Empirical Forecasting Model of Hungarian Power System Load (2014-2015)

By

Imola Horváth

Submitted to

Central European University

Department of Economics

In partial fulfilment of the requirements for the degree of Master of Arts

Supervisor: Professor Gábor Kőrösi

Budapest, Hungary

2016

Abstract

Electricity cannot be stored such as the other commodities, therefore the demand and the supply side of the power market has to be in balance all the time. As a result, an accurate forecasting model is crucial in terms of system security in order to avoid blackouts or excessive power generation. However, the currently applied model of Hungarian Transmission System Operator is inaccurate, which resulted in a high operational cost. The main objective of my thesis is to develop a handy and accurate system load forecasting model for 2015. For my research, I used the database of the Hungarian Transmission System Operator, which contains hourly system load and weather forecast data for 2014-2015. I tested the forecasting performance of double seasonal ARIMA, Holt-Winters exponential smoothing with double seasonal cycle and log-linear models. One of the key findings of my thesis is that in terms of the Hungarian system load the dummy variables can handle the seasonal pattern better than the seasonal differencing. Moreover, the calendar variables proved to be more significant explanatory variable, than the weather variables, especially those which were responsible for the effect of the holidays. Finally the log-linear model specification had the best forecasting accuracy, in the case of which the average MAPE of a day-ahead forecast is 2.4%. With the help of my model the Hungarian Transmission System Operator could increase the predictability and security of the system, furthermore decrease its operational cost.

Table of contents

List of Figures and Tables

Introduction

Electricity is all around us. All of the devices surrounding us in our everyday lives such as smart phones, laptops or household appliances work with electricity. Moreover, the industrial production and technology heavily rely on electricity as well. However, the operation of the power system on country level is difficult due to the special property of electricity, namely that it cannot be stocked like other commodities. As a result, the demand and the supply sides of the power system have to be in balance all the time. The control and schedule of electricity generation and transmission of electricity to meet the demand are the duty of the Transmission System Operator (TSO) of the given countries or regions. TSO has a high responsibility because deviation from equilibrium could have severe consequences. If the supply side exceeds the demand side, it can lead to the waste of the sources and to the increase of the cost of electricity supply. On the other hand, if the demand side is excessive, then it can be resulted in blackouts. The main stress in the system is the volatility of the demand side, which is constantly subjected to random shocks. However, it is hard to instantly adjust to the demand side, because the supply side of the system is quite inflexible. This inelasticity is due to the fact that launching a new generator has a high fix cost, its variable cost based on its efficiency level, moreover some types of generators cannot start producing at once. Furthermore, the spread of intermittent resources in the system such as solar panels and wind turbines makes it even harder to predict the supply side. Therefore, the biggest challenge of the TSO is to meet the constantly fluctuating electricity demand with a quite inflexible supply side without relying on any inventory. As a consequence, a precise model and accurate forecast of system load is very important from a system security point of view. (As electricity demand is equal to electricity supply all the time, it can be used as a synonym for system load).

Since electricity system stability is a crucial and strategical question of a given region or country, many publications in the economic literature have dealt with the problem of load forecasting. However, the currently applied model of MAVIR is inaccurate, which resulted in a high operational cost. Nevertheless, there is no relevant Hungarian publication from this field. Therefore, the main objective of my thesis is to develop a handy and accurate forecasting model for the Hungarian Transmission System Operator (MAVIR).

However, the development of a precise forecasting model is a complex task. According to Hahn et al (2009) the main purpose of the forecast has to be clarified as a first step, because that determines immediately the horizon of the forecast, which defines the set of the most influential factors, which have to be considered and the modelling approach as well. If the forecast is supposed to endorse a strategic decision of a company or the implementation of a new policy, then its breadth of vision should span a period from 1 up to 20 years. In the case of long term forecast economic related determinants, such as GDP, inflation, price of electricity have to be considered. On the other hand, when the scope of the load forecast is a year with monthly or weekly frequency then we speak about medium term load forecast. These serve for planning business operation and production or support the contract negotiation with the power trader. Finally, the short term load forecast (STLF) stands for promoting the day-to-day operation of the TSO. The horizon of STLF is usually from one day a week ahead with hourly or half-hourly frequency. (Hahn et al, 2009) For the aim of my thesis the STLF approach is the most suitable, hence I cover the literature focusing exactly on STLF.

There are two prevailing trends in the case of STLF: one is the class of conventional statistical methods and regression-based econometric techniques, the other newly emerged group is the Artificial Intelligence (AI) and Computational procedures (Hahn et al, 2009). The substantial part of the studies written in this field introduce a specific Machine Learning or Hybrid techniques for load forecasting and use only conventional methods as benchmark to prove the superiority of the computational algorithms. For this reason I regard it very important to give a broader overview about these routines and map the future possibilities, therefore I introduce some case studies as well, which solved the exact load forecast problems with AI methods. However, in my thesis I focus on conventional statistical and econometric techniques because these modelling approaches suit my research problem. Hence, I cover those papers in the literature in depth which use ARIMA, SARIMA model, exponential smoothing techniques and further multivariate regression-based models, incorporating the influential weather and calendar related exogenous variables for load forecast.

Therefore, in the next chapter I give an overview of the applied approaches and techniques on load forecast and I also introduce some case studies to demonstrate the forecasting power, the advantages and disadvantages of the given methods in a practical, real life situation. Then in chapter 2 I map the characteristic of the Hungarian system load and reveal the most influential external factors and their relationship to the system load. The available dataset for my analysis, given by MAVIR contains load and one-day-ahead weather forecast data for the 2014-2015 period. Based on the results of my graphic analysis, I consider the replication of the most suitable models among the introduced studies in my literature review. Hence, in Chapter 3 I demonstrate the goodness of fit of the models of Taylor et al (2006), Ergün and Jun (2011) and Ramanathan et al (1997) on the Hungarian load database. In chapter 4, I experiment with the development of an own model taking into account the specificity of the Hungarian system according to the main findings of the replications. Finally, I summarize the key research findings of my thesis and give recommendation for further improvements.

1. Literature review

As the main purpose of my thesis is to develop an intuitive forecasting model for MAVIR, I will review the available solutions and best practices in the literature of STLF. Since, there is no relevant Hungarian studies related to my research topic, I covered the foreign studies written in the topic. According to the categorization of Hahn et al (2009) the STLF models can be assigned to two big families: to the classical models, which use the concept of regression analysis and conventional statistical methods; and to the Artificial and Computational Intelligence models. My thesis gives a detailed overview of the different methods and techniques belonging to the two main groups. Therefore, among the classical models the ARIMA time series models, exponential smoothing techniques and regression based principal component analyses methods are explained. With respect to the computational algorithms my thesis presents the key concepts of logic and percepton based routines, the statistical learning procedures and the support vector regression according to the classification of Maglogiannis (2007). Beside the theoretical overview of the listed methods, their application for real life forecasting problems are introduced through some case studies. The main aim of the detailed literature review is to map the best practice of STLF and to use up the key findings of the studies for the development of my own model.

However, before the selection of the appropriate STLF it is necessary to clarify the set of exogenous variables which have the most influential effect on the evolution of load on short horizon (Hahn et al, 2009). In the case of short time horizon prediction the most influential external factors are the fast changing, fluctuating weather forecast variables and calendar variables. In the short term the economic related determinants, such as GDP, inflation, price of electricity are not as relevant as in longer-term forecast, because these factors are stable on a daily basis.

It is important to mention that among the weather related variables the most important one is the air temperature, which is in a non-linear connection with the load demand, because usually the electricity usage is higher in winter due to heating. On the other hand, electricity consumption become higher again in the hot summer days because of cooling. However, the explanatory power of the weather related variables depend on the climatic condition of the given country or region (Hahn et al, 2009). Beside the weather variables, the calendar ingredients play an important role in capturing the superimposed seasonal pattern of the load demand. The electricity demand has intraday, within-week and yearly cycles. In the literature weekdays, weekends, holidays and transitory days such as Mondays and Fridays are differentiated. Among the studies, collected by Hahn et al (2009) there are two modelling approaches in handling the periodic electricity demand: the local and the global approach. In relation to the local approach there are separate models for each identified features, which means that distinct regressions are specified for each hour or for weekdays and weekends. However, it requires large databases, which are often not available. On the other hand, with respect to the global approach there is a single monolithic model which captures the seasonality via the introduction of additional explanatory variables (Hahn et al, 2009). After the determination of the most essential factors affecting the power load system on short term, it can be continued with the model selection.

Based on the argumentation of Piras and Buchenel (1999), there is no ultimate forecasting model of load demand, as a result of which it is necessary to analyze the predicting power of different modelling methods. The most widespread error measure index in the industry is the MAPE, because it captures the proportionality between the forecast error and the actual load. $MAPE = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{Y_t-\widehat{Y_t}}{Y_t}\right|$, where Y_t is the actual load data, while $\widehat{Y_t}$ is the forecast (Hahn et al, 2009). I apply this index in order to choose the most promising models among elaborated ones in the case studies. Moreover, I evaluate the forecasting performance of my own developed models based on this index as well.

In order to explain which model I am going to apply to my research question, the two mentioned model families will be reviewed in more detailed. First of all, the conventional statistical and regression based methods are covered. The two main subgroups of this family are the univariate and multivariate models. The former is used for very short term forecast and it is based on the historical load data, while the latter takes into consideration also exogenous variables (Hahn et al, 2009). Among the univariate models I will give a brief insight into the ARMA and exponential smoothing techniques, while with regard to the multivariate models I will present the SARIMAX, Principal Component Analyses (PCA) based regression, monolithic and system models.

