

# Electricity Load Modeling: An Application to Italian Market

Giovanni Masala, Stefania Marica

**Abstract**—Forecasting electricity load plays a crucial role regards decision making and planning for economical purposes. Besides, in the 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. Empirical data about electricity load highlights a clear seasonal behavior (higher load during the winter season), which is partly due to climatic effects. We also emphasize the presence of load periodicity at a weekly basis (electricity load is usually lower on weekends or holidays) and at 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 non-linear regression and Fourier series while we will investigate the stochastic component through econometrical tools.

The calibration of the parameters' model will be performed by using data coming from the Italian market in a 6 year period (2007-2012). Then, we will perform a Monte Carlo simulation in order to compare the simulated data respect to the real data (both in-sample and out-of-sample inspection). The reliability of the model will be deduced thanks to standard tests which highlight a good fitting of the simulated values.

**Keyword**—ARMA-GARCH process, electricity load, fitting tests, Fourier series, Monte Carlo simulation, non-linear regression.

## I. 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, 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 properly 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.

Besides, load forecast plays also a crucial role in the price determination. The accuracy of electricity load forecasting has been intensively investigated over the past few years. Indeed, a wrong electricity load forecasting causes an increase of the cost of operations. For example, an overestimation causes a supply excess while underestimation causes insufficient

electricity supply.

Regards 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 the 15% of electricity consumption is imported from abroad. Besides, the Italian electricity sector has been recently restructured in order to follow the EC Directive 96/92 which aims to set up a single electricity market. Besides, from a political point of view, the prominent Legislative Decree n. 79 of March 16, 1999 (namely the Bersani Decree) conducted to liberalize the activities of electricity production, import, export, purchases and sales. Besides, it sets up an antitrust ceiling on the business of the dominant operator in order to advantage competition.

Let us examine the state of art. Several models have been introduced in recent literature with the purpose of modeling narrowly electricity load. More precisely, time series modeling approaches based on artificial neural networks (ANNs) and statistical methods were used. Bilgili et al. [6] apply the artificial neural network (ANN) methodology to forecast the Turkey's residential and industrial electricity consumption to analyze energy use and perform future projections for the period 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. [11] 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] 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 foresee 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,

Stefania Marica is with the Department of Economics, University of Cagliari, Via S. Ignazio, 17 – 09123 Cagliari, Italy (phone: +39.70.6753418, e-mail: gb.masala@unica.it, stefania.marica@unica.it).

prediction period, characteristics of the time series, and size of the time series [17].

Chujai et al.'s research [9] was to find a model to efficiently forecast the electricity consumption in a household by applying 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] lists in his book some general techniques for load electricity and spot prices modeling.

Saab et al. [23] use three different univariate models to forecast the monthly electric energy consumption in Lebanon, during the period 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] use a multivariate dynamic regression to forecast electricity consumption in the Brazilian southeastern submarket for one-day ahead. They compare the results from univariate dynamic regression model (including the seasonal, daily and weekly, effects only), and multivariate model (with dummy weekday type variables only) concluding that the first one performs better than the second one, although the difference in probability is not large.

Moral-Carcedo and Vicéns-Otero [19] analyze the effect of temperatures on the variability of the Spanish daily electricity demand, using different non-linear regression models.

Andersson et al. [3] 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 got with this approach are significantly better than these obtained with the mean value.

Alter and Syed [2] analyze the determinants of electricity demand in Pakistan during the period 1970-2010, using a cointegration and vector error correction approach. They find the existence of long run relationship 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.

Blázquez et al. [7] 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 that 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] 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] 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] 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 perform better than ARIMA and TF- $\epsilon$ -SVR-SA models.

Afshar and Bigdeli [1] 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] 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 allows avoiding the complexity of non-linear estimations.

Pielow et al. [22] 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] use an autoregressive least-squares regression to explore the effects of temperature and seasonality on daily Spanish electricity load. Soares and Medeiros [24] model the electricity hourly load of southeast of Brazil using rigorous statistical arguments. 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] 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.

González-Romera et al. [14] 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] 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] 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. While electric water heater and air conditioner better explain the maximum level consumption.

Collet et al. [10] 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] 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 few evidence about the short run electricity load in Italy. The aim of the present research is to fill this gap following Andersson et al.'s [4] model. 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, consumer price index, number of trips, electricity and gas prices. Furthermore, to deal with the cyclical patterns of data, we include in the daily load a Fourier series over the respective period, and estimate the deterministic component with a nonlinear multiple regression model. Besides, we investigate thoroughly the stochastic components by analyzing also heteroschedastic effects.

The paper is organized as follows. In the current Section I, we have introduced the objectives of our paper and we have presented the state of art. Section II is devoted to the main characteristics and seasonal features of the electricity consumption database. The theoretical model is described in Section III. Section IV is devoted to the empirical application and to the comparison with the real data. Finally, Section V concludes and lists some future enhancements.

