"Electricity load modeling: an application to Italian market"

AUTHORS	Giovanni Masala https://orcid.org/0000-0003-1719-641X Stefania Marica
ARTICLE INFO	Giovanni Masala and Stefania Marica (2015). Electricity load modeling: an application to Italian market. *Investment Management and Financial Innovations*, *12*(1), 35-46
RELEASED ON	Thursday, 26 February 2015
JOURNAL	"Investment Management and Financial Innovations"
FOUNDER	LLC "Consulting Publishing Company "Business Perspectives"

Giovanni Masala (Italy), Stefania Marica (Italy)

Electricity load modeling: an application to Italian market

Abstract

Forecasting electricity load plays a crucial role regarding decision making and planning for economical purposes. Besides, in light of the recent privatization and deregulation of the power industry, the forecasting of future electricity load turned out to be a very challenging problem. Electricity demand is also an important determinant for electricity spot prices which, in turn, influence the pricing of futures contracts in electricity markets.

Empirical data about electricity loads highlight a clear seasonal behavior (higher loads during the winter season), which is partly due to climatic effects. The authors also emphasize the presence of load periodicity at a weekly basis (electricity load is usually lower on weekends or holidays) and at a daily basis (electricity load is clearly influenced by the hour). Finally, a long-term trend may depend on the general economic situation (for example, industrial production affects electricity load). All these features must be captured by the model.

The purpose of this paper is then to build an hourly electricity load model. The deterministic component of the model requires linear regression while the authors investigate the stochastic component through classical time series tools.

The calibration of the model is performed by using data coming from the Italian market in a 5 years period (2007-2011). Then, the researchers perform a Monte Carlo simulation in order to compare the simulated values with respect to the real data (both in sample and out of sample inspection). The reliability of the model will be deduced using standard tests which highlight a good fit of the simulated values.

Introduction

The energy transaction operations have changed dramatically since the last decade of the 20th century due to the liberalization of the main power markets. Presently, the electricity business is negotiated in particular electricity markets and particular over-the-counter markets. A striking characteristic of these markets is that traded volumes represent energy which will be used and produced only in the future. As a consequence, a thorough forecasting of load demand and prices is a challenging problem. Indeed, electricity shares a specific property respect to other commodities, namely its demand and supply must be in balance at each time. Additionally, load forecast plays also a crucial role in price determination.

The accuracy of electricity load forecasting has been intensively investigated over the past few years. Indeed, an inaccurate electricity load forecast causes an increase in the cost of operations. For example, an overestimation causes a supply excess while underestimation causes insufficient electricity supply.

Regarding the Italian market, which is the object of this paper, we note that Italy is not self-sufficient concerning energy. At this purpose, the National Energy Agency report asserts that Italy depends on foreign suppliers for about 85 percent of its needs. For example, about 15% of electricity consumption is imported from abroad. Besides, the Italian electricity sector has been recently restructured in order to follow EC Directive 96/92, which aims to set up a single electricity market. Besides, from a political point of view, the prominent Legislative Decree No. 79 of March 16, 1999 (the Bersani Decree) was implemented to liberalize the activities of electricity production, import, export, purchases and sales. The Decree also sets up an antitrust ceiling on the business of the dominant operator in order to promote competition.

Several models have been introduced in recent literature with the purpose of precisely modeling electricity loads. More accurately, time series modeling approaches based on artificial neural networks (ANNs) and statistical methods were used. Bilgili et al. ([6], 2012) apply the artificial neural network (ANN) methodology to forecast Turkey's residential and industrial electricity consumption, analyze energy use and perform future projections for the period of 2008-2015. According to the ANN model, Turkey's residential and industrial sector electricity consumption will increase by 2015. They find that the performance values of the ANN method are better than the performance values of the linear regression (LR) and nonlinear regression (NLR) models.

Deihimi et al. ([10], 2012) use a wavelet echo state network (WESN) to forecast short term load and temperature. They demonstrate that WESN improve the accuracy of both load and temperature short term forecast compared to wavelet neural network (WNN) model.

Nagi et al. ([20], 2008) forecast the electricity demand using a hybrid artificial intelligence scheme

based on self-organizing maps (SOMs) and support vector machines (SVMs). The results show that this approach gives good prediction accuracy for mid-term electricity load forecasting.

Statistical models include moving average and exponential smoothing methods such as multi-linear regression models, stochastic process, data mining approaches, autoregressive and moving averages (ARMA) models, GARCH models, Box-Jenkins methods and Kalman filtering-based methods. These techniques provide forecasting models of different accuracy. The accuracy of the prediction depends on the minimum error of the forecast. The appropriate prediction methods are considered from several factors such as prediction interval, prediction period, characteristics of the time series, and size of the time series (Makridakis et al., 1998, [17]).

Chujai et al. ([9], 2013) research was to find a model to efficiently forecast the electricity consumption in a household by applying the Box and Jenkins method. The results show that the ARIMA model was the best for finding the most suitable forecasting period in monthly and quarterly. On the other hand, ARMA was the best model for finding the most suitable forecasting period in daily and weekly. For example, Weron ([26], 2006) lists in his book some general techniques for load electricity and spot prices modeling.

Saab et al. ([23], 2001) use three different univariate models to forecast the monthly electric energy consumption in Lebanon, during the period of 1990-1999. Different measures used by the authors show that the AR(1) highpass filter model was the best forecasting model for the electrical energy Lebanese data set.

Migon and Alves ([18], 2013) use a multivariate dynamic regression to forecast electricity consumption in the Brazilian southeastern submarket for one-day ahead. They compare the results from a univariate dynamic regression model (including the seasonal, daily and weekly, effects only), and a multivariate model (with dummy weekday type variables only). They conclude that the first one performs better than the second one, although the difference in probability is not large. Moral-Carcedo and Vicéns-Otero ([19], 2005) analyze the effect of temperatures on the variability of the Spanish daily electricity demand, using different non-linear regression models.

Andersson et al. ([3], 2011) propose an Hourly Price Forward Curve (HPFC) based on the median estimation to evaluate the hourly, daily and yearly energy price profiles. The authors show that the results obtained via this approach are significantly better than these obtained with the mean value.

Alter and Syed ([2], 2011) analyze the determinants of electricity demand in Pakistan during the period of 1970-2010, using a cointegration and vector error correction approach. They find the existence of a long run relation among electricity demand and its determinants in aggregate, residential, industrial, commercial and agricultural sectors.