The ARMA (p, q) models are one of the univariate time series approaches applied for load forecasting which are a (p, q)-th ordered mixed autoregressive and moving average processes can be described in the form if satisfy both the stationarity and the invertibility conditions. $\phi_p(L)Y_t = \theta_q(L)\varepsilon_t$, where Y_t is the load, ε_t is the white noise at time t, while $\phi_p(L)$ and $\theta_q(L)$ are the autoregressive and the moving lag polynomials. If the load time series is not stationary and contains unit root(s), then the appropriate extension of the model has to be applied, which is the ARIMA (p, d, q) specification, where d is the order of integrity of the time series, in other words the d-th difference of the series become stationary. In order to take into account the seasonality in load data the seasonal version of the AR(I)MA models or the periodic autoregressive models can be applied (Hahn et al, 2009). The other conventional univariate statistical methods are the exponential smoothing techniques. The simple exponential smoothing is used for short range forecasting, when there is no observed trend and seasonality in the time series, only the data fluctuates around a stable or slowly evolving mean (Chatfield and Yar, 1988). Consequently, the smoothed value is the weighted averages of the former values with more weight on the current values.

$S_t = \alpha Y_t + (1-\alpha)S_{t-1}$, where S stand for the smoothed values, while X for the real data and $\alpha > 0$. When there is a straightforward trend in the series, than the Holt's exponential smoothing method has to be employed (Chatfield and Yar, 1988).

$$S_t = \alpha Y_t + (1-\alpha)(S_{t-1} - T_{t-1}),\ \alpha \in [0,1]$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\alpha)T_{t-1},\ \gamma \in [0,1]$$

When seasonal pattern can be observed also in the time series then Holt-Winters exponential smoothing technique has to be put into force. If the electricity demand pursues a steady seasonal fluctuation then the additive version is the appropriate. Otherwise if the amplitude of the fluctuating seasonality is increasing then the multiplicative version is the suitable one. In connection with load forecast the multiplicative version is the relevant (Chatfield and Yar, 1988).

$$S_t = \alpha \frac{Y_{t-1}}{I_{t-s}} + (1-\alpha)(S_{t-1} - T_{t-1})$$
$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$$
$$I_t = \delta \frac{Y_t}{S_t} + (1-\delta)I_{t-s}$$

, where I_t is the seasonal factor and $\delta \in [0,1]$. However, most of the exponential smoothing models have the equivalence with (S)AR(I)MA models with the exception of the multiplicative Holt-Winters method. For example, the Simple Exponential Smoothing is similar to ARIMA(0,1,1) and Holt's Exponential Smoothing to ARIMA(0,2,2), while additive Holt-Winters exponential smoothing is corresponding to a SARIMA model (Chatfield and Yar, 1988).

In order to explain the application of the formerly explained univariate regression-based techniques for load forecast, I present the studies of Hippert et al (2005) and Taylor et al (2006). Hippert et al (2005) tested the performance of the different smoothing techniques, while Taylor et al (2006) compared the goodness of fit of double-seasonal ARMA models to the modified

Holt-Winters exponential smoothing method. One of the filtering models of Hippert et al (2005) was the application of 24 separate smoothing filters of Winters for each hour of the day, because hourly load series are seasonal as the consumption profile varies over weekdays and weekends. To eliminate the weekly seasonality as well, 168 separate filters can be introduced for each hours of the day but the drawback of this process is that forecasts based on one week old data (Hippert et al, 2005).The proposed solution of the authors was the combination of the two filtering techniques. As a result, not only the information which occurred a week ago was taken into account but also the information of the previous day and the within-week trend at the load forecast. Hence, load was defined as daily average and hourly deviation. The authors tested the forecasting performance of the Holt's filter and the simple exponential smoothing method. The elaborated technique did not outperform the simple exponential smoothing, because in short term the trend does not matter and the essential difference of these models is their trend prediction mechanism. (Hippert et al, 2005).

On the other hand, Taylor et al (2006) developed a double-seasonal ARIMA model to handle the seasonality. The general form of a double-seasonal ARIMA model is the following: $ARIMA(p,d,q)X(P_1,D_1,Q_1)_{s_1}X(P_2,D_2,Q_2)_{s_2}$, where P is the number of seasonal autoregressive terms, D stands for number of seasonal differences and Q represents the number of seasonal moving average terms. With respect to their hourly intraday Brazilian database $s_1 = 24$, while $s_2 = 168$. The authors applied the Box-Jenkins method for the development sample in order to figure out the most suitable SARIMA specification. Therefore, they studied the autocorrelation and partial autocorrelation functions to identify the order of the model. Taylor et al (2006) used the Schwartz-Bayesian Information Criteria to compare the fit of the different SARIMA models, furthermore at each case they tested the serial correlation of the residuals. The most appropriate specification of the Brazilian dataset was $ARIMA(3,0,3)X(3,0,3)_{24}X(3,0,3)_{168}$.

As it was mentioned earlier the only exponential smoothing method which cannot be substituted by ARIMA models is the multiplicative Holt-Winters exponential smoothing technique, therefore it is often tested which method has a stronger predicting power (Hahn et al, 2009). Taylor et al (2006) had to modify the standard Holt-Winters exponential smoothing formula in order to incorporate both the daily and weekly seasonal cycles. Hence, the authors introduced an additional seasonal index and an extra smoothing equation. Therefore the model contained a separate intraday and intraweek seasonal cycles beside the smooth and trend equations. The authors extended the forecast equation with an extra term to adjust the first-order autocorrelation (Taylor et al, 2006).

$$S_t = \alpha\left(\frac{y_t}{D_{t-s_1}W_{t-s_2}}\right) + (1-\alpha)(S_{t-1} + T_{t-1})$$
$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$$
$$D_t = \delta\left(\frac{y_t}{S_t W_{t-s_2}}\right) + (1-\delta)D_{t-s_1}$$
$$W_t = \omega\left(\frac{y_t}{S_t D_{t-s_1}}\right) + (1-\omega)W_{t-s_2}$$
$$\hat{y}_t(k) = (S_t + kT_t)D_{t-s_1+k}W_{t-s_2+k} + \phi^k(y_t - \left((S_{t-1} + T_{t-1})D_{t-s_1}W_{t-s_2}\right)),$$

where S stands for smoothing, T for capturing trend, D is responsible for the daily seasonality, while W for the weekly. The parameters (α, γ, δ, ω, ϕ) were estimated by minimizing the one-step-ahead sum of forecasted error squares (Taylor et al, 2006). Based on the result of the minimizing procedure, the optimal value of the parameters suggest that the load demand time series is dominated by seasonality and first-order autocorrelation, hence the weight of trend and smooth equations are low in the forecast (Taylor et al, 2006). Taylor et al (2006) regarded practical this exponential smoothing procedure because it does not require any model specification, therefore it is technically a simple method and give robust results in long term.

It is important to mention that both Hippert et al (2005) and Taylor et al (2006) smoothed out those atypical data, when idiosyncratic shock or special holiday occurred, because their main purpose was to model the routine days.

In forecasting studies it is a commonly used technique to create a naïve benchmark model to be able to evaluate not only the absolute but the relative forecasting power of the developed model as well (Hahn et al, 2009). Both Hippert et al (2005) and Taylor et al (2006) followed the consideration that the forecast of a given day can be predicted by the load of the previous day, which is resembled to Random Walk Theory. However, according to Hippert et al (2005) this procedure leads to a high forecast error after weekends due to the different load profile. Therefore, an improvement of this naïve benchmark is to take into account the type of a given day, whether that given day is a weekday or a weekend. From practical perspective it means that Monday load is forecasted by the electricity demand on the previous Friday, while Saturday load is forecasted by the demand on last Sunday (Hippert et al, 2005). Further fine-tuning of the benchmark model could be the consideration of the weekly periodicity of the electricity consumption pattern. Hence the load forecast of a given day would be based on the load value a week ago (Hippert et al, 2005). Taylor et al (2006) applied the latest approach of Hippert et al (2005), hence they also defined the predicted value as the value of the corresponding period of the previous week. $\widehat{Y(k)}_t$=$Y_{t+k-168}$ where k≤168 and it is the forecast lead time, while Y_t is the demand in period t. The authors regarded insensible to incorporate the daily periodicity as well.

The other subgroup of the classical regression based methods are the multivariate regression models. Compare to the univariate models when the forecasting horizon is wider, then it is advisable to use multivariate models, which considers the effect of exogenous variables as well.

The most relevant exogenous variables are the calendar and weather related information. The AR(I)MA models can be extended with exogenous variables, which are called AR(I)MAX or their seasonal version as SAR(I)MAX. Another regression based method is the application of the principal component analyses, which "…is a standard statistical method that is used for reducing the dimension of multivariate datasets, where variables are highly correlated, to a smaller set of variables. These are linear combinations of the original variables; they are uncorrelated and explain most of the variation in the data and are thus called principal components. In a regression context, one can therefore focus on this smaller number of independent variables, rather than dealing with the large number of original variables with complex interrelationships" (Taylor et al, 2006, p8.). Therefore, in the case of an enormous dataset a PCA based regression could be a more parsimonious solution than a (S)AR(I)MAX model. However it could be also possible that some of the influential explanatory variables are not observable or not available, then with the help of state-space models and Kalman-filter the forecast can be still reliable in terms of linear relationship between input and output data (Gary and Grag, 2001).