## II. DATA DESCRIPTION

The database containing hourly data on loads and spot electricity prices can be freely retrieved from the Italian market operator's website [27] as well as daily data on forward electricity price and gas price. We selected data from 2007 to 2012. On the whole we have 52,608 hourly records (unit load is MWh) and 2,192 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 Fig. 1 the hourly values for the first week of June 2010 (starting with Tuesday, note that June, 2 is a holiday) while Fig. 2 exhibits the values for whole June 2010.

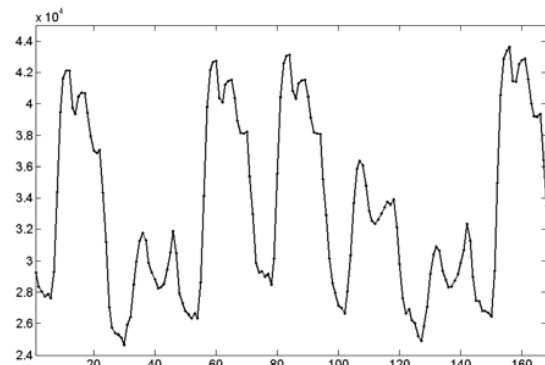


Fig. 1 Hourly values (MWh) 1-7 June 2010

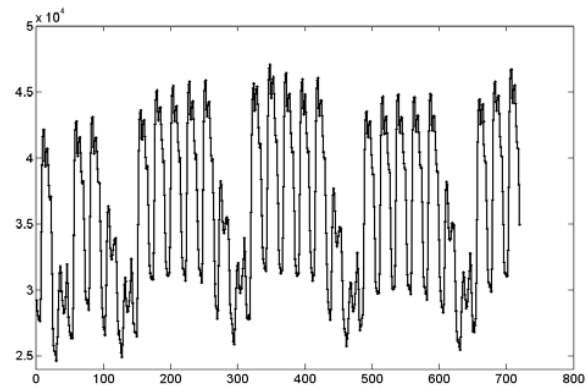


Fig. 2 Hourly values (MWh) June 2010

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.

Besides, we also estimate the mean load values for each hour (for the whole database). We show the results in the following Fig. 3.

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 the human activities. The following tri-dimensional Fig. 4 represents load values with respect to day and hour (in this case we have restricted to February, 2010). The hourly peaks are again evident.

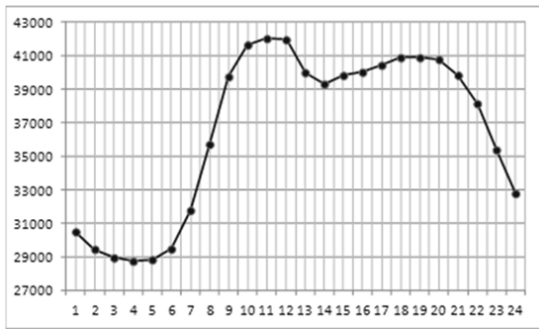


Fig. 3 Hourly mean load (hour vs. MWh)

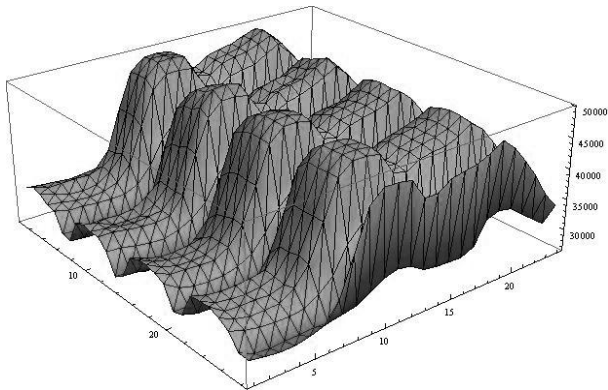


Fig. 4 Load (MWh) vs. Hour/Day

The daily cyclical feature may also be highlighted when considering daily loads. At this purpose, Fig. 5 represents the daily loads for June 2010 and Fig. 6 presents the daily loads for 2010.

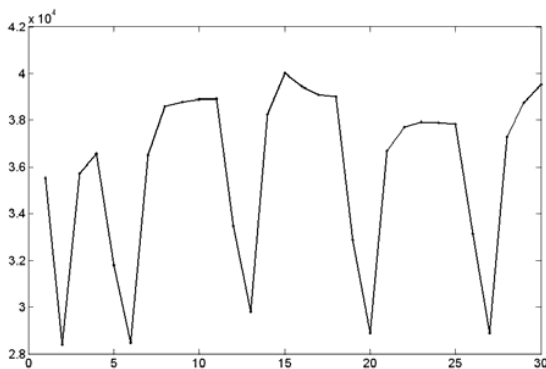


Fig. 5 Daily values (MWh) June 2010

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). We show the results in the following Fig. 7 (the first value corresponds to Monday).

We note 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).