Generally, models and forecasts on energy load are considered at three different levels of time horizon: short-, medium-, and long-term using different frequency of the data. In the short- and medium-term, energy demand are considered in hourly, weekly or monthly interval range, whereas yearly load is normally performed on a yearly average basis.

Blàzquez et al. ([7], 2012) examine the residential electricity demand paying particular attention to the influence of price, income, and weather conditions using a panel data approach for 47 Spanish provinces. The result shows the higher sensitivity of electricity demand to cold than to hot days, because Spanish households use gas heating systems more than electric heating systems, and only a small fraction of them use air conditioning.

Bianco et al. ([5], 2009) estimate the elasticities of Italian domestic and non-domestic electricity consumption on GDP, GDP per capita and price. Using annual series from 1970 to 2007, the authors find that variations in GDP and GDP per capita explain quite well domestic and non-domestic electricity consumption. Furthermore, they find that the electricity price is an irrelevant variable in forecasting models for the case of Italian electricity consumption.

Filik et al. ([13], 2011) propose a nested methodology able to make short, medium, and long-term hourly load forecast within a single framework. The authors show the accuracy of the model using hourly actual Turkish load demand values.

Owing to the importance of load forecasting, various models have been proposed for the short-term load forecasting, applied to intervals ranging from one hour to one week. Furthermore, different approaches are applied to deal with the daily, weekly and annual seasonality problem.

Hong ([15], 2011) to deal with seasonality estimates an electric load forecasting model applying the support vector regression (SVR) approach with chaotic artificial bee colony algorithm. The results show that this model performs better than ARIMA and TF-ε-SVR-SA models.

Afshar and Bigdeli ([1], 2011) forecast the short-term load electricity in Iran using a spectral analysis approach (SSA). They show that this method has better prediction ability than SSA-AR and SSA-LoLiMot methods.

Wang et al. ([25], 2013) use a decomposition approach to model different levels of electricity demand in South East Queensland (Australia), deriving from distinct seasonal climate. This method is relatively easy to implement and avoids the complexity of non-linear estimations.

Pielow et al. ([22], 2012) modeled short- and long-term electricity demand in commercial and industrial sectors of the United States applying a percentage difference autoregressive approach. They describe daily, weekly and monthly calendar variables by Fourier series with two frequencies at each time scale, reducing the number of predictors.

Pardo et al. ([21], 2002) use an autoregressive least-squares regression to explore the effects of temperature and seasonality on daily Spanish electricity load. Soares and Medeiros ([24], 2008) model the electricity hourly load of southeast of Brazil using rigorous statistical methods. They model daily and weekly seasonality with different dummy variables, and annual cycle with Fourier decomposition where the number of trigonometric functions is determined by the Bayesian Information Criterion (BIC).

Bruhns et al. ([8], 2005) compare a non-parametric model with local regression (LOESS) with an alternative model combining two Fourier series, one with dependency on the hour and one with dependency on day-type in order to deal with the modification of the daily load shape throughout the year. The results show that this model is the best, even if it requires great care in the day-type typology.

Gonzàles-Romera et al. ([14], 2008) investigate the behavior of Spanish monthly electric demand using a hybrid approach. They forecast the periodic component and the trend in a different way. The former one is predicted with Fourier series whereas the latter with a neural network.

Fan and Hyndman ([12], 2012) propose an additive model with nonlinear and non-parametric terms to forecast the short-run electricity load. They assumed the time of year effect, temperature effects and lagged demand effects to be smooth functions and estimate them as a cubic regression spline.

Kavousian et al. ([16], 2013) investigate the determinants of residential electricity consumption, by developing separate models for daily (maximum) peak and minimum (idle) consumption. They found that daily minimum consumption has a lower variation compared with daily maximum level, and is best explained by invariant time factors, such as degree days, zip code, house size, and number of refrigerators. Alternatively, electric water heaters and air conditioners better explain the maximum level consumption.

Dordonnat et al. ([11], 2008) present a model for hourly electricity load forecasting based on stochastically time varying processes. They model the short-run French electricity load with trends, seasons, weather and heating effects focusing on two hours, 9 AM and 12 PM in 9 years. The empirical evidence also highlight that the forecasting function depends strongly on the hour of the day.

In their survey, Andersson et al. ([4], 2013) examine the German electricity market. Their model captures the deterministic component as well as stochastic variations of electricity load. The load data is then decomposed into daily and hourly level in order to get more accurate forecasting. Then, each part is modeled separately. A linear regression approach was used to model the deterministic component, and an autoregressive model was used for modeling the stochastic component.

In the existing literature, there is little evidence about the electricity load in Italy. The aim of the present research is to fill this gap following the model of Andersson et al. ([4], 2013). The novelty of this study is to model the electricity load not only as a function of calendar and meteorological data, but also as a function of economic variables such as industrial production and number of trips. Furthermore, to deal with the cyclical patterns of data, we estimate the deterministic component with a linear multiple regression model. Additionally, we investigate thoroughly the stochastic components by also analyzing heteroscedasticity effects.

The paper is organized as follows. In the Introduction we have presented the objectives of our paper and we have discussed various state of the art empirical approaches. Section 1 is devoted to the main characteristics and seasonal features of the electricity consumption database. The theoretical model is described in Section 2. Section 3 is devoted to the empirical application and to the comparison with the real data. Final section concludes and lists some future enhancements.

1. Data description

The database containing load hourly data can be freely retrieved from the Italian market operator's website (http://www.mercatoelettrico.org/En/Default.aspx). We selected data from 2007 to 2011. On the whole we have 43,824 hourly records (unit load is MWh) and 1,826 daily aggregate records. Note that daily values are obtained by considering the mean hourly values for this day.