As an illustration of the application of the introduced multivariate methods, I present the paper of Taylor et al (2006), Ergün and Jun (2011) and Ramamnathan et al (1997). I selected these studies because they give an insight into different approaches and modelling routines. Taylor et al (2006) used the combination of regressions and PCA, Ergün and Jun (2011) applied a monolithic model with exact day-matching technique, while Ramanathan et al (1997) employed the local modelling approach. I give a detailed overview of the work of the authors, because I rely on the setup and experience of these studies during the development of my own model.

The other break-through of Taylor et al (2006) was the introduction of the PCA approach as an alternative of regression, beyond the exponential smoothing with double seasonal cycles method. First of all, the authors transformed the dataset into a Y_{ij} matrix, where i stands for the given days, while j for the hours, therefor j varies between 1 and 24 in the hourly Rio load data. Then C matrix was created in the following way: $C = (N-1)^{-1/2}Y^TY$. After it the eigenvalue decomposition was executed: $C = V\Omega^2V^T$, as a result $P = YV$ (Taylor et al, 2006). Therefore, the principal components were obtained by the basis transformation of the daily electricity demand. Only Q number of components with the strongest explanatory power were kept according to a given cut-off rate. However, the components could capture only the observed intraday pattern in the datasets. Hence, in order to handle the weekly seasonality as well, the components which belonged to the given days were regressed on dummies, representing the days of the weeks (Taylor et al, 2006). $p(i)_q = \alpha_{q_0} + \alpha_{q_1}d_1 + \cdots + \alpha_{q_6}d_6 + \beta_q i + \lambda_q i^2 + error(i)$, where $\beta_q i + \lambda_q i^2$ term captured the growth trend over the observed period. The forecasting formula was the following: $\widehat{Y_{i+1,j}} = \sum_{q=1}^{Q} p(i+1)_q\, V_{jq}$ (Taylor et al, 2006). However, the authors revealed serial correlation in the forecasted error term, therefore an additional AR process was added to the forecasting procedure, where the k step ahead forecasted error was predicted by the linear combination of the error occurred 24 hours ago and the last known prediction error. The weights were depended on the length of the forecast horizon. The formula of the error adjusting process: $E_t(k) = \alpha_0(k) + \alpha_1(k)E_{t-24}(k) + \alpha_2(k)E_{t-1}(1)$ (Taylor et al, 2006).

The second case study is interesting becasue Ergün and Jun (2011) built a simple monolithic multivariate model with a between estimator paneling technique as a treatment for seasonality. The main objective of Ergün and Jun (2011) was to develop an intuitive benchmark model for forecasting electricity demand, which can be used by non-expert industrial actors as well.

For that reason the authors applied the exact matching-day technique instead of difficult mathematical methods and seasonal adjusting procedures. They claimed that each day within a year has a unique demand, therefore the cycle of seasonality is a year. As a result, Ergün and Jun (2011) predicted the daily demand for electricity of a given day by taking the average demand of the same day in the former years. The available dataset for the authors was the daily system load of New England region between 2004 and 2008. One of the advantages of this construction is that there is no restriction on the form of seasonality but on the other hand it can still capture the calendar effect in a parsimonious way according to the argumentation of Ergün and Jun (2011).

This means that the solution of Ergün and Jun (2011) is equivalent of the introduction of 365 dummy variables for each day in each year to identify the varying electricity consumption pattern but by this way the computational time would significantly increase in line with the noise in the coefficient estimations. The previous N year matching-day average demand for electricity was defined in the following way in the paper (Ergün and Jun, 2011):

$$Mean_t = \frac{1}{N}\sum_{n=1}^{N} D_{t-n*364}$$

Based on the result of the studies dealing with STLF, the authors took into account the effect of weather condition, beside the calendar impact. As, the authors examined the relationship of temperature and daily demand for electricity, they identified the classic V-shape connection with turning point at 60.2°F. For simplicity the authors incorporated separate interacted terms of temperature above and below 60.2°F instead of introducing a higher order term to capture the non-linear effect of the temperature. The effect of those hidden factors, which did not change on daily level only over a year were also considered. Therefore, Ergün and Jun (2011) implemented another term in the demand forecast equation, which was responsible for capturing the systematic deviation of the expected value of the demand on a given day from the corresponding calculated mean value. The developed electricity load forecasting equation is the following (Ergün and Jun, 2011):

$$D_t = \beta_1 Mean_t + \beta_2 (D_{t-1} - Mean_t) + \beta_3 Temperature_t + \beta_4 Temperature_{t-1} + \varepsilon_t \quad ,$$

where $Temperature_t = \alpha_1 (Temperature_t - 60.2) d_1 + \alpha_2 (60.2 - Temperature_t) d_2$

$d_1 = 1 \; if \; Temperature_t > 60.2$°F, while $d_2 = 1 \; if \; Temperature_t < 60.2$°F

The significance of the regression-based method of Ergün and Jun (2011) that it parsimoniously and efficiently exploits the available information and it has an intuitive simple structure. However, it is worth to mention that one of the main drawback of this exact day-matching technique is that if the weekly seasonality is dominant in the given load dataset, then this weekly cycle will be smoothed via averaging, which leads to information loss and noisier, less accurate predictions (Ergün and Jun, 2011).

The third case study is the most advanced one among the introduced classic regression based studies, because the developed model of Ramanathan et al (1997) is already a multivariate model, which handles the seasonality with local modelling approach and dummy variables. Puget Sound Power & Light company organized a "forecasting competition" for which several teams of researchers were invited. The aim of the project was to model short run forecasts for hourly system load between 1983 and 1990 for winter and fall terms. Engle, Granger, Ramanathan, and Vahid-Arraghi formed a team and their model specification was called EGRV and the model specification is detailed in their Ramanathan et al (1997) study. The authors decided to follow a simple approach instead of nonparametric, time-varying parameter or general dynamic regression models. Therefore, they chose hour by hour modelling strategy, which means that hourly dataset was not considered as a single chronological ordered consecutive time series.

As the forecast was separated by hours, the authors specified 48 models because the weekdays and weekends were distinct. The general form of a forecast equation of Ramanathan et al (1997) was the following:

$Load_h = aDeterministic + bTemperature + cLoad + dPasterrors + \varepsilon$, where

- $Load_h$: the predicted demand for electricity in that given hour on any day
- $Deterministic$: predictable variables whose values are known in advance, like calendar variables, such as the year, day of the week, month. The authors specified a day after holiday variable to capture the deviant consumption pattern in transition between weekends or holidays and weekdays. All of these deterministic variables were dummy variables (Ramanathan et al, 1997)
- $Temperature$: it is a common fact that the relationship between temperature and load follows a nonlinear pattern but in this case, regard to the winter and fall terms, the connection is generally monotonic. As the temperature decreases, the demand for electricity is increasing. Although, it is not necessary to take into account the impact of cooling in the summer, the EGRV model did not neglect the nonlinear effect of temperature on load, hence squared variables were also introduced. Furthermore, the authors claimed that the effect of temperature is not constant over the months, therefore interacted terms were also incorporated into the model. Besides, the level and squared value of the maximum temperature of the given and previous days were also added to the forecasting equation as well as the moving average of past seven days midnight temperature in order to capture the effect of long cold spell. There was one-day-ahead weather forecast data available for the estimation (Ramanathan et al, 1997)
- $Load$8AM: As the forecast made at Puget at the beginning of the day at 8 am, Load8AM variable stood for representing the state of the system at time of forecasting. If the coefficient of this variable is close to 1 that would mean the validity of Random Walk

hypothesis. In that case the best prediction of tomorrow load is the today load. In order to take into account the pattern transition between weekends or holidays and weekdays, Load8AM*Monday and Load8AM*Dayafterholiday interacted variables were defined (Ramanathan et al, 1997)

- $Pasterrors$: Ramanathan et al (1997) assumed that missed factors have a recurring nature, as a consequence the forecast errors of the previous week were also considered in their model. But by this way the error terms of the model become heavily serially correlated. Therefore the authors applied the Cochrane-Orcutt autoregressive error structure at the estimation.

At the practical implementation and estimation of the model the authors had a dilemma. They considered the advantages and drawbacks of hour by hour estimation versus grouping the variables and create a seemingly unrelated regressions system (SUR). The main benefit of SUR is the parsimonious estimation process. On the other hand, the error terms in the error vector has to be contemporaneously uncorrelated in order to estimate a SUR model with OLS. However, in this case this requirement might not fulfill because a shock that hits the system has carry-over effect to the following hours as well. Besides, the hourly residuals of the consecutive days are also correlated due to the incorporated past errors. Moreover, the correlation between the first hour of a given day and the last hour of the previous day is highly correlated (Ramanathan et al, 1997). Furthermore, the reaction of load to the calendar and temperature effects is varying in the different hours, therefore the authors regarded it advisable to drop the insignificant variables and let slightly different specification in terms of each hour. By this way the coefficients varied in a systematic way hours by hours, moreover the model was more accurate, less multicollinear and less noisy (Ramanathan et al, 1997). In order to make the model more dynamic, Ramanathan et al (1997) developed an adaptive version of the EGRV model as well. This model was designed to systematic error correction.