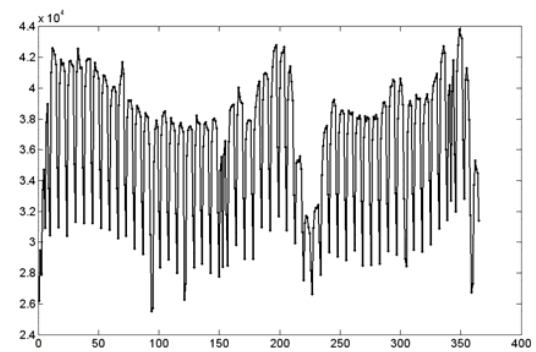


Fig. 6 Daily values (MWh) Year 2010

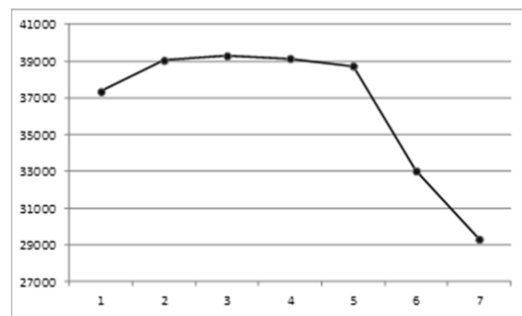


Fig. 7 Daily mean load (MWh)

If we wish to determine longer cyclical trends, we have to consider loads data on a monthly basis. We exhibit these values in Fig. 8 for the whole database.

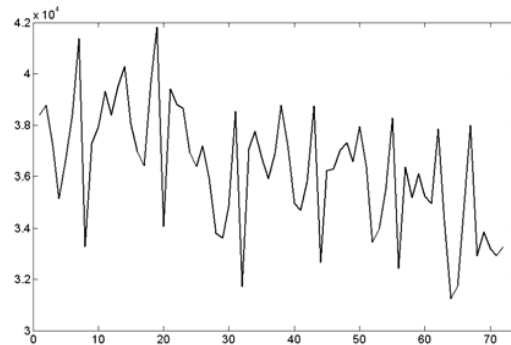


Fig. 8 Monthly values (MWh) Years 2007-2012

This plot highlights a cyclical behavior with respect to the month. For example, a lower peak occurs each year in August while the higher peak is rather variable. We note that each point in the x-axis represents a month.

Next, we estimate the mean load values for each month (for the whole database). We show the results in the following Fig. 9.

We deduce from this plot a lower peak in August (when industrial activities are reduced) and two higher peaks in February and July which represent respectively the colder winter month and the warmer summer month.

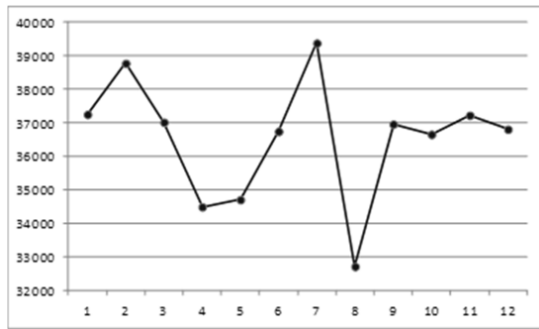


Fig. 9 Monthly mean load (MWh)

We finally give in Fig. 10 the yearly loads for the period 2007-2012.

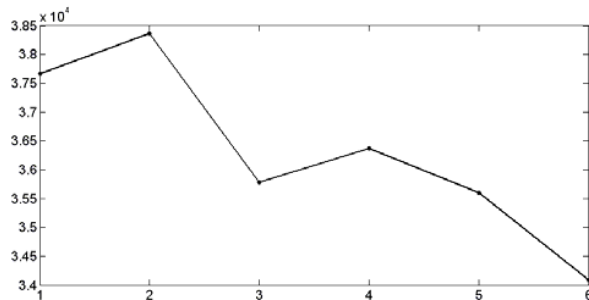


Fig. 10 Yearly values (MWh) 2007-2012

We deduce from Fig. 10 that the yearly load has been increasing in the period 2007-2008 and then the trend inverted (except for year 2010). This behavior has to be explained in the light of economics' indicators such as the Industrial Indicator Index (see at this purpose Section III).

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

TABLE I UNIT ROOT TESTS (ADF)	
Variable	Test value ( <i>p</i> value)
daily load	-7.8466 (0.000)
trips	-3.5638 (0.033)
electricity price	-3.0040 (0.131)
oil price	-1.5089 (0.827)
gas price	-2.5411 (0.308)

Table I 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 values are taken to be stationary. The opposite occurs for electricity price, oil price and gas price. So we can conclude that these variables are not stationary at levels. Anyway, the variable oil price showed a non-stationary nature caused by a stochastic trend, whereas despite the variables electricity price and gas price showed the same characteristics of the previous one, their trend is not of stochastic nature. For this reason oil price was omitted from this study.

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 [26]).

We remember that the periodogram associated with a vector of observations  $\{x_1, \dots, x_n\}$  is given by:

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

where  $\omega_k = 2\pi \cdot (k/n)$  denote the Fourier frequencies (rad/unit time). We illustrate in the next Fig. 11 the periodogram of daily and hourly data.