A first graphical inspection permits to deduce some features of the hourly data. At this purpose, we show in Figure 1 the hourly values for the first week of June 2010 (starting with Tuesday, note that June, 2 is a holiday) while Figure 2 exhibits the values for all of June 2010.
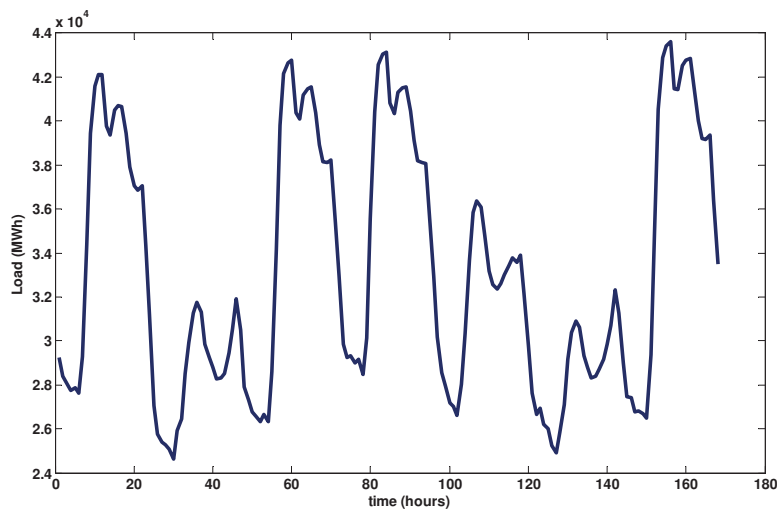
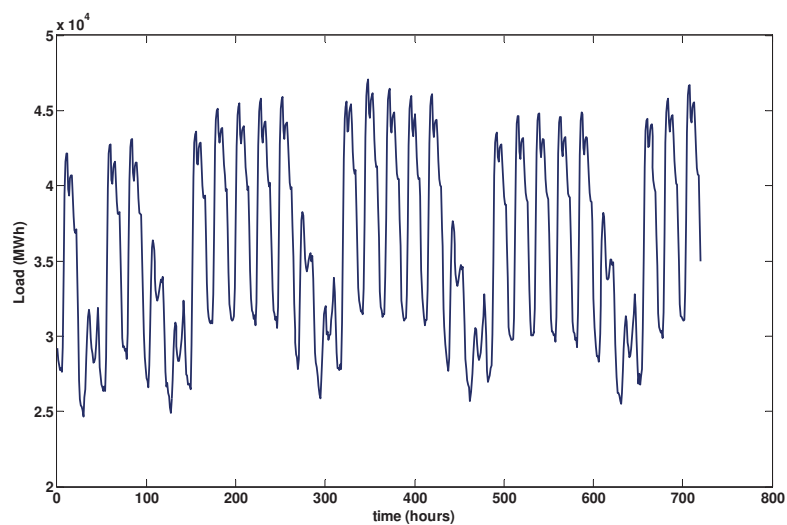**Fig. 1. Hourly values (MWh) 1-7 June 2010**

Fig. 2. Hourly values (MWh) June 2010

From the two plots we highlight some cyclical aspects. The intra-day values present a usual higher peak at about 11 a.m. and a usual lower peak at about 4 a.m. When considering a longer horizon, we note a cyclical behavior with respect to the type of days. We have essentially three types of days: working days, Saturdays and Sundays. Nevertheless, from a more careful analysis of the database, we point out that some special days (which we call semi-holidays) must be considered. For example, Fridays following a holiday or Mondays before a holiday. We also estimate the mean load values for each hour (for the whole database). We show the results in Figure 3.
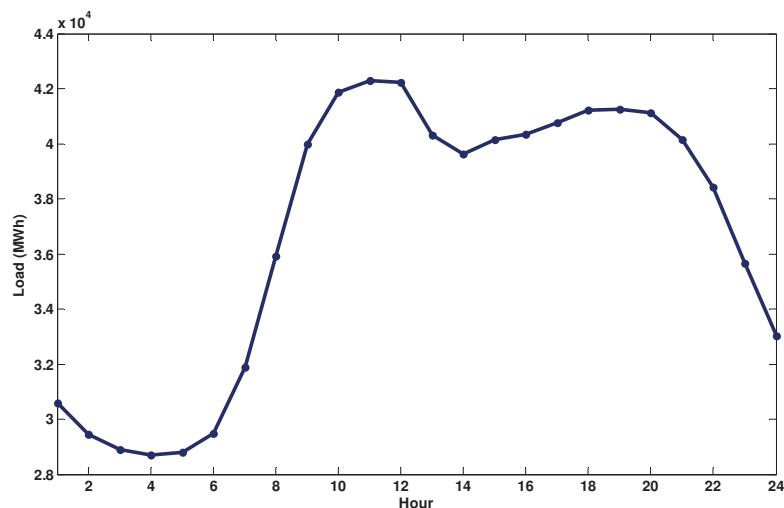

Fig. 3. Hourly mean load (Load vs. hour)

The higher and lower peaks with respect to the hour are now more evident. During the central hours of the day the electricity load is higher than during the night, reflecting patterns in human activities.

We then estimate the mean load values for each day of the week and for the whole database (we did not take into account holidays or other special days). It appears that electricity load on Saturdays and Sundays represent the lowest values while the five working days are rather equivalent (except for Monday which experiences a slightly lower value).

Next, we estimate the mean load values for each month (for the whole database). We show the results in Figure 4.
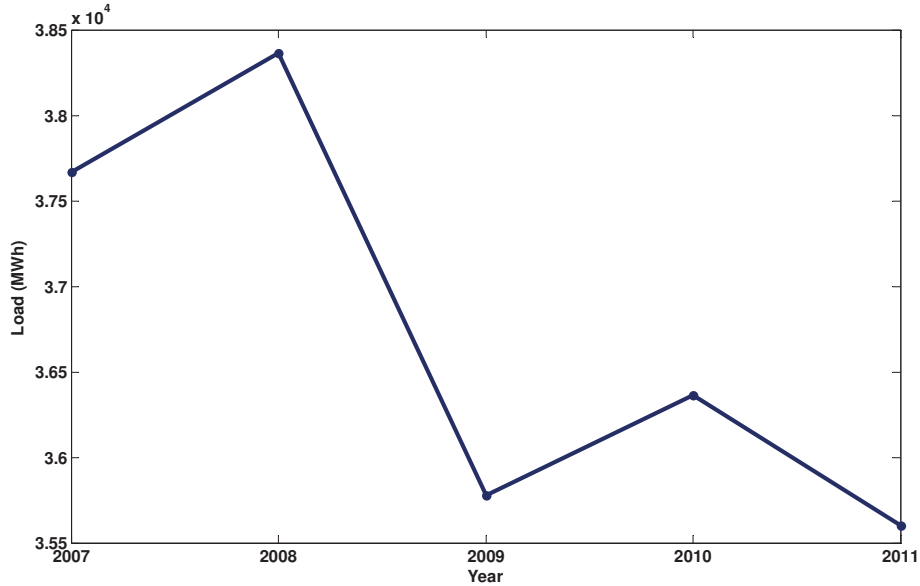

Fig. 4. Yearly values (MWh) 2007-2011

We deduce from Figure 4 that the yearly load has been increasing in the period of 2007-2008 and then the trend inverted (except for year 2010). This behavior has to be explained in the light of economic indicators such as the Industrial Production Index (further explained in section 3).

