demand is incorporated by simply adding lagged temperature variables. The number of nodes is chosen via cross-validation; in this implementation, 3 equidistant nodes are used. It has been proven experimentally that the overall behavior of the method is quite robust.

Hour	Type of day							All days
	Sun	Mon	Tue	Wed	Thu	Fri	Sat	
1 am	0.76 (1.55)	0.82 (1.28)	0.89 (1.47)	0.79 (1.36)	0.74 (1.47)	0.75 (1.04)	0.72 (1.23)	0.78 (1.34)
2 am	1.02 (1.70)	0.93 (1.05)	0.99 (1.52)	0.87 (1.33)	0.87 (1.45)	0.99 (1.17)	0.95 (1.37)	0.95 (1.37)
3 am	1.93 (1.78)	1.07 (1.08)	1.08 (1.66)	1.00 (1.33)	0.91 (1.33)	1.09 (1.06)	0.82 (1.23)	1.13 (1.35)
4 am	1.13 (1.80)	1.09 (1.19)	1.09 (1.61)	0.91 (1.23)	0.88 (1.28)	1.02 (1.02)	0.91 (1.32)	1.00 (1.35)
5 am	1.17 (1.80)	1.21 (1.22)	1.23 (1.69)	1.03 (1.41)	0.97 (1.26)	1.05 (0.99)	0.93 (1.34)	1.09 (1.39)
6 am	1.11 (1.61)	1.20 (1.17)	1.21 (1.51)	0.99 (1.34)	0.94 (1.30)	0.99 (0.99)	0.99 (1.25)	1.06 (1.31)
7 am	1.10 (1.47)	1.35 (1.24)	1.23 (1.53)	1.01 (1.33)	0.78 (1.22)	1.07 (1.18)	1.09 (1.33)	1.09 (1.33)
8 am	1.31 (1.64)	1.56 (1.56)	1.49 (1.52)	1.02 (1.39)	0.97 (1.30)	1.28 (1.28)	1.24 (1.46)	1.27 (1.45)
9 am	1.66 (2.03)	1.72 (1.87)	1.75 (1.78)	1.14 (1.66)	1.03 (1.42)	1.39 (1.31)	1.27 (1.52)	1.42 (1.66)
10 am	1.67 (2.08)	1.77 (2.14)	1.63 (1.57)	1.05 (1.60)	0.94 (1.43)	1.20 (1.32)	1.16 (1.26)	1.35 (1.63)
11 am	1.58 (2.06)	1.73 (2.01)	1.40 (1.46)	1.08 (1.55)	0.95 (1.45)	1.14 (1.40)	1.09 (1.24)	1.28 (1.60)
12 pm	1.48 (1.89)	1.73 (2.02)	1.31 (1.52)	0.97 (1.53)	0.86 (1.45)	1.22 (1.52)	1.07 (1.26)	1.23 (1.60)
13 pm	1.40 (1.80)	1.53 (1.86)	1.32 (1.52)	1.06 (1.59)	0.94 (1.49)	1.23 (1.58)	1.24 (1.41)	1.25 (1.61)
14 pm	1.48 (1.77)	1.48 (1.78)	1.30 (1.57)	1.18 (1.58)	1.05 (1.52)	1.28 (1.60)	1.39 (1.50)	1.31 (1.62)
15 pm	1.50 (1.79)	1.35 (1.75)	1.34 (1.54)	1.22 (1.57)	1.13 (1.65)	1.22 (1.55)	1.41 (1.53)	1.31 (1.63)
16 pm	1.56 (1.91)	1.40 (1.95)	1.29 (1.60)	1.31 (1.63)	1.24 (1.76)	1.42 (1.74)	1.53 (1.77)	1.39 (1.77)
17 pm	1.63 (1.97)	1.52 (1.97)	1.33 (1.68)	1.45 (1.88)	1.36 (1.72)	1.54 (2.05)	1.67 (2.09)	1.50 (1.91)
18 pm	1.58 (1.82)	1.63 (1.81)	1.34 (1.70)	1.58 (2.03)	1.33 (1.65)	1.54 (1.95)	1.78 (2.32)	1.54 (1.90)
19 pm	1.55 (1.91)	1.74 (1.66)	1.32 (1.66)	1.63 (2.08)	1.31 (1.56)	1.55 (1.69)	1.79 (2.39)	1.56 (1.85)
20 pm	1.55 (1.84)	1.70 (1.67)	1.32 (1.68)	1.57 (1.89)	1.44 (1.55)	1.46 (1.53)	1.48 (2.03)	1.50 (1.74)
21 pm	1.62 (1.93)	1.56 (1.70)	1.19 (1.49)	1.40 (1.81)	1.46 (1.36)	1.27 (1.39)	1.26 (1.72)	1.39 (1.63)
22 pm	1.33 (1.66)	1.41 (1.57)	0.94 (1.22)	1.12 (1.47)	1.14 (1.09)	1.18 (1.16)	1.32 (1.60)	1.20 (1.40)
23 pm	1.13 (1.40)	1.31 (1.59)	0.98 (1.21)	1.22 (1.62)	1.13 (1.21)	1.08 (1.15)	1.41 (1.57)	1.18 (1.39)
24 pm	1.38 (1.80)	1.27 (1.61)	1.07 (1.44)	1.35 (1.67)	1.24 (1.24)	1.16 (1.35)	1.55 (1.73)	1.29 (1.55)
All hours	1.40 (1.79)	1.42 (1.62)	1.25 (1.55)	1.16 (1.58)	1.07 (1.42)	1.21 (1.38)	1.25 (1.56)	1.25 (1.56)