They applied a simple exponential smoothing techniques, where the adaptive forecast was the linear combination of the original forecast and the adjusted error (Ramanathan et al, 1997).

$$\widetilde{y_{t+1}} - \widehat{y_{t+1}} = \tilde{y}_t - \hat{y}_t + \theta(y_t - \tilde{y}_t)$$

The drawback of this setup is a decrease in accuracy when the structure of the underlying dataset remains constant over the observed period. However in this case it performed well especially in connection with the weekend model (Ramanathan et al, 1997). The main critic of the model was that it is only a simple multiple regression model with numerous correlating variables without any economic related term. The argumentation of the authors was that their model was not based on economic theory, the main objective of the specification was to capture the short run behavior of households and industrial consumers' responses to weather conditions and calendar effects, moreover in short run the economic variables are unchanged and not influential (Ramanathan et al, 1997). Besides, the authors believed that in the case of a forecasting model the interpretation of the coefficients are not necessary and the multicollinearity does not hinder the forecasting ability of a model, the estimate remains consistent and unbiased. It is due to the fact that the coherence and connection among the explanatory variables sustains, only their effect could not be separated, therefore the estimation of their coefficient will be uncertain (Ramanathan et al, 1997).

To sum it up the main advantages of the classical regression based and statistical models that they easily can cover the connection between the input and output data, furthermore their results can be simply interpreted. However, the main drawbacks of these regression based models that they are suffered from numerical instabilities, they are unable to capture complex and non-linear relationships (Hahn et al, 2005).

Therefore, in line with the development of the computational technology, the authors from the beginning of 2000 turned to the artificial intelligence methods in case the of load forecasting. According to John McCarthy, the father of the artificial intelligence (AI) is "(t)he science and engineering of making intelligent machines, especially intelligent computer programs". The resounding success of these techniques is the ability to cope with theoretically poor but big and meaningful databases (Hahn et al, 2005). Moreover, they can reveal complex and non-linear relationship between input and output data, in a way to learn a set of rule via instances without defining a pre-specified form (Hahn et al, 2005). However, these techniques are time and capacity consuming, moreover at most cases they are black box models, which means that the results are impenetrable and hardly interpretable (Taylor et al, 2006). The most commonly used techniques in studies focusing on STLF are the: (1) logic based and (2) statistical learning algorithm, (3) percepton based techniques, (4) support vector regressions and (5) further hybrid procedures (Hahn et al, 2005). However, the AI solutions for STLF are rather belongs to the computational and operational research field than to economics, besides their modelling approach is not suitable for my research objective. Despite, I would like to give a brief overview about the main concepts of these methods and some real life forecast solutions, because most of the papers dealing with STLF apply these computational routines.

Among the logic based algorithm the decision tree is the most widespread, because it is the most interpretable and it has the highest transparency. Furthermore, this process is quite fast and robust to missing values and noise in the dataset. The procedure describes the potential range of the values of outputs in the function of the traits of the input set (Maglogiannis, 2007). Generally the structure of a decision tree consists of hierarchically ordered nodes and edges without any loop. Each node is a decision point, which stores a test function for incoming data.

The result of the test is binary and based on the outcome, the dataset is divided to two outgoing edges, called left child and right child. The final results are contained in the terminal nodes, in the so called leaves (Lahouar and Slama, 2015). For the development of the most accurate decision tree structure it is necessary to build a training and testing procedures. In the framework of this experimenting phase the algorithm tests the number of optimal nodes and the relevance of the decision questions belonged to the nodes based on the features of the input dataset (Lahouar and Slama, 2015). The process optimizing for the goodness of fit of the predictions to the training sample. In order to avoid an infinite cycle it is necessary to define the maximum number of leaves and a termination rule (Lahouar and Slama, 2015). The enhanced version of the decision tree is the random forest. The random forest, as it name suggests is the combination of the predictions of many decision trees (Lahouar and Slama, 2015). In the framework of this procedure there are q randomly selected training sample from the given dataset. For this q training samples, q predicting trees are developed according to the explained algorithm. Finally, the outputs of all these predictors are averaged (Lahouar and Slama, 2015). The added value of the random forest machine learning technique to the simple decision tree method is that by this way predictions are more immune to noise, because during the randomization of the different training sample uncorrelated trees were also generated. However, this technique performs well only with discrete features and simple decision structures, besides duplicates can be also occurred (Lahouar and Slama, 2015). An excellent instance for the use of random forest methods for STLF is the study of Lahouar and Slama (2015). The main objective of Lahouar and Slama (2015) was to develop a forecasting model for the electricity demand of Tunisia. The Tunisian Power Company provided a dataset of hourly load data between 2009 January 1 and 2014 August 31, the expected horizon of the forecast was a day-ahead prediction. The authors decided to design a forecasting model with the help of the random forest machine learning technique.

The consideration of Lahouar and Slama (2015) was based on the fact that the conventional statistical methods cannot handle the non-linear features of the load data time series, hence they turned to the more sophisticated, currently widespread artificial intelligence method. According to the argumentation of the authors among the machine learning techniques the random forest is the most suitable ones, because it is less sensitive to parameter values, more robust and resistant to irrelevant inputs, therefore it can cope with any load profile and complex consumption pattern. Furthermore this method provides a good interpretation as well. The decision tree defined by Lahouar and Slama (2015) for the Tunisia load dataset can be seen on the figure A1.1. Lahouar and Slama (2015) incorporated the months, day types, minimum and maximum temperature of weather forecast, morning and peak load of the previous day, furthermore electricity demand of the preceding 24 and 48 hours of the given hour into the model as input data for decision points based on the graphic analyses of the historical Tunisian load. Finally, the authors decided to build 24 separate modes for each hours in order to avoid the accumulation of forecasting errors. In this way the one-day-ahead forecast requires 24 times one-step-ahead forecast instead of a 24-steps-ahead forecast at a monolithic model, which is subjected to much larger error accumulation than the solution of the authors. Besides, the authors kept the poorly correlated input variables to increase the immunity of the algorithm. Furthermore, to be the model as long lasting as possible the authors chose the online version of this machine learning technique, which means a step by step broadening training sample base, because the previous day become part of the training sample for the next forecast. With this approach the model can handle changes emerged in the electricity demand pattern and consumers behavior (Lahouar and Slama, 2015). During the fine-tuning phase of the model the authors figured out that electricity demand on Mondays were consistently underestimated, therefore a new rule were added to forecast Monday load from the previous Friday instead of Sunday.

The model was also inaccurate in the case of the moving holidays and on the extreme hot days, as a consequence the algorithm was trained to use the former historical data belonged to these special days instead of the standard forecasting procedure (Lahouar and Slama, 2015).

The second subgroup of the most commonly used STLF AI techniques is the statistical learning algorithm. In the case of the statistical learning algorithm there is an explicit underlying model, which provides a probability that an example belongs to a given group (Maglogiannis, 2007). For example the Bayesian Network possesses a conditional probabilistic structure. The other well-known routine is the k-nearest neighbor method, which is based on the principle that "instances within a dataset will generally exist in close proximity to other instances that have similar properties" (Maglogiannis, 2007, p11.). The advantage of this procedure that it is fast and not memory consuming, however very sensitive to irrelevant inputs (Maglogiannis, 2007).

The most popular percepton based process is the artificial neural network (ANN), which is the third subgroup. It simulates the working of a human brain. The structure of the ANN model consists of nodes, which imitate the neurons of human brains and these are connected by links and interact with each other via these links (Maglogiannis, 2007). The nodes can take input data and execute simple manipulations on the data, while weights are assigned to the links. The learning procedure is developed by changing the weights in the sake of generating outputs fitting the most accurate way to the real values (Maglogiannis, 2007). It is the most frequently applied method for load forecasting, because it can model complex and multidimensional relationships, copes with multicollinearity and there is no need to understand the underlying data (Hahn et al, 2009). However, this method is prone to overfitting, it is a time and capacity consuming technique, furthermore the inside optimization process is a black box, hence the model is uninterpretable (Maglogiannis, 2007). For the realization of ANN in STLF Taylor et al (2006) is a good illustration. Taylor et al (2006) constructed a single hidden layer feedforward network, which consists of input data, one hidden layer and input data.

In this case the input data functions as explanatory variable, which was the lagged load demand data, while the output was the load demand of a given hour. During the training process the system optimized the following loss function (Taylor et al, 2006):

$$min_{v,w}\left(\frac{1}{n}\sum_{t=1}^{n}(y_t - f(x_t, v, w))^2 + \lambda_1 \sum_{j=0}^{m}\sum_{i=0}^{k} w_{ji}^2 + \lambda_2 \sum_{j=0}^{m} v_i^2\right), \text{ where}$$

- f: the resultant sigmoidal model
- v and w are the weights
- n: the number of in-sample observations
- k: units of inputs
- m: units in the hidden layer
- $\lambda_1$and λ_2 stand for penalization terms in order to avoid overfitting

Although, theoretically this technique should perform the best forecast by construction but in the case of Taylor et al (2006) it was outperformed by the classical regression based methods. The authors explained the poor performance of their artificial neural network with the fact that their dataset was short and they did not separate the weekdays and weekends.

The fourth subgroup is the support vector machine or regression, which is the state of the art machine learning procedure. This technique is "used for data classification and regression with non-linear kernel-based approaches, which means that instead of regressing in the (x,y) space, x is mapped into a higher-dimensional space by a mapping function in order to make the optimization process numerically easier and find the maximum distance among the separated group of input data" (Hahn et al, 2009, p4.). Like ANN this routine is also robust to noise and it can cope with complex and higher dimensional problems. From the other side, it is capacity consuming and parameters of the model is hard to interpret (Hahn et al, 2009).

Further hybrid computational techniques have already appeared in the literature, which main aim is to alloy the mentioned methods to overcome their weaknesses. These procedures are usually stochastic algorithm that try to find a good solution to a hard optimization problem by sampling the objective function, like particle swarm optimization and genetic, evolutionary algorithms (Hahn et al, 2009). These routines are applied to fine-tune the parametrization and the training period of a machine learning model or to help in the determination of the optimal setup of explanatory variables. These algorithms can find the global optimum and they are more resistant toward noise (Lahouar and Slama, 2015).