Furthermore, Table 1 shows the Augmented Dickey Fuller (ADF) unit root tests of the variables (dependent and economic variables) used to estimate the daily component.

Tab. 1 Unit root tests (ADF)

Variable	Test value (p value)
daily load	-7.8466 (0.000)
trips	-3.5638 (0.0330)

This table indicates the value of the ADF test. From the results, we can see that the ADF test for daily load and trips rejects the null hypothesis that the variables are non-stationary at a 95% confidence level, so these variables are taken to be stationary.

Finally, a spectral (or harmonic) analysis permits to highlight cyclical patterns of data. The purpose of this analysis is to decompose a time series with cyclical components into a few underlying sinusoidal functions of particular wavelengths (see Weron [26], 2006). Recall that the periodogram associated with a vector of observations $\{x_1,\ldots,x_n\}$ is given by:

$$I_n\left(\omega_k\right) = \frac{1}{n}\left|\sum_{t=1}^{n} x_t \cdot e^{-i(t-1)\cdot\omega_k}\right|^2, \tag{1}$$

where $\omega_k = 2\pi\cdot\left(k/n\right)$ denotes the Fourier frequencies (rad/unit time). We illustrate in the next Figure 5 the periodogram of daily and hourly data.
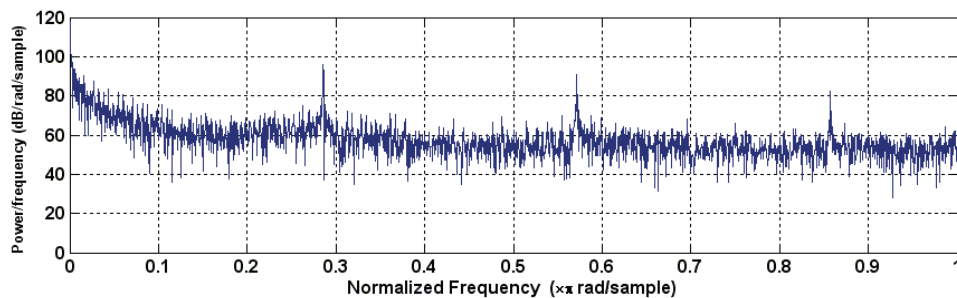
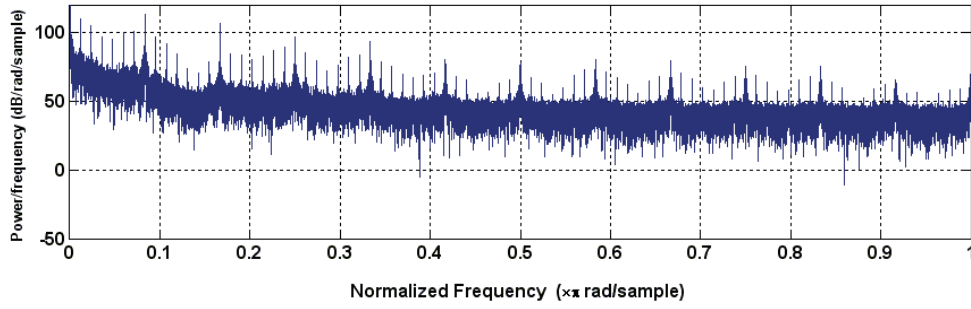

Fig. 5a. Periodogram daily data

Fig. 5b. Periodogram hourly data

The daily plot presents a peak at frequency $\omega_k =$ 0.2856 (associated to $T = 7$ days period). The smaller peaks ($\omega_k = 4/7$ and $\omega_k = 6/7$) are the harmonics multiples. The hourly data present peaks associated to 24 hours and 168 hours and their harmonics multiples.

2. The theoretical model

We denote $L(t)$ the electricity load (with t expressed in hours, accordingly with the database feature). We set up the following decomposition:

$$L(t) = f(t) + \tilde{x}(t), \qquad (2)$$

where $f(t)$ is the deterministic component and $\tilde{x}(t)$ is the stochastic one.

Alongside we can decompose the load level between the hourly component L_h and the daily component L_d:

$$L(t) = L_d(t) + L_h(t). \qquad (3)$$

Remark. In order to examine the daily characteristics, the database has to be adapted by computing the average hourly loads for each day. The daily component assumes then the same value of each hour t associated to the same day.

The daily component.

We can decompose it as the sum of the deterministic part and the stochastic part as follows:

$$L_d(t) = f_d(t) + \tilde{x}_d(t). \qquad (4)$$

The variable t in equation (4) is expressed in days. We can model the deterministic component through multiple linear regression. The regressors can be divided into the following categories:

♦ calendar variables (days of week, months, holidays, semi-holidays). These are dummy variables;

♦ economic variables (industrial production index, trips);

♦ non economic variables (temperature index).

We get then:

$$f_d(t) = \alpha_d + \sum_{i=1}^{7} \beta_{di} \cdot D_i(t) + \sum_{i=1}^{12} \gamma_{di} \cdot M_i(t) +$$
$$+ \delta_d \cdot H(t) + \varsigma_d \cdot SH(t) + \qquad (5)$$
$$+ \lambda_{d1} \cdot IP(t) + \lambda_{d2} \cdot Tr(t) + \lambda_{d3} \cdot T(t).$$

We choose the following regressors: *IP* (industrial production), *T* (temperature index), *Tr* (trips), *M* (month), *H* (holiday), *SH* (semi-holiday), *D* (week day).

The data concerning economic variables (on a monthly or quarterly basis) are freely available at the following link: http://www.istat.it/en/. We show in Figure 6 the yearly industrial production for the period 2007-2011. Note the similarities with Figure 5.
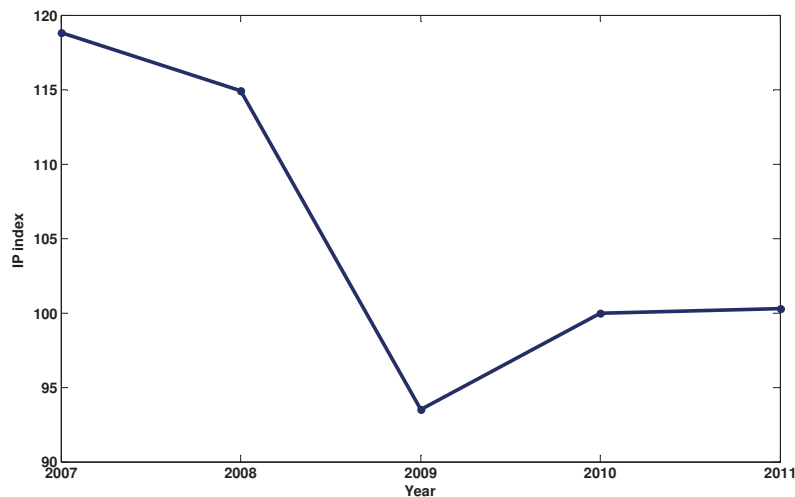

Fig. 6. Yearly industrial production index period of 2007-2011

The close binding between load values and the IP index can be emphasized through a correlation analysis. At this purpose, the correlation with respect to monthly values is 73% and the correlation with respect to yearly values is 90%. We deduce that the industrial production index is a good regressor, which can explain the yearly trend of load values.