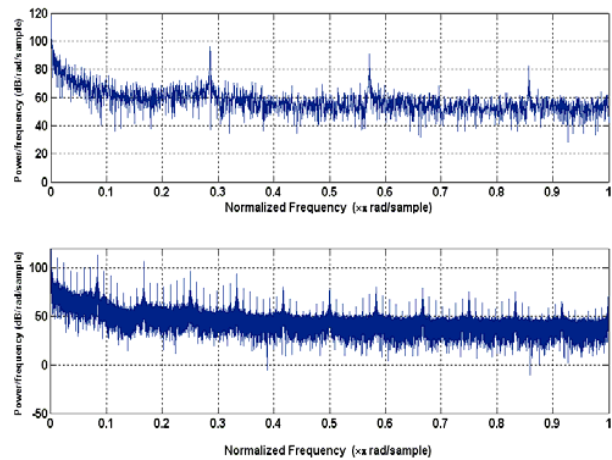


Fig. 11 Periodogram Daily/Hourly data

The daily plot presents a peak at frequency  $\omega_k = 0.1428$  (associated to a  $1/\omega_k = 7$  day period). The smaller peaks ( $\omega_k = 7/2$  and  $\omega_k = 7/3$ ) are the harmonics multiples. The hourly data presents peaks associated to 24 and 168 hours and their harmonics multiples.

### III. 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:

$$L(t) = f(t) + \tilde{x}(t) \quad (1)$$

where  $f(t)$  is the deterministic component and  $\tilde{x}(t)$  is the stochastic one. We further decompose the load level between the hourly component  $\tilde{L}_h$  and the daily component  $\tilde{L}_d$ :

$$\tilde{L}(t) = \tilde{L}_d(t) + \tilde{L}_h(t) \quad (2)$$

**Remark:** In order to examine the daily characteristics, the database has to be adapted by computing the average hourly loads for each day.

#### A. The Daily Component

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

$$\tilde{L}_d(t) = f_d(t) + \tilde{x}_d(t) \quad (3)$$

We can model the deterministic component through multiple non-linear regressions. 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, consumer price index, trips, gas price, electricity price);
- non economic variables (temperature index).
- Fourier terms.

Besides, we also consider some quadratic variables (square of  $IP$ ) as well as some interactions between some of the variables (trips/December and  $IP$ /August). Finally, some lagged variables (temperature index and electricity price) turn out to be significant. We get then:

$$\begin{aligned} L_d(t) = & \alpha_d + \sum_{i=1}^7 \beta_{di} \cdot D_i(t) + \sum_{i=1}^{12} \gamma_{di} \cdot M_i(t) + \\ & + \delta_d \cdot H(t) + \zeta_d \cdot SH(t) + \lambda_{d1} \cdot IP(t) + \lambda_{d2} \cdot CPI(t) + \\ & + \lambda_{d3} \cdot Tr(t) + \lambda_{d4} \cdot P_G(t) + \lambda_{d5} \cdot P_E(t) + \lambda_{d6} \cdot T(t) + \\ & + \mu_{d1} \cdot P_E(t-1) + \mu_{d2} \cdot T(t-1) + \mu_{d3} \cdot IP^2(t) + \\ & + \mu_{d4} \cdot Tr(t) \cdot M_{12}(t) + \mu_{d5} \cdot IP(t) \cdot M_8(t) + \\ & + \sum_{m=1}^3 \left( a_m \cdot \sin \frac{2\pi m}{7} \cdot t + b_m \cdot \cos \frac{2\pi m}{7} \cdot t \right) \end{aligned} \quad (4)$$

We choose the following regressors:  $IP$  (industrial production),  $CPI$  (consumer price index),  $P_G$  (gas price),  $P_E$  (electricity price),  $T$  (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 [28]. We show in Fig. 12 the yearly industrial production for the period 2007-2012. Note the similarities with Fig. 10.

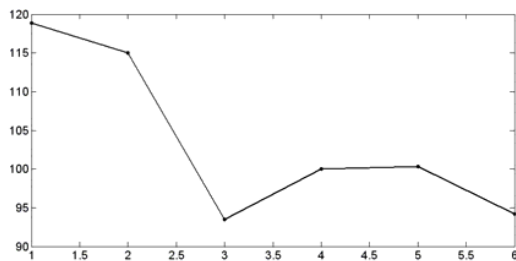


Fig. 12 Yearly Industrial Production Index period 2007-2012

The close binding between load values and  $IP$  index can be emphasized through a correlation analysis. At this purpose, the

correlation with respect to monthly values is 69% and the correlation with respect to yearly values is 89%. 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 Fig. 13 the load values respect to temperature.

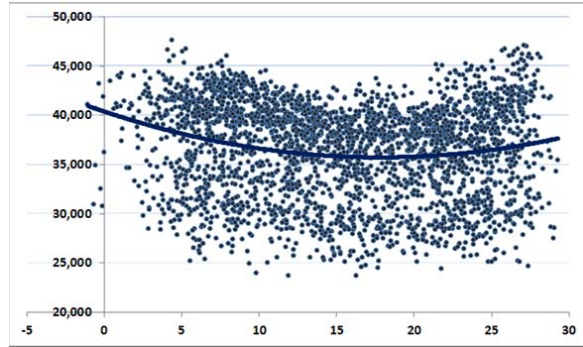


Fig. 13 Load values (MWh) vs. temperature (°C)