One case study for the utilization of hybrid technique is the paper of Zhang and Luo (2015). Zhang and Luo (2015) developed a hybrid method for short term load forecast of Jiangxi province in China. The authors combined the Gaussian Process machine learning technique with transfer learning procedure for this reason. Their underlying consideration was that most of the exogenous variables which influence the short term electricity demand are hard to obtain and difficult to quantify, however Gaussian Process can handle the latent variables well, as it is the only non-parametric machine learning technique and due to this feature it is less time and capacity consuming. Furthermore, based on the argumentation of the authors these mentioned hidden variables are resembled to each other within short distance area, therefore they came up with the idea to complete the Gaussian Process with knowledge transfer and apply the available dataset of the neighboring cities for the load forecast of the target city. The main assumption of the transfer learning technique is that two groups of tasks can be defined, target and source ones. Besides, the performance of the target task can be improved due to the exploitation of the accumulated information of the selected source tasks (Zhang and Luo, 2015). The notion of the authors is similar to the combination and the development of the Kalman-filter and k-nearest neighbor method. In this special case the selection of the source task is based on the similarity of the load profile of the cities in the province (Zhang and Luo, 2015). The available dataset of the authors consisted of the 15 minutes frequency load data of the 12 cities, which meant 16000 data points in case of each city. The power load prediction was taken as a random variable which was correlated with the previous load values.

The extent of the correlation was measured by a defined covariance function. The predictor of the load output sequence was the joint distribution of the predictor itself and the historical load data (Zhang and Luo, 2015). The structure of the model of Zhang and Luo (2015) can be observed on the figure A1.2.

Finally I summarize aspects and findings of the introduced studies, which are relevant related to the development of a customized forecasting model of the Hungarian system net load. First of all I chose the conventional regression based STLF methods and not the computational techniques, because the former is faster and more intuitive, therefore more suitable for a handy forecasting model. Then it can be stated that the biggest challenge of the conventional STLF methods is to cope with the superimposed seasonality of the power load, which is emerged on daily, weekly and yearly level. The most often used techniques in the covered studies for the elimination of the seasonality were the introduction of seasonal dummy variables or executing double seasonal differencing. The further important question was the consideration of the modelling approach beside the handle of the seasonality. The two approaches were the local and the global modelling approach. The former one used separate model for each hour or for weekdays and weekends, while the latter applied only a monolithic model, which treated the intraday load data as chronologically ordered consecutive time series.

The advantage of the local modelling approach is that it can easily handle the intraday seasonality, however it leads to the loss of interrelated information. On the other hand, in the case of a single model the one-day-ahead forecast means 24-steps-ahead forecast, which is resulted in the accumulation of the forecast error and in decreasing accuracy. Therefore, it is an ambiguous question that which approach should be employed for the Hungarian load forecasting model. Further important aspects of the models were the appropriate selection of the explanatory variables. In the short term forecasting horizon the economic variables are not elementary, because they do not change significantly during a short time interval.

However, the weather and calendar variables are very influential. The most important weather related variable is the temperature, which has a nonlinear connection with the electricity demand due to active heating in cold winter days and cooling in case of hot summer days. It is also important to consider the geographical fact that the measured or forecasted temperature of a given hour has a delayed effect on the system load, which is approximately three hours. Regarding to the calendar effects the most relevant experience of the papers were that the load profile is different on weekdays, weekends and holidays. Furthermore Mondays and Fridays behave as transition days and they are characterized by specific electricity consumption patterns. There was an agreement in the studies that the load profile of holidays is similar to the load profile on Sundays. Furthermore, some of the papers pointed out that the load on Mondays has to be forecasted from the previous Friday load, otherwise it would be consistently underestimated. However, it was not straightforward whether it is worth to be estimated a separate weekend model or forecasting Saturday from Friday does not cause consistent bias. In the next section I analyze the key features of the Hungarian system net load profile and check the relevance of the discoveries of the covered studies.

2. Data and Methodology

In this chapter I would like to reveal the key features and specificity of the Hungarian system load with the help of the database made available by the Hungarian Transmission System Operator (MAVIR). In the framework of the graphic analyses I present the evolution of the electricity usage over a year on monthly and daily frequency. Besides, I also give an insight into the connection of the weather factors and the system load. Furthermore, I investigate the atypical behavior of the electricity demand on national holidays and long weekends. Finally, according to the key findings of the graphic analyses, I explicate the main aspects of the chosen modelling approach and methods.

I got hourly net power system load data and meteorological data from MAVIR for 2014 and 2015, however the historical load data are also available on the website of MAVIR. From the terminological point of view the industry distinguishes gross from net system load. The difference between them is that net system load does not contain the self-consumption of the generators and further technical losses. As I have no data for controlling these phenomena, I decided to work with the system net load time series. The meteorological dataset, made available by MAVIR contains hourly day-ahead-forecasts of wind speed, humidity, light and temperature for 2014 and 2015. The dataset does not involve the winter holiday period (20th December to 5th January), because based on the industrial experience of MAVIR it is a very special part of the year, therefore forecasts for these weeks are made by a separate and special model. Furthermore, according to the advice of MAVIR some outlier days when sudden and unexpected issues happened such as big unforeseen sky tears were excluded. The consideration of the Hungarian TSO is in line with the international practice, because in the case of the covered literature the authors also smoothed out the outlier days.

As it can be observed on the charts A2.1 the classical W-shape yearly load pattern can be identified in relation to the Hungarian net load data in 2014 and in 2015. The electricity demand is decreasing as spring is coming, but in the middle of the summer it reaches its peak again due to cooling. After the summer dog days the load decreases until the beginning of the heating period. In 2015, this peak period was prolonged, besides it was more volatile due to various extreme weather conditions in summer and in September. In order to get a deeper insight into the nature of the periodicity of the system load, I executed monthly and daily aggregation in the function of hour. Based on the graphs A2.2 it can be concluded that the daily electricity demand pattern is the same in each month with parallel shifts. The highest load occurs during winter and in the middle of summer, while the lowest electricity consumption can be observed in spring. A slight difference can be noted between the two years with regard to the hierarchical order of average monthly load, especially in the case of summer, which is due to the different weather condition in the two years. Considering the daily run of the load, the global minimum is around 3-4 am, then a steep increase can be identified due to the "busy morning", such as the beginning of the morning shift and the morning schedule of the public transportation. After the start of the office hours, the electricity demand pursues a dampened growth up until the lunch break. The second local peak can be connected to the end of the business day and to sunset. As a result, this evening peak is varying over the months between 5-9 pm. The importance of this double peaks phenomenon in the intraday pattern was mentioned in the study of Labouar and Salma (2015) as well. With respect to the intraweek periodicity of the system load, the electricity demand on weekdays and weekends deviates in a significant way (see on A2.3). On Tuesdays, Wednesdays, Thursdays a similar pattern can be marked, however Friday and Monday behave in a different way. Until 3 am the Monday load is coincided with the Sunday load and during the consecutive hours a slow catching up effect can be detected. On the other hand, the deviation from the weekdays load figure starts from Friday afternoon.

Furthermore, there is a distinct load profile even on Saturdays and on Sundays. Saturdays behave similar to the weekdays until 2 am, while on Sundays the demand is lower than on Saturdays. Besides, weekends and weekdays have different evening peaks as well: the former has it around 7 pm, while the latter at 8 pm. The correlation matrix of the net load and the calendar variables also confirms the former findings because the days of the week have the strongest relationship with net load (see on A2.4). Beside the days of the week, the national holidays are also highly correlated with the net load. Therefore, I conduct further analyses focusing on the impact of holidays on net system load.

According to the studied literature further special behavior pattern can be noticed in terms of holidays compare to simple weekdays or weekends. In Hungary it is a usual practice that the holidays are completed with an extra day for the sake of long weekends. Of course, the additional holiday has to be worked off on a given Saturday. As a first step, it is important to examine the relationship of the calendar variables with the net load from dynamic perspective as well. Therefore, I study in detail the cross-correlation functions of net load and the calendar variables. The charts A2.5 show the strength of the daily connections. The calendar variables change only on daily level, furthermore their lead values are also known. As it can be noticed on the cross-correlograms A2.5, the working Saturday has almost no impact on the net load, only its 6th lag is significantly different from 0. However, the current, previous and consecutive values of the "National holiday" variable are influential. In order to understand the spillover effect of the national holidays and working Saturdays, I analyze some special cases. Six types of holiday can be differentiated: (1) a long weekend consists of four days including Thursday and Friday, (2) a long weekend consists of four days including Monday and Tuesday, (3) a long weekend consists of three days including Friday, (4) a long weekend consists of three days including Monday, (5) single holiday and (6) working Saturdays. Although my sample is too short to draw general consequences, however some interesting observations can be made based on the investigation of some special cases. The main findings are summarized in the following table:

Type of holiday/Influences	Given day	Further deviant behavior
single day	2014.08.20 – Wednesday (A2.6) • Until 2am the pattern was similar to the previous day • The global minima shifted from 3 to 6 am • During the day load profile showed the same pattern with a 25% downward shift	The demand on Thursday and Tuesday was lower than on the following week, especially in the case of Thursday. Moreover, the dawn of Thursday still resembled to a holiday load profile
	2014.11.01 – Saturday (A2.7) • The figure of the load profile was the same with a downward shift compare to the following Saturday	The load on Friday resembled to the forthcoming week, in case of Sunday the same tendency could be observed, but the dawn was still like a holiday pattern
	2015.11.01 – Sunday (A2.8) • Pattern remained the same with a substantial downward shift versus forthcoming Sunday	Both in case of Monday and Saturday the demand was lower than a week later, especially in the middle of the day. Moreover, the dawn of Monday acted like a holiday
3 days with Friday	2015.05.01 and 2015.10.23 - (A2.9) • The dawn of Friday until 2 am behaved like other Fridays but then during the whole day the demand was significantly lower. Besides, there was a time shift in case of the global minima as well	In case of Thursday from the middle of the day some deviation can be observed in the load profile versus other Thursday. The dawn of Saturday acted like a holiday and the total load profile slightly differentiated from the usual ones.
3 days with Monday	2014.04.20 and 2015.04.05 - (A2.10) • The same trend was observed as in case of the other 3-days long weekend	Besides the same trend as in the former case. When this long weekend was at Eastern the previous weekdays and the following Tuesday slightly deviated the common profile which perhaps due to the spring school holiday
4 days with Thursday and Friday	2014.10.23 and 2015.08.20 - (A2.11) • The formerly noticed phenomena were valid in this case as well, hence dawn behave like on normal days,	The electricity demand on the afternoon of the previous day was already lower. The dawn of the following day of the holiday

	then time shift in the global minima and during the day downward shift in the load profile	was still as low as during holiday. The extra day mostly behaved like a Saturday
Working Saturday	2014.10.18 and 2015.10.18 - (A2.12) • The profile was resembled to Fridays	Sunday showed similar pattern to normal Saturdays

1. Table - Spillover effects of holidays in 2014-2015

Unfortunately, there was no 4-day-long weekend, where Tuesday and Monday would have been the holidays, therefore I could not reveal the specificity of this type of holiday. Although my time series is too short to draw general rules with regard to the specific load profile on holidays, however the main findings of table 1 could be built into the model as a rule-of-thumb:

- The demand on holiday is roughly 25% lower than otherwise would be
- The load profile of holiday at dawn is similar to its standard daily value until 2 am
- There is a time shift with regard to the global minimum of electricity usage on holidays
- Some spillover effect after a holiday can be noticed, because the load of the previous day shows a decreasing tendency in the afternoon and evening versus the usual value of that day and hour. Furthermore, the load is also as low as it was on the holiday at dawn of the following day
- In the case of working Saturdays the electricity demand is as high as on Fridays, as a result the Sunday profile acts similar to a normal Saturday

After the deep analyses of the influence of the calendar variables, I would like to discuss the relevant impacts of the weather related variables on the system net load. As I have mentioned, the given meteorological dataset contained one-day-ahead forecast values for temperature, humidity, illumination and wind speed on average for the whole country. According to the correlation matrix of the net load and the weather variables, temperature and illumination have the largest impact on net load (see on A2.13). However, there is a strong negative correlation between illumination and temperature, moreover humidity also highly correlates with illumination and temperature. However, it is essential from the modelling point of view to scrutinize the exact relationship between net load and the weather variables. Based on the empirical results of the covered studies, the most influential weather variable is the temperature, which is in a non-linear relationship with the load due to the heating activity in winter and cooling activity during the summer. As it can be observed on the chart A2.14 a U-shape pattern is valid in the case of Hungary, but the dispersion of the data points is quite high. A further relevant remark is that the turning point is around 290°Kelvin. As for the forecasted wind speed, no strong polynomial connection to load could be identified (see on A2.15). Furthermore, the illumination is in a negative linear relationship with load during the winter below 20 klux, while the humidity is in a positive linear relationship with the load above 80%, however these connections are not strong (see on A2.16 and A2.17). Beside the static graphic analyses I study in detail again the cross-correlation functions of net load and the weather variables, to reveal the strength of the hourly and the daily connections (see on A2.18-A2.25). It is important to keep in mind that only one-day-ahead weather forecast values are available for the Hungarian TSO, hence only the information about the lag values can be utilized during the modelling. In the case of temperature, the formerly revealed strong correlation can be identified from dynamic prospect as well. Temperature on the past 7 days and the past 8-18 hours have significant impact on the net load. With regard to humidity and illumination, the past 7 days also count. However, with reference to illumination past 1-7 lags have positive, while the past 12-20 lags have negative effect on net load within the day. A reverse pattern can be observed in relation to humidity, where within day the past 1-4 lags have negative and the past 9-18 lags have positive influence on net load, however the correlation is weak. Furthermore, in line with the key remarks of the correlation matrix, wind speed is not a significant factor neither on hourly, nor on daily level. Based on the findings of the graphic analyses of the weather related factors, it can be concluded that the weather variables have strong interday and intraday spillover impact on net load. Furthermore, beside temperature humidity and illumination also have to be built into the model, especially at the identified relevant periods.

After the exploration of the key traits of the Hungarian system load dataset in 2014 and 2015, as a next step the concept of the modelling have to be set in line with the studied papers of the literature review part. As I have already emphasized, the main aim of my thesis is to develop a useful forecasting model for the Hungarian TSO. For this reason the model should have a transparent structure with meaningful explanatory variables, with intuitive coefficients and results. These considerations are important because the dispatchers in the head offices have to understand which factors are the most essential from the operation point of view and toward which exogenous effect is the system the weakest. Furthermore, sometimes the industrial officers have to make expertise judgement and conduct modifications in the forecasts by hand. As a consequence, I choose to use the conventional regression based methods, because these are the most suitable modelling approaches to my project because the new machine learning techniques require outstanding programming knowledge and computational capacity, special software license which are not necessarily available at MAVIR.

The most widespread regression based technique is the ordinary least squares estimation (OLS) but this estimator requires numerous conditions to be consistent. Therefore, I check the validity of the OLS requirements as for my dataset. First of all, I test the stationarity of the time series with the help of the ADF test (see A2.26). None of the variables contain unit root, as a result of which cointegration is not an issue. Besides, there is no threat of perfect multicollinearity among the calendar and meteorological variables based on the correlation matrices. In addition, as it was formerly argued by Ramanathan et al (1997), multicollinearity does not hinder the forecasting power, only an obstacle to interpreting the effect of the explanatory variables. Furthermore, I also consider the presence of simultaneity because developing a forecasting model for system load is related to some extent to a demand-supply estimation problem.

There is no simultaneity in the case of power system load estimation, because it was shown in the literature that in the short term price has no influence on demand. It is due to the fact that excessive demand or supply has to be immediately counterbalanced at any price, besides price does not change within day at most cases. In the case of Hungary the price cannot change within the day because the Hungarian power market is a half-regulated market. For the residential consumer the electricity price is time-invariant and set by the authority with yearly revision. There is free market for the industrial consumers but in their contracts the cost of the electricity supply is also fixed in advance. Dynamic electricity price is present only on the power exchange market, although here can be only day-ahead and future products traded which means that the current disturbance of the system will not influence the current price. Finally, I take into account the occurrence of endogeneity. One of the main sources of endogeneity could be the simultaneity but as I explained there is a low probability of simultaneity in the current case. Omitted variables could be a further potential source of endogeneity but it is also not an issue in the current case because all of the relevant explanatory variables, which were mentioned in the literature, are available and can be incorporated into the analysis. Therefore I exclude the presence of endogeneity as well. After the justification of the necessary conditions for OLS, based on the main findings of the graphic analyses I regard the model specification of Taylor et al (2006), Ergün and Jun (2011) and Ramanathan et al (1997) relevant, because these authors applied classical regression based techniques specified to capturing the impact of weather and calendar variables on system load. Therefore as a starting point I will experiment with the replication of the mentioned studies before the creation of my own model. I will estimate the model for 2014 and tested its forecasting performance on 2015 based on the MAPE index.

3. Replication

In this chapter I will test the fit of the model specification of Taylor et al (2006), Ergün and Jun (2011) and Ramanathan et al (1997) to the Hungarian system net load dataset for 2014 and 2015. The object of the replication is to reveal those approaches and modelling techniques which can capture the specific pattern of the Hungarian net load dataset. Moreover, I would like to recognize the deficiencies that cannot be solved in the framework of the mentioned model specifications and by this way mean a field for development. As a result, I will carry out the original model specification of each study, then I will check the correctness of the models with the help of statistical tests (Durbin-Watson test, Breusch-Godfrey Serial Correlation LM Test, RESET test and Heteroscedasticity test), after if it is necessary I will execute amendments in order to develop a statistically correct model. Finally I will test the forecasting performance of the established model specifications and summarize the findings. In logical order I will demonstrate the univariate naïve benchmark model, the double-seasonal ARIMA model and the exponential smoothing with double seasonal cycles technique of Taylor et al (2006), then I will present the model specification of Ergun and Jun(2011) and Ramanathan et al (1997) among the multivariate models.

3.1 Naïve benchmark model

In the case of the Hungarian power system net load data, the presence of unit root was rejected by ADF tests (see on A3.1). Therefore, instead of seasonal random walk model specification, a seasonal version of an AR process can be still applied as a naïve benchmark model. First of all I ran the hourly net load data on their corresponding week ago value (see on A3.2). Although the week ago hourly electricity demand seemed to be a significant explanatory variable, however the Durbin-Watson statistics and LM-test proved the presence of serial correlation (see on A3.3).