Table 3. Errors for one-day-ahead hourly forecasts for the year 2006, compared with Cancelo et al. (2008).

Estimating changes in demand for holidays is very complex, and correct modeling is important when comparing methods. The model incorporates a large number of dummy variables that consider all peculiarities of the problem. Subsequently, the number of parameters is reduced using the usual regression model tests. The joint representation of the 24-hour effects for a day shows consistent and interesting results.

The model has been adapted (with the supervision of the Spanish TSO, REE) to make real-time predictions for 12 smaller electrical systems in Spain (e.g., Balearic Islands, Canary Islands, Ceuta and Melilla). The results indicate that the model can be used in other systems, with very satisfactory results.

Future work will focus on an enhanced temperature model, an alternative implementation for nonworking days, and the inclusion of other weather variables (solar radiation, wind speed, etc.).

Acknowledgements

The Authors would like to thank to Jesús Rupérez, Pablo Martín, Carlos Rodríguez, Ana Rodríguez, Juanjo Abellán and Iago Veiras, for their help and technical support. This work has been funded by Red Eléctrica de España (REE) as a R&D project.

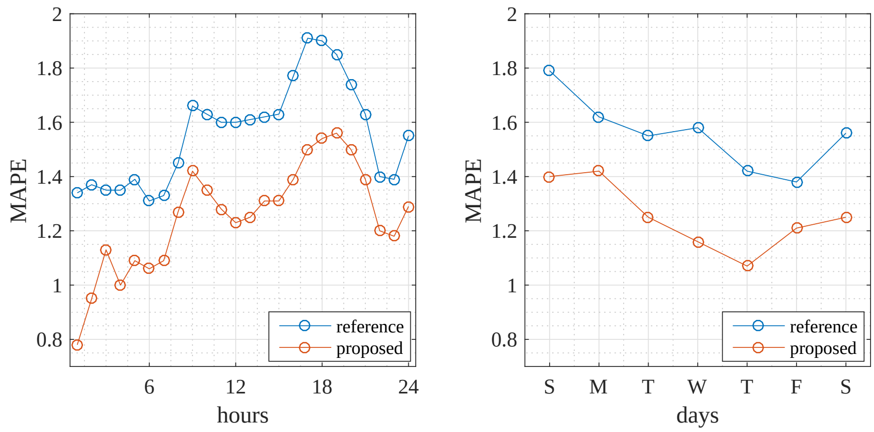

Figure 5. Hourly and daily errors for the reference [9] and proposed methods.

References

[1] J. Cancelo, A. Espasa, Forecasting daily demand for electricity with multiple-input nonlinear transfer function models: a case study, Tech. rep., Carlos III University (1991).

[2] D. Bunn, E. Farmer, Comparative models for electrical load forecasting., New York: Wiley, 1985.

[3] J. W. Taylor, L. M. de Menezes, P. E. McSharry, A comparison of univariate methods for forecasting electricity demand up to a day ahead, International Journal of Forecasting 22 (1) (2006) 1 – 16.

[4] J. W. Taylor, Triple seasonal methods for short-term electricity demand forecasting, European Journal of Operational Research 204 (1) (2010) 139 – 152.

[5] S. J. Koopman, M. Ooms, M. A. Carnero, Periodic seasonal reg-ARFIMA-GARCH models for daily electricity spot prices, Journal of the American Statistical Association 102 (477) (2007) 16–27.

[6] A. Clements, A. Hurn, Z. Li, Forecasting day-ahead electricity load using a multiple equation time series approach, European Journal of Operational Research 251 (2) (2016) 522 – 530.