Nevertheless, the temperatures' 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:

$$\bar{T} = \begin{cases} T - 20 & \text{if } T > 20 \\ 0 & \text{if } 16 < T < 20 \\ 16 - T & \text{if } T < 16 \end{cases}$$

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 Autoregressive-GARCH process  $AR(p) - GARCH(1,1)$ :

$$\tilde{x}_d(t) = c + \sum_{i=1}^p \delta_i \cdot \tilde{x}_d(t-i \cdot \Delta t) + \tilde{\varepsilon}(t) \quad (5)$$

with

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

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

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

#### B. The Hourly Component

We can deduce the hourly component from (2):

$$\tilde{L}_h(t) = \tilde{L}(t) - \tilde{L}_d(t) \quad (6)$$

We have again the deterministic and the stochastic part:



$$\tilde{L}_h(t) = f_h(t) + \tilde{x}_h(t) \quad (7)$$

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

$$L_h(t) = \sum_{i=1}^7 \sum_{j=1}^{12} \left[ d_{ij} + \sum_{k=1}^{24} \xi_{ijk}(t) \cdot H_{ijk}(t) \right] \quad (8)$$

where the regressor  $H$  denotes the hour. 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  $AR(p)$ -GARCH(1,1) as before for the same reasons:

$$\tilde{x}_h(t) = \sum_{i=1}^7 \sum_{j=1}^{12} \left[ c_{h,ij} + \sum_{k=1}^p \delta_{h,ijk} \cdot \tilde{x}_{h,ij}(t-k \cdot \Delta t) + \tilde{\varepsilon}_{h,ij}(t) \right] \quad (9)$$

with

$$\begin{aligned} \tilde{\varepsilon}_{h,ij}(t) &= \eta_{h,ij}(t) \cdot \sqrt{h_{h,ij}(t)}, \quad \eta_{h,ij}(t) \stackrel{i.i.d.}{\sim} N(0,1) \\ h_{h,ij}(t) &= \omega_{h,ij} + \alpha_{h,ij} \cdot \tilde{\varepsilon}_{h,ij}^2(t-1) + \beta_{h,ij} \cdot h_{h,ij}(t-1) \end{aligned}$$

### C. The Overall Model

We get the final model by aggregating all the components described before:

$$\begin{aligned} \tilde{L}(t) &= \alpha_d + \sum_{i=1}^7 \beta_{di} \cdot D_i(t) + \sum_{i=1}^{12} \gamma_{di} \cdot M_i(t) + \\ &+ \delta_d \cdot H(t) + \zeta_d \cdot SH(t) + \lambda_{d1} \cdot IP(t) + \lambda_{d2} \cdot CPI(t) + \\ &+ \lambda_{d3} \cdot Tr(t) + \lambda_{d4} \cdot P_G(t) + \lambda_{d5} \cdot P_E(t) + \lambda_{d6} \cdot T(t) + \\ &+ \mu_{d1} \cdot P_E(t-1) + \mu_{d2} \cdot T(t-1) + \mu_{d3} \cdot IP^2(t) + \\ &+ \mu_{d4} \cdot Tr(t) \cdot M_{12}(t) + \mu_{d5} \cdot IP(t) \cdot M_8(t) + \\ &+ \sum_{m=1}^3 \left( a_m \cdot \sin \frac{2\pi m}{7} \cdot t + b_m \cdot \cos \frac{2\pi m}{7} \cdot t \right) + \\ &+ c + \sum_{i=1}^p \delta_i \cdot \tilde{x}_d(t-i \cdot \Delta t) + \sigma_d \cdot \tilde{\varepsilon}(t) + \\ &+ \sum_{i=1}^7 \sum_{j=1}^{12} \left[ d_{ij} + \sum_{k=1}^{24} \xi_{ijk}(t) \cdot H_{ijk}(t) \right] + \\ &+ \sum_{i=1}^7 \sum_{j=1}^{12} \left[ c_{h,ij} + \sum_{k=1}^p \delta_{h,ijk} \cdot \tilde{x}_{h,ij}(t-k \cdot \Delta t) + \sigma_{h,i,j} \cdot \tilde{\varepsilon}_{h,ij}(t) \right] \end{aligned} \quad (10)$$

with

$$\begin{aligned} \tilde{\varepsilon}(t) &= \eta(t) \cdot \sqrt{h(t)}, \quad \eta(t) \stackrel{i.i.d.}{\sim} N(0,1) \\ h(t) &= \omega + \alpha \cdot \tilde{\varepsilon}^2(t-1) + \beta \cdot h(t-1) \\ \tilde{\varepsilon}_{h,ij}(t) &= \eta_{h,ij}(t) \cdot \sqrt{h_{h,ij}(t)}, \quad \eta_{h,ij}(t) \stackrel{i.i.d.}{\sim} N(0,1) \\ h_{h,ij}(t) &= \omega_{h,ij} + \alpha_{h,ij} \cdot \tilde{\varepsilon}_{h,ij}^2(t-1) + \beta_{h,ij} \cdot h_{h,ij}(t-1) \end{aligned}$$

## IV. EMPIRICAL APPLICATION

In this section, we will unveil the estimation of the

parameters of the model and then we present the evaluation of the performance of the model in the calibrating window (2007-2011) and out of sample for the year 2012.