Therefore, I applied the error correction mechanism used by Taylor et al (2006) at their PCA models, thus I incorporated the last prediction error and the one, which occurred 24 hours ago (see on A3.4). However, this modification could not solve still the problem of the autocorrelated error terms (see on A3.5). Then with the help of the ACF graph I tried to identify the relevant MA terms, which were the errors of the last 1.5 day (see on A3.6). In the final model specification, the following terms remained significant:

Variable	Coefficient	P-value
C	2537.692	0.0000
NET_LOAD(-168)	0.859835	0.0000
MA(1)	1.518117	0.0000
MA(2)	1.601950	0.0000
MA(3)	1.492930	0.0000
MA(4)	1.364681	0.0000
MA(5)	1.214870	0.0000
MA(6)	1.084793	0.0000
MA(7)	0.923570	0.0000
MA(8)	0.745217	0.0000
MA(9)	0.593087	0.0000
MA(10)	0.447893	0.0000
MA(11)	0.285762	0.0000
MA(12)	0.111064	0.0000
MA(24)	0.385616	0.0000
MA(25)	0.593214	0.0000
MA(26)	0.599864	0.0000
MA(27)	0.541596	0.0000
MA(28)	0.490606	0.0000
MA(29)	0.420439	0.0000
MA(30)	0.313818	0.0000
MA(31)	0.186045	0.0000
MA(32)	0.064765	0.0000
R-squared	0.993547	
Adjusted R-squared	0.993529	
Akaike information criterion	13.52342	
Durbin-Watson stat	1.957238	

2. Table - The first correct specification of the naïve benchmark model

As it can be observed in Table 2, the test statistics improved significantly but the model still suffered from serial correlation (p-value of LM-test F-statistics was 0). Therefore my conclusion was that in the case of the Hungarian load data, it is not enough to take into account only the weekly cycle. In order to get a statistically correct model, it is necessary to get rid of the intraday periodicity. As a result, I re-estimated the original setup of Taylor et al (2006) on the daily seasonally differenced dataset (see on A3.7). However, the residuals of the estimates remained serially correlated in the case of the added error correction process as well, even when I further extended the original model and incorporated the past errors of the last 1.5 day (see on A3.8). Therefore I relaxed the assumption of Taylor et al (2006) and beside the week ago corresponding load demand I also included further lags. The additional lags solved the problem of serial correlation, but could not eliminate the heteroscedasticity (see on A3.9). Hence, I used the White heteroscedasticity-consistent standard errors in order to get a consistent estimate. However, the drawback of the White heteroscedasticity-consistent standard errors is that it complicates the calculation of the confidence intervals, furthermore it increases the forecast errors. I tested whether the explanatory variables of the final model specification remained significant after the use of the White heteroscedasticity-consistent standard errors (see on A3.10). Furthermore, I tested the validity of the correct model specification with the RESET-test (see on A3.11). Based on the test results, the best model specification was already statistically valid with the application of the White heteroscedasticity-consistent standard errors, because its error terms were not serially correlated, the explanatory variables were jointly significant, and the RESET-test could not undermine the correctness of the specification.

Variable	Coefficient	P-value
D24_NET_LOAD(-167)	0.226744	0.0000
D24_NET_LOAD(-168)	-0.216492	0.0000
D24_NET_LOAD(-1)	2.508853	0.0000
D24_NET_LOAD(-2)	-3.140097	0.0000
D24_NET_LOAD(-3)	2.842575	0.0000
D24_NET_LOAD(-4)	-1.876450	0.0000
D24_NET_LOAD(-5)	0.829308	0.0000
D24_NET_LOAD(-6)	-0.180802	0.0000
D24_NET_LOAD(-23)	0.023930	0.0000
D24_NET_LOAD(-26)	-0.092504	0.0000
D24_NET_LOAD(-27)	0.164791	0.0000
D24_NET_LOAD(-28)	-0.172303	0.0000
D24_NET_LOAD(-29)	0.122430	0.0000
D24_NET_LOAD(-30)	-0.048349	0.0000
MA(1)	-1.060635	0.0000
MA(2)	0.957771	0.0000
MA(3)	-0.618984	0.0000
MA(4)	0.196859	0.0000
MA(6)	0.062944	0.0005
MA(7)	-0.062221	0.0000
MA(8)	0.047652	0.0000
MA(11)	0.058931	0.0000
MA(24)	-0.772072	0.0000
MA(25)	0.753567	0.0000
MA(26)	-0.695775	0.0000
MA(27)	0.450985	0.0000
MA(28)	-0.136039	0.0000
MA(30)	-0.044159	0.0013
R-squared	0.981754	
Adjusted R-squared	0.981692	
Akaike information criterion	13.35742	
Durbin-Watson stat	1.997533	

3. Table - The second correct specification of the naïve benchmark model

Finally I tested the forecasting power of the naïve benchmark model to set a reference point for the evaluation of the further model specifications. I executed a static one-step-ahead forecasting procedure. As it can be observed on the charts A3.12 the naïve benchmark model is extremely imprecise as the average MAPE is 107%, while in Taylor et al (2006) it was around 4.5%.

The biggest errors occurred on Thursdays, on Fridays, in spring and in autumn. The Chow forecast test also signaled that this specification cannot handle well the changing load profile in autumn (see on A3.13). Therefore, it can be concluded that this model specification is neither statistically robust, nor accurate from the forecasting point of view.

To sum it up, the Hungarian net load dataset does not contain unit root, therefore the seasonal version of the Random Walk model was not appropriate as a naïve benchmark model. As a result, I estimated an AR process based on the consideration of Taylor et al (2006), hence I used the week ago hourly net load data as an explanatory variable. However, the model was serially correlated even with respect to the application of the error correction procedure developed by Taylor et al (2006). The problem could not be eliminated with the extension of the model with further past errors. To resolve the intraday periodicity, I executed the 24 hour seasonal differencing and re-estimated the former models, but the autocorrelated error terms did not disappear. Finally, I relaxed the hypothesis of Taylor et al (2006) and incorporated further past net load data beside the week ago based on the PACF and ACF graphic analyses. After the development phase I also tested the forecasting performance of the model, which was very weak especially when seasons changed. The main conclusion was that in the case of the Hungarian electricity demand, the intraday seasonality is very substantial and not only the week ago demand is relevant but also the electricity need during the past 1.5 day.

3.2 Double-seasonal ARIMA model

In the case of the Hungarian system net load dataset, I also followed the Box-Jenkins methodology to find the most suitable SARIMA specification. In order to identify the lags I plotted PACF and ACF of net load. ACF showed a seasonality in spite of the fact that the series was double seasonal differenced but PACF broke down after 3 lags (see on A3.14). As a result, the AR part consisted of 3 lags, while the number of relevant MA terms is ambiguous hence I was experimenting with more lags. Even though I set the Newey-West seasonal error correction, the 24^{th} and 168^{th} lags of the residuals of the estimated model remained significant. Consequently, I tried to incorporate these lags of the MA and AR terms in the regression (see on A3.15). However, the Durbin-Watson statistic, the Serial Correlation LM-test and the residual diagnostics (autocorrelation, normality test) proved the presence of serial correlation (see on A3.16). Therefore, I doubted the correct specification of the model, as a result of which I executed the RESET test, but it could not confirm the misspecification of the model (see on A3.17). Hence, I thought that maybe some structural break cause the problem. Therefore, I tested the Chow Breakpoint Test for the 2400^{th} hour and the 5300^{th} hour, which are the middle of spring and the beginning of autumn. The null hypothesis could have been rejected in both cases, so there are structural breaks in the relationship (see on A3.18). As a next step I checked the same model specification for the subsamples. The results were not unanimous:

- 1-2400th hours: serial correlation was not rejected by the Durbin-Watson test statistic and the Serial Correlation LM test, according to the RESET test the model was misspecified
- 2400-5300th hours: serial correlation could not be eliminated but the RESET test null hypothesis remained significant
- 5300-8136th hours: the model remained well specified and serial correlation could be eliminated according to the tests

Finally, I resolved the restriction of the SARMA specification. Based on the result of the development of the naïve benchmark model, I reran the model on the lags and past errors of the past 1.5 day. The final specification passed the Durbin-Watson, LM-test, RESET-test but among the different heteroscedasticity tests only the Harvey-test could undermine the presence of heteroscedasticity, therefore I applied again the White heteroscedasticity-consistent standard errors and checked again the joint significance of the variables of the last specification with Wald-test (see on A3.19).

Variable	Coefficient	P-value
DTOT_NET_LOAD(-1)	1.517865	0.0000
DTOT_NET_LOAD(-2)	-0.446246	0.0000
DTOT_NET_LOAD(-3)	-0.155827	0.0001
DTOT_NET_LOAD(-4)	0.077533	0.0000
DTOT_NET_LOAD(-168)	-0.305443	0.0000
DTOT_NET_LOAD(-169)	0.482028	0.0000
DTOT_NET_LOAD(-170)	-0.180587	0.0000
DTOT_NET_LOAD(-23)	0.114546	0.0000
DTOT_NET_LOAD(-11)	-0.014418	0.0010
DTOT_NET_LOAD(-18)	-0.068195	0.0026
DTOT_NET_LOAD(-19)	0.053116	0.0086
DTOT_NET_LOAD(-16)	0.021063	0.0078
DTOT_NET_LOAD(-25)	-0.177737	0.0000
DTOT_NET_LOAD(-27)	0.062009	0.0000
MA(23)	-0.124965	0.0000
MA(167)	0.086158	0.0000
MA(168)	-0.103925	0.0000
MA(1)	-0.046213	0.0000
MA(2)	-0.291605	0.0000
MA(3)	-0.102289	0.0000
MA(6)	0.056405	0.0233
MA(7)	-0.012679	0.0340
MA(24)	-0.797394	0.0000
MA(26)	0.260387	0.0000
MA(27)	0.124648	0.0000
MA(30)	-0.049571	0.0430
MA(9)	0.011507	0.0037
MA(12)	-0.005468	0.0304
R-squared	0.975551	

Adjusted R-squared	0.975466
Akaike information criterion	13.33888
Durbin-Watson stat	1.993292

4. Table - The final correct specification of the "double-seasonal ARMA"

After the specification phase I tested again the forecasting performance of the final version. As it can be noted on the charts A3.20, the accuracy of the model is extremely weak even worse than the naïve benchmark. The average MAPE of the one-step-ahead forecast is 175%, while the double-seasonal ARIMA model of Taylor et al (2006) had 2.3% MAPE. The highest error occurred on Mondays, in April and in September. The Chow forecast test signaled once more a structural break at the beginning of autumn.