[7] R. Ramanathan, R. Engle, C. W. Granger, F. Vahid-Araghi, C. Brace, Short-run forecasts of electricity loads and peaks, International Journal of Forecasting 13 (2) (1997) 161 – 174.

[8] R. Cottet, M. Smith, Bayesian modeling and forecasting of intraday electricity load, Journal of the American Statistical Association 98 (464) (2003) 839–849.

[9] J. R. Cancelo, A. Espasa, R. Grafe, Forecasting the electricity load from one day to one week ahead for the Spanish system operator, International Journal of Forecasting 24 (4) (2008) 588 – 602, Energy Forecasting.

[10] V. Dordonnat, S. Koopman, M. Ooms, A. Dessertaine, J. Collet, An hourly periodic state space model for modelling French national electricity load, International Journal of Forecasting 24 (4) (2008) 566 – 587, Energy Forecasting.

[11] R. F. Engle, C. W. J. Granger, J. Rice, A. Weiss, Semiparametric estimates of the relation between weather and electricity sales, Journal of the American Statistical Association 81 (1986) 310–320.

[12] N. Charlton, C. Singleton, A refined parametric model for short term load forecasting, International Journal of Forecasting 30 (2) (2014) 364—368.

[13] A. Pierrot, Y. Goude, Short-term electricity load forecasting with generalized additive models, Proceedings of ISAP power (2011) 410—416.

[14] S. Fan, R. J. Hyndman, Short-term load forecasting based on a semi-parametric additive model, IEEE Transactions on Power Systems, 27 (2012) 134—141.

[15] P. Gaillard, Y. Goude, R. Nedellec, Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting, International Journal of Forecasting 32 (3) (2016) 1038–1050.

[16] F. Ziel, Modeling public holidays in load forecasting: a German case study, Journal of Modern Power Systems and Clean Energy 6 (2) (2018) 191–207.

[17] J. Nowicka-Zagrajek, R. Weron, Modeling electricity loads in California: ARMA models with hyperbolic noise, Signal Processing 82 (2002) 1903–1915.

[18] M. Smith, Modeling and short-term forecasting of New South Wales electricity system load, Journal of Business & Economic Statistics 18 (4) (2000) 465–478.

[19] M. López, S. Valero, A. Rodriguez, I. Veiras, C. Senabre, New online load forecasting system for the Spanish transport system operator, Electric Power Systems Research 154 (2018) 401–412.

[20] A. Azadeh, S. Ghaderi, S. Sohrabkhani, Forecasting electrical consumption by integration of neural network, time series and ANOVA, Applied Mathematics and Computation 186 (2) (2007) 1753–1761.

[21] A. Azadeh, S. Ghaderi, S. Tarverdian, M. Saberi, Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption, Applied Mathematics and Computation 186 (2) (2007) 1731–1741.

[22] G. A. Darbellay, M. Slama, Forecasting the short-term demand for electricity: Do neural networks stand a better chance?, International Journal of Forecasting 16 (1) (2000) 71 – 83.

[23] H. Hahn, S. Meyer-Nieberg, S. Pickl, Electric load forecasting methods: Tools for decision making, European Journal of Operational Research 199 (3) (2009) 902 – 907.

[24] E. G. Gladyshev, Periodically correlated random sequences, Soviet Mathematics (1961) 385 – 388.

[25] M. Pagano, On periodic and multiple autoregressions, The Annals of Statistics 6 (6) (1978) 1310–1317.

[26] G. C. Tiao, M. R. Grupe, Hidden periodic autoregressive-moving average models in time series data, Biometrika 67 (2) (1980) 365–373.

[27] T. Hastie, R. Tibshirani, Generalized Additive Models, Chapman and Hall/CRC, 1990.

[28] G. Wahba, Spline models for observational data, SIAM: Society for Industrial and Applied Mathematics, 1990.

[29] S. N. Wood, Generalized Additive Models: an introduction with R, Chapman and Hall/CRC, 2006.

[30] G. Box, G. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day. (1970)

[31] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov, F. Csaki, Proceedings of the 2nd International Symposium on Information Theory (1973) 267–281.

[32] G, Schwarz, Estimating the Dimension of a Model, Annals of Statistics, 6 (1978) 461–464.

[33] P. H. Franses, A note on the mean absolute scaled error, International Journal of Forecasting 32 (1) (2016) 20–22.

[34] F. Ziel, B. Liu, Lasso estimation for GEFCom2014 probabilistic electric load forecasting, International Journal of Forecasting 32 (3) (2016) 1029–1037.