### A. 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).

#### 1. Deterministic Part of the Daily Component

This part is given by a non-linear regression with 39 regressors as already discussed in Section III.

The coefficients of the main variables have expected signs. It means that calendar, economic and not economic variables matter to explain electricity load behavior. The temperature index coefficient is positive (its value is 85.464) and statistically significant ( $p$  value 0.001). It means that when temperature index moves away from the given threshold value (16 °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 (the coefficient size is 260.82 with  $p$  value 0.001).

The estimation results also confirm the trend depicted in Fig. 1: during working days electricity load is higher than Saturdays and Sundays. Furthermore, holidays' days are like weekends' days, negatively related to the electricity load (the coefficient size is -5088 with  $p$  value 0.000). We omit the description of the other parameters for sake of brevity.

#### 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 (Fig. 14).

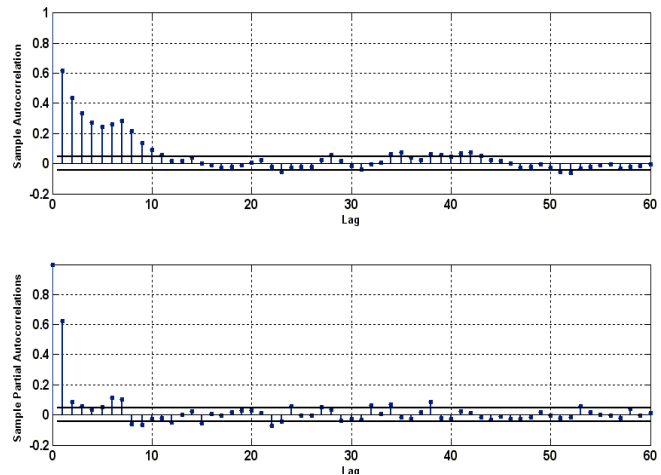


Fig. 14 ACF and PACF for daily residuals

These plots show the presence of autocorrelation. Besides, the Engle's test detects the presence of residual heteroschedasticity. We propose then to model the residuals through an AR(8)-GARCH(1,1) process. The new residuals are given in the following Fig. 15.

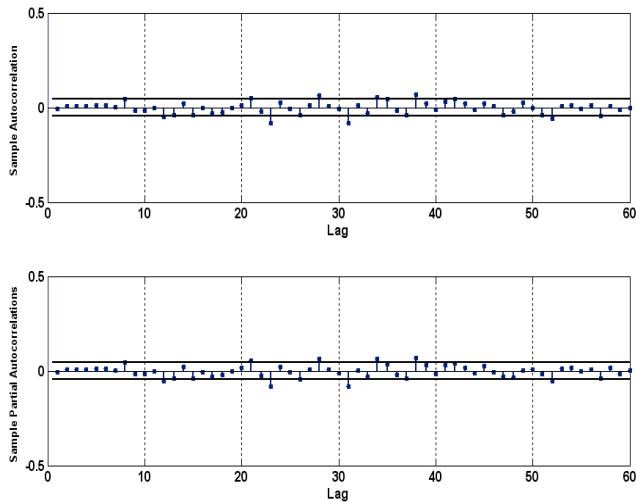


Fig. 15 ACF and PACF for daily AR(8)-GARCH(1,1) residuals

We deduce from these plots that residuals' autocorrelation has been removed.

Then, we also exhibit in the following Table II the parameters of the process with the associated test statistics.

Parameter	Value	t Statistic
Constant	10	0.8000
AR(1)	0.527506	19.5076
AR(2)	0.035728	1.0156
AR(3)	0.061980	1.8281
AR(4)	0.002647	0.0836
AR(5)	-0.030396	-1.0685
AR(6)	0.079031	2.7634
AR(7)	0.201525	8.6721
AR(8)	-0.084461	-4.5428
Constant	63068.4	8.8105
GARCH(1)	0.654649	31.3656
ARCH(1)	0.286045	12.6495

### 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 Section III. The regression is performed after fixing the type of day and month (for this reason, we do not apply here the Fourier analysis). For sake of brevity, we do not exhibit the results.

### 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 plot the autocorrelation function and the partial autocorrelation function (Fig. 16). As an example we consider type of day 1 and month January.

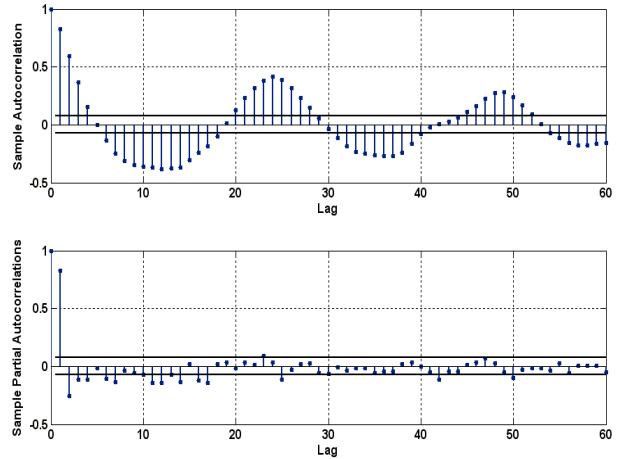


Fig. 16 ACF and PACF for hourly residuals

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

We propose to model these residuals through an AR(10)-GARCH(1,1) process. The autocorrelation function and the partial autocorrelation function of the residuals are given in the following Fig. 17.