The database regarding daily temperatures can be obtained from the Mathematica 9.0 software (weather data). In order to produce a reliable temperature index, we performed the mean value of the ten biggest Italian cities' temperatures. We represent in the next Figure 7 the load values respect to temperature.
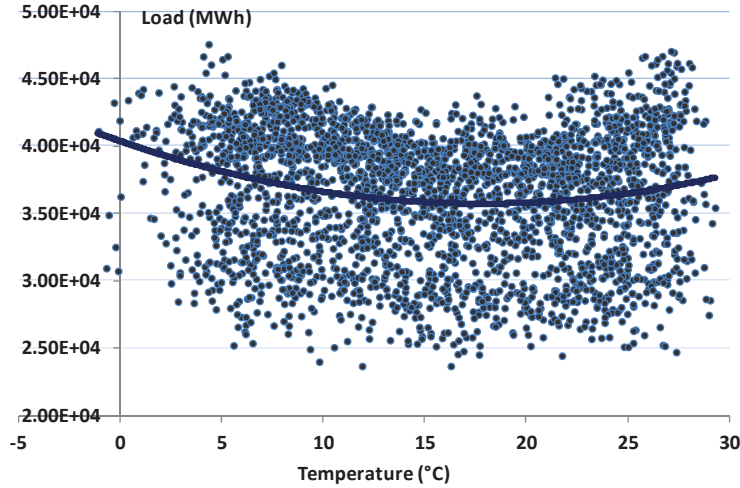

Fig. 7. Load values (MWh) vs. temperature (°C)

Nevertheless, the temperature's values unveil a correlation of about -9% with respect to daily electricity load. In order to overtake this problem, we have constructed a new temperature index based on *CDD* and *HDD* values as follows:

$$\overline{T} = \begin{cases} T - 20 & \text{if} \quad T > 20 \\ 0 & \text{if} \quad 16 < T < 20. \\ 16 - T & \text{if} \quad T < 16 \end{cases} \tag{6}$$

This new indicator has a correlation of about +20% with respect to daily electricity load. Indeed, heating and cooling systems start when temperature reaches these thresholds.

The residuals of this regression represent the stochastic part. We propose for this component an Auto-regressive-GARCH process $AR(p) - GARCH(1,1)$.

$$\tilde{x}_d(t) = c + \sum_{i=1}^{p} \delta_i \cdot \tilde{x}_d\left(t - i \cdot \Delta t\right) + \tilde{\varepsilon}(t) \text{ with}$$

$$\tilde{\varepsilon}(t) = \eta(t) \cdot \sqrt{h(t)}, \quad \eta(t) \stackrel{i.i.d.}{\sim} N(0,1) \tag{7}$$

$$h(t) = \omega + \alpha \cdot \tilde{\varepsilon}^2(t-1) + \beta \cdot h(t-1),$$

in order to take into account the autocorrelation and heteroscedasticity effects in the residuals.

2.1. The hourly component. We can deduce the hourly component from (3): (*t* expressed in hours)

$$L_h(t) = L(t) - L_d(t). \tag{8}$$

We have again the deterministic and the stochastic part:

$$L_h(t) = f_h(t) + \tilde{x}_h(t). \tag{9}$$

The deterministic component can be modelled through a multiple linear regression as follows:

$$f_h(t) = \sum_{i=1}^{7} \sum_{j=1}^{12} \left[d_{ij} + \sum_{k=1}^{24} \xi_{ijk}(t) \cdot H_{ijk}(t) \right], \tag{10}$$

where the regressor H denotes the hour (dummy variable). The first sum with index i is the type of day (working days, Saturday and semi-holidays or Sunday and holidays), the second sum with index j denotes the month, and the last one with index k denotes the hour.

The residuals of this regression are the stochastic component. We propose here again an Autoregressive-GARCH(1,1) process as before for the same reasons:

$$\tilde{x}_h(t) = \sum_{i=1}^{7} \sum_{j=1}^{12} \left[c_{h,ij} + \sum_{k=1}^{p} \delta_{h,ijk} \cdot \tilde{x}_{h,ij}(t - k \cdot \Delta t) + \tilde{\varepsilon}_{h,ij}(t) \right] \text{ with}$$

$$\tilde{\varepsilon}_{h,ij}(t) = \eta_{h,ij}(t) \cdot \sqrt{h_{h,ij}(t)}, \quad \eta_{h,ij}(t) \stackrel{i.i.d.}{\sim} N(0,1) \tag{11}$$

$$h_{h,ij}(t) = \omega_{h,ij} + \alpha_{h,ij} \cdot \tilde{\varepsilon}_{h,ij}^2(t-1) + \beta_{h,ij} \cdot h_{h,ij}(t-1).$$

2.2. Overall model. We get the final model by aggregating all the components described before: (t expressed in hours):

$$\tilde{L}(t) = \alpha_d + \sum_{i=1}^{7} \beta_{di} \cdot D_i(t) + \sum_{i=1}^{12} \gamma_{di} \cdot M_i(t) + \delta_d \cdot H(t) + \varsigma_d \cdot SH(t) +$$

$$+ \lambda_{d1} \cdot IP(t) + \lambda_{d2} \cdot Tr(t) + \lambda_{d3} \cdot T(t) + c + \sum_{i=1}^{p} \delta_i \cdot \tilde{x}_d(t - i \cdot \Delta t) + \sigma_d \cdot \tilde{\varepsilon}(t) +$$

$$+ \sum_{i=1}^{7} \sum_{j=1}^{12} \left[d_{ij} + \sum_{k=1}^{24} \xi_{ijk}(t) \cdot H_{ijk}(t) \right] + \sum_{i=1}^{7} \sum_{j=1}^{12} \left[c_{h,ij} + \sum_{k=1}^{p} \delta_{h,ijk} \cdot \tilde{x}_{h,ij}(t - k \cdot \Delta t) + \sigma_{h,ij} \cdot \tilde{\varepsilon}_{h,ij}(t) \right]$$

(12)

with

$$\tilde{\varepsilon}(t) = \eta(t) \cdot \sqrt{h(t)}, \quad \eta(t) \overset{i.i.d.}{\sim} N(0,1)$$

$$h(t) = \omega + \alpha \cdot \tilde{\varepsilon}^2(t-1) + \beta \cdot h(t-1)$$

$$\tilde{\varepsilon}_{h,ij}(t) = \eta_{h,ij}(t) \cdot \sqrt{h_{h,ij}(t)}, \quad \eta_{h,ij}(t) \overset{i.i.d.}{\sim} N(0,1)$$

$$h_{h,ij}(t) = \omega_{h,ij} + \alpha_{h,ij} \cdot \tilde{\varepsilon}^2_{h,ij}(t-1) + \beta_{h,ij} \cdot h_{h,ij}(t-1).$$

3. Empirical application

In this section, we describe the estimation of the parameters of the model and then we present the evaluation of its performance in the calibrating window (2007-2010) for the out of sample estimation concerning the year 2011. Lastly, we perform an in sample estimation concerning a shorter two-year time horizon (2010-2011).

3.1. Model parameters. In this subsection we examine separately the four steps of the model (namely the deterministic and stochastic part of the daily and hourly component respectively).

3.1.1. Deterministic part of the daily component. This part is given by a linear regression with 24 regressors as already discussed in the previous section 2.

The coefficients of some striking variables are given in the Table 2 (the number of observations is 1,461 and the adjusted R-squared is 93.40%, finally the F-statistic vs. the constant model is 855 with zero *p*-value).

Table 2. Parameters estimation

Variable	Coefficient	*p* value
Temperature	209.14***	0.000
Industrial production	79.11***	0.000
Trips	0.32***	0.000
Monday	5.14***	0.000
Sunday	-4.09***	0.000
Holidays	-5.25***	0.000

Note: ***Statistically significant at 1% level.

We note that these variables have the expected signs and they are all statistically significant. It means that calendar, economic and non-economic variables matter to explain electricity load behavior. The temperature index coefficient is positive. It means that when temperature index moves away from the given threshold value (16/20 °C), the electricity load also increases, related to using cooling and heating systems. The same positive relationship is found for industrial production, which absorbs more energy and for trips.

The estimation results also confirm the trend depicted in Figure 1: during working days electricity load is higher (positive coefficients, for example Monday) than those in Saturday and Sunday (negative coefficients). Furthermore, holidays' days are like weekends' days, negatively related to the electricity load. We omit the description of the other parameters for sake of brevity (they are available upon request).

3.1.2. Stochastic part of the daily component. This component is given by the residuals of the previous regression. In order to examine the characteristics of the residuals, we plot the autocorrelation function and the partial autocorrelation function (Figure 8).
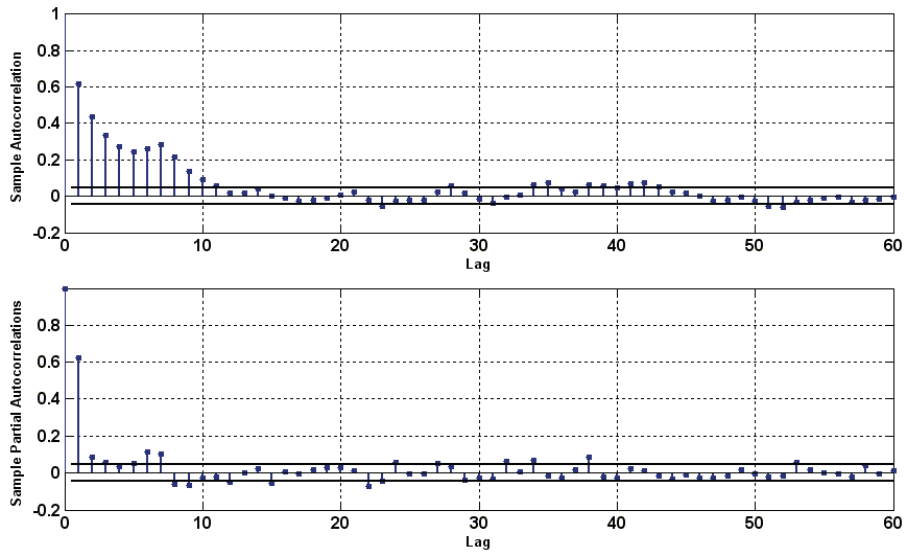
Fig. 8. ACF and PACF for daily residuals

These plots show the presence of autocorrelation. Besides, the Engle's test detects the presence of residual heteroscedasticity. We propose then to model the residuals through an AR(7)-GARCH(1,1) process.

The new residuals emphasize no more autocorrelation (we omit the associated plots) so that seven days lags are sufficient to remove autocorrelation.

Then, we also exhibit in the following Table 3 the parameters of the process with the associated test statistics (the lags 2, 3, 4 and 5 are not statistically significant).

Table 3. ARMA-GARCH parameters

Parameter	Value	*t* Statistic
Constant	-0.6736	-0.0421
AR(1)	0.6515***	37.2394
AR(6)	0.1197***	6.5151
AR(7)	0.0980***	5.1084
Constant	277107***	5.0176
GARCH(1)	0.2912***	4.3052
ARCH(1)	0.6349***	4.7766
DoF	3.4450***	8.6996

Note: ***Statistically significant at 1% level.

3.1.3. Deterministic part of the hourly component. This component is given by a multi-linear regression over the hourly component with 24 regressors as already discussed in the previous section 2. The regression is performed after fixing the type of day and month. For sake of brevity, we do not exhibit the results.

3.1.4. Stochastic part of the daily component. This component is given by the residuals of the previous regressions for each type of day and each month. In order to examine the characteristics of the residuals, we have examined the autocorrelation function and the partial autocorrelation function.

The Engle's test detects again the presence of residual heteroscedasticity.

We propose to model these residuals through an AR(2)-GARCH(1,1) process. The autocorrelation function and the partial autocorrelation function of the residuals reveal now absence of residual autocorrelation.

We omit again the results.