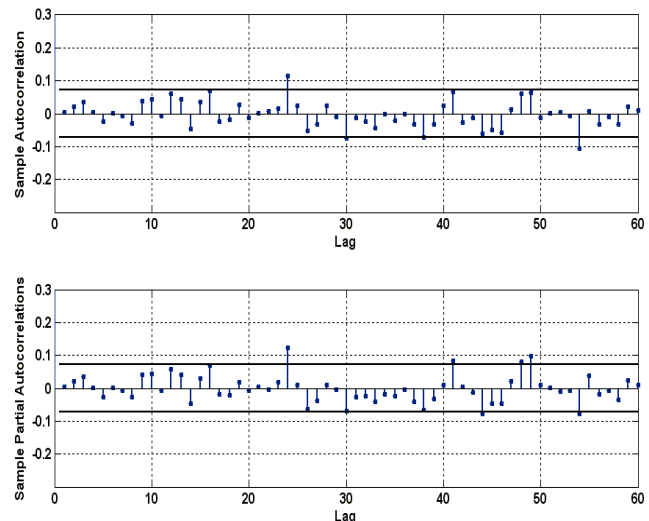


Fig. 17 ACF and PACF for daily AR(10)-GARCH(1,1) residuals

We deduce from these plots that residual autocorrelation has been removed.

We omit again the results.

Let us finally compare in the following Fig. 18 the simulated load values (dotted line) with respect to the empirical ones (solid line) at hourly level for the period 1-7 February, 2009.

The latter plot highlights a good accordance between simulated hourly loads and real data for the given period.

Regards the out-of-sample simulation, we compare again in the following Fig. 19 the simulated load values (dotted line) with respect to the empirical ones (solid line) at hourly level



for the period 1-7 February, 2012.

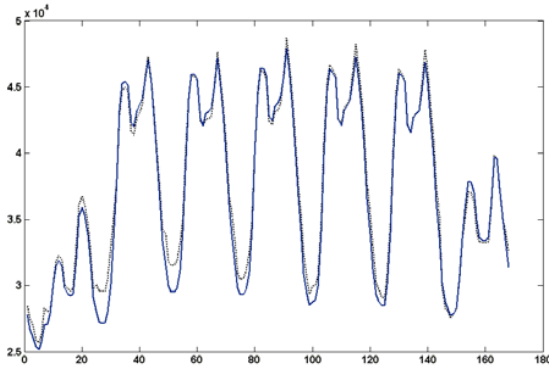


Fig. 18 Real load data vs. empirical ones

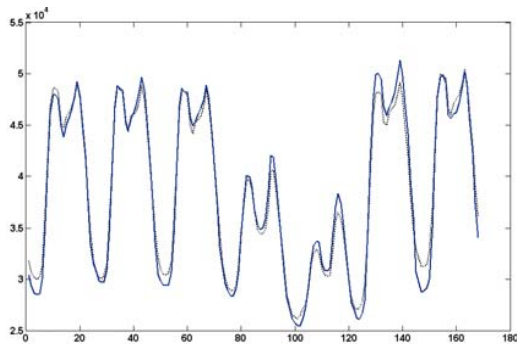


Fig. 19 Real load data vs. empirical ones

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

#### B. 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)

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| = \frac{1}{n} \sum_{i=1}^n |\varepsilon_i|$$

- MAPE (mean absolute percentage error)

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

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

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

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

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

$$MAPE_N^+ = \frac{100}{\mathbf{1}_{\hat{y}_i > y_i}} \sum_{i=1}^n \frac{\text{Max}(\hat{y}_i - y_i; 0)}{y_i}$$

$$MAPE_N^- = \frac{100}{\mathbf{1}_{\hat{y}_i < y_i}} \sum_{i=1}^n \frac{|\text{Min}(\hat{y}_i - y_i; 0)|}{y_i}$$

where  $\mathbf{1}_X$  denotes the counting function and  $\mathbf{1}_{\hat{y}_i > y_i} + \mathbf{1}_{\hat{y}_i < y_i} = n$ .

The relation  $\frac{\mathbf{1}_{\hat{y}_i > y_i}}{n} \cdot MAPE_N^+ + \frac{\mathbf{1}_{\hat{y}_i < y_i}}{n} \cdot MAPE_N^- = MAPE$  also holds.

- RMSE (root mean square error):

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

- R square:

$$R^2 = 1 - \frac{\text{Var}(\hat{y} - y)}{\text{Var}(y)}$$

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

$$\bar{R}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-p-1}$$

where  $p$  denotes the number of variables of the model. The in-sample results (1,000 simulations) for the period 2007-2011 are summarized in the Table III below:

TABLE III FITTING RESULTS 2007-2011	
Indicator	Value
MAE	988.77 MWh
MAPE	2.89%
MAPE <sup>+</sup>	1.46%
MAPE <sup>-</sup>	1.43%
RMSE	1,315.30 MWh
R SQUARE	96.80%

Finally, the out-of-sample results for the year 2012 are summarized in the Table IV below.