We finally compare in the following Figure 9 the simulated load values (dotted line) with respect to the empirical ones (solid line) at hourly level for the period 14-20 February, 2011.
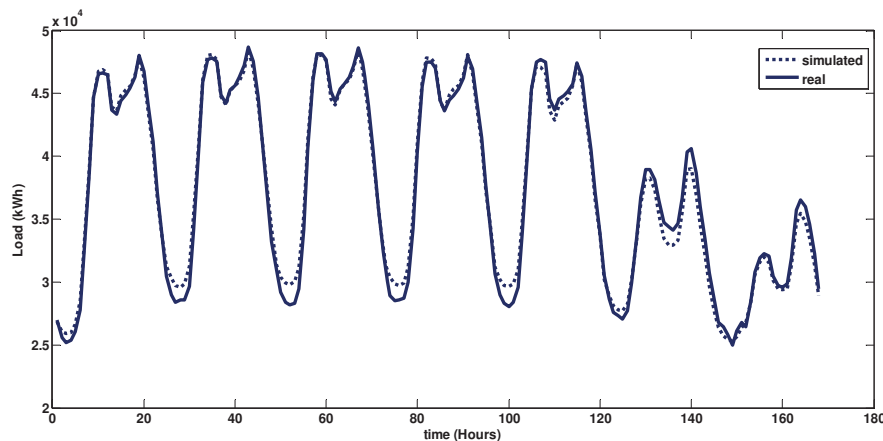

Fig. 9. Real load data vs. empirical ones

Figure 9 highlights a good accordance between simulated hourly loads and real data for the given period.

Regards the in sample simulation, we omit for sake of brevity the parameters' discussion. We compare

again in the following Figure 10 the simulated load values (dotted line) with respect to the empirical ones (solid line) at hourly level for the period of 14-20 February, 2011.
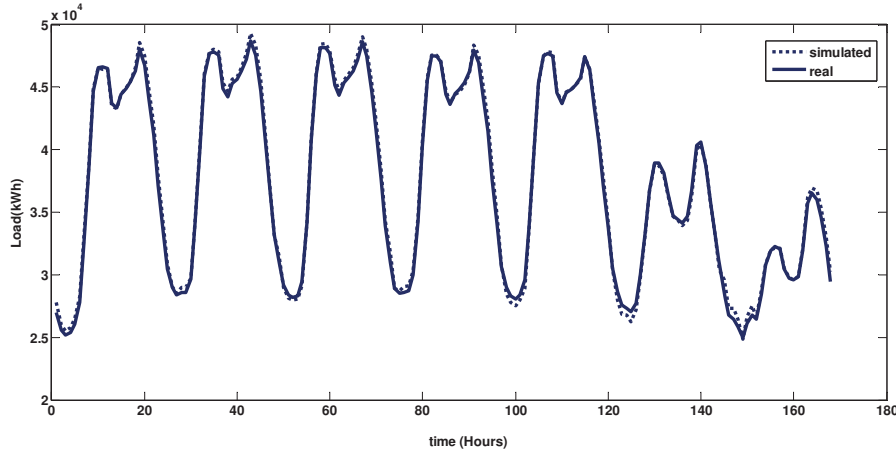

Fig. 10. Real load data vs. empirical ones

Figure 10 highlights again a good accordance between simulated hourly loads and real data for the given period. The fitting is obviously better with respect to the out of sample case.

We now describe in the next section the numerical fitting tests.

3.2. Model performance. In order to perform this task we use the following indicators (where we denote $\hat{y}_i$ the simulated values, y_i the real values, $\varepsilon_i = \hat{y}_i - y_i$ and n is the length of the sample).

♦   MAE (mean absolute error)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| = \frac{1}{n}\sum_{i=1}^{n}\left|\varepsilon_i\right|. \tag{13}$$

♦   MAPE (mean absolute percentage error)

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{y_i}\right|. \tag{14}$$

The MAPE does not take into account over-estimated or under-estimated values. At this purpose we elaborate the following two indicators:

$$MAPE^{+} = \frac{100}{n}\sum_{i=1}^{n}\frac{Max\left(\hat{y}_i - y_i;0\right)}{y_i},$$

$$MAPE^{-} = \frac{100}{n}\sum_{i=1}^{n}\frac{\left|Min\left(\hat{y}_i - y_i;0\right)\right|}{y_i}, \tag{15}$$

which satisfy the condition MAPE^{+} + MAPE^{-} = = MAPE. These latter can be normalized as follows

$$MAPE_{N}^{+} = \frac{100}{1_{\hat{y}_i > y_i}}\sum_{i=1}^{n}\frac{Max\left(\hat{y}_i - y_i;0\right)}{y_i},$$

$$MAPE_{N}^{-} = \frac{100}{1_{\hat{y}_i < y_i}}\sum_{i=1}^{n}\frac{\left|Min\left(\hat{y}_i - y_i;0\right)\right|}{y_i}, \tag{16}$$

where 1_X denotes the counting function and

$$1_{\hat{y}_i > y_i} + 1_{\hat{y}_i < y_i} = n.$$

The relation

$$\frac{1_{\hat{y}_i > y_i}}{n}\cdot MAPE_{N}^{+} + \frac{1_{\hat{y}_i < y_i}}{n}\cdot MAPE_{N}^{-} = MAPE$$

also holds.

♦   RMSE (root mean square error):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2} \tag{17}$$

♦ R-squared:

$$R^2 = 1 - \frac{Var\left(\hat{y} - y\right)}{Var\left(y\right)}. \tag{18}$$

As the R-squared increases with the number of variables, the adjusted R-squared $\bar{R}^2$ has been introduced as follows:

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\cdot\frac{n-1}{n-p-1}, \tag{19}$$

where p denotes the number of variables of the model.

The in sample results (1,000 simulations) for the period of 2010-2011 are summarized in the Table 4 below.

Table 4. Fitting results of 2010-2011

Indicator	Value
MAE	819.31 MWh
MAPE	2.46%
MAPE+	1.26%
MAPE-	1.20%
RMSE	1,150.70 MWh
R-SQUARED	97.42%

Finally, the out of sample results for the year 2011 (and estimation window 2007-2010) are summarized in the Table 5 below.

Table 5. Fitting results 2011

Indicator	Value
MAE	1,103.10 MWh
MAPE	3.25%
RMSE	1,460.90 MWh
R-SQUARED	95.83%

Remark. If we consider only calendar variables in the linear regression of the deterministic part, the overall fitting clearly worsens. For the in sample simulation, the MAPE is 2.79% (while R-squared is 96.86%) and regards the out of sample simulation, the MAPE is 5.02% (while R-squared is 95.36%).

Conclusions

The purpose of this paper is to estimate a long-term hourly load model for the Italian market. The first step of the paper consists in decomposing the data into a daily part and an hourly part, respectively. At this stage, we modeled the deterministic and the stochastic component separately for the daily part and the hourly part.

Regarding the deterministic component, we used linear regression. At this purpose, we highlighted that the challenging task was to determine the most suitable explanatory variables for both the daily and the hourly part in order to capture seasonal effects.

For what concerns the stochastic component, coming from the residuals of the previous regressions, we used classical time series tools in order to establish the more adequate processes.

The parameters of the model were calibrated thanks to publicly available hourly load values for the Italian market in the period of 2007-2010 for the 2011 out of sample estimation (and the period of 2010-2011 for the in sample estimation).

After determining the characteristics and the parameters of the model, we performed a Monte Carlo simulation and we compared simulated values with real data. The classical fitting tests suggested a good accordance between the simulated values and the empirical ones.

The model proposed is then a suitable way to forecast Italian electricity usage. Indeed, the economic and non-economic variables (specific to Italian market) permit to take into account the long term trend and the calendar and hourly dummy variables capture the cyclical patterns of the data.

The model may be improved by identifying more eligible regressors for the deterministic component and by using more refined processes for the stochastic part. Further research will also be dedicated to the modeling of the price of electricity in order to face the problem of electricity derivatives pricing.

References

1. Afshar, K., Bigdeli, N. (2011). Data analysis and short term load forecasting in Iran electricity market using singular spectral analysis (SSA), *Energy*, 36 (5), pp. 2620-2627.
2. Alter, N., Syed, H.S. (2011). An empirical analysis of electricity demand in Pakistan, *International Journal of Energy Economics and Policy*, 1 (4), pp. 116-139.
3. Andersson, G., Cornel, J., Hezog, F., Hildmann, M., Stokic, D. (2011). Robust calculation and parameter estimation of the Hourly Price Forward Curve, *17th Power Systems Computation Conference*, August 22-26, 2011, Stockholm, Sweden.
4. Andersson, G., He, Y., Hildmann, M., Sotiropoulos, E. (2013). Modeling of Electricity Load for Forward Contract Pricing, *Institute of Electrical and Electronics Engineers*, IEEE, July 21, 2013.
5. Bianco, V., Manca, O., Nardini, S. (2009). Electricity consumption forecasting in Italy using Linear Regression Models, *Energy*, 34 (9), pp. 1413-1421.
6. Bilgili, M., Sahin, B., Yasar, A., Simsek, E. (2012). Electric energy demands of Turkey in residential and industrial sectors, *Renewable and Sustainable Energy Reviews*, 16, pp. 404-414.
7. Blàzquez, L., Nina, B., Filippini, M. (2012). Residential electricity demand for Spain: new empirical evidence using aggregated data, *Centre for Energy Policy and Economics, Swiss Federal Institutes of Technology*, CEPE Working Paper No. 82.
8. Bruhns, A., Deurveilher, G., Roy, J.S. (2005). A non-linear regression model for mid-term load forecasting and improvements in seasonality, *15th PSCC*, Liege, 22-26 August 2005.
9. Chujai, P., Kerdprasop, N., Kerdprasop, K. (2013). Time series analysis of household electric consumption with ARIMA and ARMA models, *IMECS 2013*, March 13-15, 2013, Hong Kong.

10. Deihimi, A., Orang, O., Showkati, H. (2012). Short-term electric load and temperature forecasting using wavelet echo state networks with neural reconstruction, *Energy*, 57, pp. 382-401.
11. Dordonnat, V., Koopman, S.J., Ooms, M., Dessertaine, A., Collet, J. (2008). An hourly periodic state space model for modelling French national electricity load, *International Journal of Forecasting*, 24 (4), pp. 566-587.
12. Fan, S., Hyndman, R.J. (2012). Short-term load forecasting based on a semi-parametric additive model, *IEEE Transactions on Power Systems*, 27 (1), pp. 134-141.
13. Filik, U.B., Gerek, Ö.N., Kurban, M. (2011). A novel modeling approach for hourly forecasting of long-term electric energy demand, *Energy Conversion and Management*, 52, pp. 199-211.
14. Gonzàles-Romera, E., Jaramillo-Moràn, M.A., Carmona-Fernàndez, D. (2008). Monthly electric energy demand forecasting with neutral networks and Fourier series, *Energy Conversion and Management*, 49, pp. 3135-3142.
15. Hong, W.-C. (2011). Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm, *Energy*, 36 (9), pp. 5568-5578.
16. Kavousian, A., Rajagopal, R., Fischer, M. (2013). Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy, 55, pp. 184-194.
17. Makridakis, S., Wheelwright, S.C., Hyndman, R.J. (1998). *Forecasting: methods and applications*, New York: John Wiley & Sons.
18. Migon, H.S., Alves, L.C. (2013). Multivariate dynamic regression: modeling and forecasting for intraday electricity load, *Applied Stochastic Models in Business and Industry*, 29, pp. 579-598.
19. Moral-Carcedo, J., Vicéns-Otero, J. (2005). Modelling the non-linear response of Spanish electricity demand to temperature variations, *Energy Economics*, 27 (3), pp. 477-494.
20. Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K. (2008). Electrical power load forecasting using hybrid self-organizing maps and support vector machines, *The 2nd International Power Engineering and Optimization Conference (PEOCO 2008)*, Shah Alam, Selangor, Malaysia, 4-5 June 2008.
21. Pardo, A., Meneu, V., Valor, E. (2012). Temperatures and seasonality influences on Spanish electricity load, *Energy Economics*, 24, pp. 55-70.
22. Pielow, A., Sioshansi, R., Roberts, M.C. (2012). Modeling short-run electricity demand with long-term growth rates and consumer price elasticity in commercial and industrial sectors, *Energy*, 46, pp. 533-540.
23. Saab, S., Badr, E., Nasr, G. (2001). Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon, *Energy*, 26 (1), pp. 1-14.
24. Soares, L.J., Medeiros, M. (2008). Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data, *International Journal of Forecasting*, 24, pp. 630-644.
25. Wang, C.-H., Grozev, G., Seo, S. (2012). Decomposition and statistical analysis for regional electricity demand forecasting, *Energy*, 41 (1), pp. 313-325.
26. Weron, R. (2006). *Modeling and forecasting electricity loads and prices. A statistical approach*, John Wiley & Sons Ltd.