TABLE IV FITTING RESULTS 2012	
Indicator	Value
MAE	1,459.20 MWh
MAPE	4.41%
RMSE	1,824.90 MWh
R SQUARE	94.51%

#### V. CONCLUSIONS

The purpose of this paper is to set up a long term hourly load model for Italy. 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.

Regards the deterministic component, we used non-linear regression. At this purpose, we highlighted that the

challenging task was to determine the more suitable regressors for both the daily and the hourly part in order to capture both seasonal and periodic effects.

For what concerns the stochastic component, coming from the residuals of the previous regressions, we used econometric 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 2007-2011.

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

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 the modeling of the price of electricity in order to face the problem of electricity derivatives pricing.

#### REFERENCES

- [1] Afshar K. and Bigdeli N. (2011) "Data analysis and short term load forecasting in Iran electricity market using singular spectral analysis (SSA)", *Energy*, Vol. 36, pp. 2620-2627.
- [2] Alter N. and Syed H. S. (2011) "An empirical analysis of electricity demand in Pakistan", *International Journal of Energy Economics and Policy*, Vol. 1, No. 4, pp. 116-139.
- [3] Andersson G., Cornel J., Hezog F., Hildmann, M. and 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. and 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. and Nardini S. (2009) "Electricity consumption forecasting in Italy using Linear Regression Models", *Energy*, Vol. 34, pp. 1413-1421.
- [6] Bilgili M., Sahin B., Yasar A. and Simsek E. (2012) "Electric energy demands of Turkey in residential and industrial sectors", *Renewable and Sustainable Energy Reviews*, Vol. 16, pp. 404-414.
- [7] Blázquez L., Nina B. and 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., and Roy J. S. "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. and Kerdprasop K. "Time series analysis of household electric consumption with ARIMA and ARMA models", *IMECS 2013*, March 13-15, 2013, Hong Kong.
- [10] Collet J., Dessertaine A., Dordonnat V., Koopman S. J. And Ooms M., (2008) "Journal of Forecasting", Vol. 24, pp. 566-587.
- [11] Deihimi A., Orang O. and Showkati H. (2012) "Short-term electric load and temperature forecasting using wavelet echo state networks with neural reconstruction", *Energy*, Vol. 57, pp. 382-401.
- [12] Fan S. and Hyndman R. J. (2012) "Short-term load forecasting based on a semi-parametric additive model", *IEEE Transactions on Power Systems*, Vol. 27, No. 1, pp. 134-141.
- [13] Filik U. B., Gerek Ö., N. and Kurban M. (2011) "A novel modeling approach for hourly forecasting of long-term electric energy demand", *Energy Conversion and Management*, Vol. 52, pp. 199-211.
- [14] González-Romera E., Jaramillo-Morán M. A. and Carmona-Fernández D. (2008) "Monthly electric energy demand forecasting with neural networks and Fourier series", *Energy Conversion and Management*, Vol. 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*, Vol. 36, pp. 5568-5578.
- [16] Kavousian A., Rajagopal R. and 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*, Vol. 55, pp. 184-194.
- [17] Makridakis S., Wheelwright S. C., and Hyndman R. J. (1998) "Forecasting: methods and applications" New York: John Wiley & Sons.
- [18] Migon H. S., and Alves L. C., (2013) "Multivariate dynamic regression: modeling and forecasting for intraday electricity load", *Applied Stochastic Models in Business and Industry*, Vol. 29, pp. 579-598.
- [19] Moral-Carcedo J. and Vicéns-Otero J. (2005) "Modelling the non-linear response of Spanish electricity demand to temperature variations", *Energy Economics*, Vol. 27, Issue 3, pp. 477-494.
- [20] Nagi J., Yap K.S., Tiong S.K., Ahmed, S.K. "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., and Valor E. (2002) "Temperatures and seasonality influences on Spanish electricity load", *Energy Economics*, Vol. 24, pp. 55-70.
- [22] Pielow A, Sioshansi R. and Roberts M.C. (2012) "Modeling short-run electricity demand with long-term growth rates and consumer price elasticity in commercial and industrial sectors", *Energy*, Vol. 46, pp.533-540.
- [23] Saab S., Badr E., and Nasr G. (2001) "Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon", *Energy*, Vol. 26, pp. 1-14.
- [24] Soares L. J. and Medeiros M. (2008) "Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data", *International Journal of Forecasting*, Vol. 24, pp. 630-644.
- [25] Wang C-H., Grozev G. and Seo S. (2012) "Decomposition and statistical analysis for regional electricity demand forecasting", *Energy*, Vol. 41, pp. 313-325.
- [26] Weron R., (2006), "Modeling and forecasting electricity loads and prices. A statistical approach", John Wiley & Sons Ltd.
- [27] <http://www.mercatoelettrico.org/En/Default.aspx>
- [28] <http://www.istat.it/en>