All in all, even with double seasonal differencing the lags of the 1.5 days and MA terms had to be incorporated into the model because the double seasonal differencing could not solve the problem of daily and weekly periodicity and the error terms remained serially correlated. As a consequence, the final specification was similar to the naïve benchmark model, even though it was outperformed by the naïve benchmark model. It implicates that in this case double seasonal differencing led to information loss and turned out to be counterproductive.

3.3 Exponential smoothing with double seasonal cycles

At the beginning it is important to mention that I executed the replication of the exponential smoothing model of Taylor et al (2006) with a special dwhs() forecast program in R software which was developed according to the Taylor (2003) process. Consequently, it was a black box procedure, hence I did not conduct any adjustment on the original setup. Therefore, the forecasting performance of the technique can be immediately discussed. As it can be observed on the chart A3.22, the forecasted values fit almost perfectly the real dataset. The average MAPE of the one-step-ahead forecast was 0.89%, while it varied in the range between 0.6% and 1.25% (see on A3.23).

The result is not unexpected, because with reference to Taylor et al (2006), the exponential smoothing with double seasonal cycles performed the best. It is also interesting to mention that the least predictable months are April, May and September, when the weather conditions are usually the most variable, and at the turn of seasons. From the daily aggregation point of view, the method was the least accurate on Mondays and weekends, which are special days according to the studies. These days have special consumption patterns, especially on national holidays and long weekends. Finally, I checked the correctness of the model, therefore I tested the ACF and PACF of the residuals of the forecast. Based on the diagrams A3.24, it can be concluded that despite the incorporation of the intraday and intraweek seasonal cycles, the residuals of the Holt-Winters exponential smoothing with double cycles remained highly autocorrelated, therefore the model is statistically not correct.

Based on the results, I argue that the univariate techniques are not appropriate for the forecasting of the Hungarian net load, because these methods cannot handle the influence of the weather conditions and the varying residential and industrial consumption patterns in the function of the weekdays and holidays. Consequently, the univariate models missed relevant information and in spite of the incorporation of the past values, they produced inaccurate predictions. Hence, I continued with the testing of the forecasting performance of the multivariate models, Ergün and Jun (2011) and Ramanthan et al (1997).

3.4 Model with exact-day matching technique

At the beginning it is necessary to declare that the authors worked with daily data and the Hungarian system net load database consists of hourly data. According to the argumentation of the authors the yearly periodicity counts, as a result their model specification should also work on the hourly level as well. Historical load data are publicly available on the website of the Hungarian Transmission System Operator, therefore I could calculate the average demand of a given hour on a given day from 2010 up to 2015, and execute the exact-day matching technique on the hourly level. Regarding the specification of the temperature variable, I relied on the findings of the graphic analyses, hence I declared 290°K as the turning point. As a result $d_1 = 1\ if\ Temperature_t > 290$°K, while $d_2 = 1\ if\ Temperature_t < 290$°K

After the generation of the variables defined in the study, I tested the goodness of fit of the model of Ergün and Jun (2011) on the Hungarian dataset (see on A3.25). The lags of the temperature and the increasing interval of the temperature function were not significant, therefore I made sure with F-test whether these variables are jointly significant or not (see on A3.26). The lags of the temperature were not significant jointly, hence I dropped them from the equation and re-estimated the model (see on A3.27). The variables were strongly significant, moreover the R-square was high but the Durbin-Watson statistics and the serial correlation LM test indicated autocorrelation in the residuals (see on A3.28). Therefore, I tried to adjust the model with the application of the Newey-West seasonal error correction, but it could not eliminate the identified autocorrelation. After the check of the ACF and PACF of the residuals, it can be concluded that the first three lags of the net load variable are deviated the highest extent from 0, hence these should be incorporated into the model (see on A3.29). Although, all of the lags were significant, they could not solve the problem of serial correlation. Therefore, the correctness of the model was doubtful, as a result of which I executed the RESET test which could confirm the misspecification of the model (see on A3.30). Therefore, I hypothesized that the model is rather sensitive to the mean or the demeaned net load variables. Consequently, I was experimenting with the first 3 lags of demeaned net load, mean net load and simple net load. In order to avoid multicollinearity, only two out of three mentioned variables were incorporated into the equations. I applied LSE type of modelling, put all the relevant variables in the equation, then drop the most insignificant ones, after that checking F-test statistics whether these are jointly insignificant as well. The first priority was to develop a statistically correct model, therefore in the case of each specification I executed the serial correlation LM test and tested the Durbin-Watson statistic as well as heteroscedasticity. My secondary objective was to reach the highest explanatory power of the developed model based on the Schwartz and Akaike information criteria (AIC and BIC values). At the end of the iteration the best specification could not solve still the problem of autocorrelation in the residuals based on the LM test, however the Durbin-Watson statistics was already convincing (see on A3.31). Therefore, I plotted again the ACF and PACF of the residuals, the 24th lag remained highly correlated (see on A3.32). With the incorporation of the 24th lag, the statistics of the model improved. The LM-test could undermine the presence of serial correlation (see on A3.33). I do not think that this result is robust, therefore my final conclusion is that the original model does not fit the Hungarian dataset because it cannot handle the intraday seasonality.

As a consequence, I decided to develop further the model of Ergün and Jun (2011) and figure out the way to handle the intraday seasonality. My first attempt was to conduct seasonal differencing in order to get rid of the intraday periodicity and then reran the original model specification. The seasonal differencing did not solve the problem of the autocorrelation in the error terms, but the first lag of the temperature variable turned partly significant (see on A3.34). I checked again the plot of the ACF and PACF of the residuals and repeatedly the first three lags were significant at the highest extent. Therefore I started again the LSE-type of iteration with all combinations of the three lags of the net load, mean load and demeaned load variables taking into account the possibility of multicollinearity. The best performing specification was still serially correlated based on the LM-test, although the Durbin-Watson statistics was again promising. After the check of the ACF and PACF the 6th lag of the residuals showed high partial autocorrelation, as a result of which I decided to extend the model with those lags as well. Finally with the exclusion of the 4th lag, all of the explanatory variables remained significant. Besides, both the Durbin-Watson statistics and the LM-test rejected the presence of serial correlation but the model was heteroskedastic according to the test result, therefore instead of the Nevey-West correction I applied the White heteroscedasticity-consistent standard error correction, with the check of the joint significance of the variables in the latest model with Wald-test. The model is still not robust, but at least theoretically gives a consistent estimate.

Variable	**Coefficient**	**P-value**
D24_MEAN_NET_LOAD	-0.156932	0.0003
D24_DEMEAN_NET_LOAD(-1)	-0.133546	0.0009
D24_TEMP_V*D2	19.30455	0.0006
D24_TEMP_V(-1)*D2	-17.96129	0.0016
D24_TEMP_V(-1)*D1	5.008207	0.0003
D24_NET_LOAD(-1)	1.761574	0.0000
D24_NET_LOAD(-2)	-0.919797	0.0000
D24_NET_LOAD(-3)	0.240234	0.0000
D24_NET_LOAD(-5)	0.044787	0.0003
D24_NET_LOAD(-6)	-0.020762	0.0260
R-squared	0.968470	
Adjusted R-squared	0.968435	
Akaike information criterion	13.89841	
Durbin-Watson stat	2.000854	

5. Table - The first correct specification of the model with exact-day matching

Consequently, I experimented with another approach to handle the intraday seasonality. Instead of daily seasonal differencing, I introduced the days of the week in a form of dummy variables. First of all, I ran the original model specification of Ergün and Jun (2011) with the incorporation of the day dummies, taking into account the possibility of multicollinearity (see on 3.35). The lags of temperature variables and Tuesday, Thursday dummies were not significant, thus I tested with F-statistics whether they are jointly insignificant as well (see on 3.36). After the test proved their jointly insignificance, they could have been removed from the equation. However, after the re-estimation of the model, in spite of the significant explanatory variables and the high R^2, the Durbin-Watson statistics and the LM test proved the presence of serial correlation in the residuals (see on 3.37). In order to identify which left out lags caused the serial correlation, I plotted the ACF and PACF (see on 3.38). Again the problem was with the first 6^{th} and 24^{th} lags.

However, the extension of the model with these variables could not solve the problem of serial correlation. Therefore, as a next attempt I took the interaction of the 23^{th} – 25^{th} lags and the day dummies and re-estimated the last specification. Even though, in the case of the most appropriate version, when all of the variable were significant and the BIC, AIC was the lowest, the Durbin-Watson statistics and LM-test indicated the presence of serial correlation (see on 3.39). My final guess was to eliminate the autocorrelation in the residuals with the incorporation of MA terms. My idea was based on the consideration that the evolution of the load data within day might be due to the specific conditions of that given hours, thus the lags of the different load variables are not as strong indicators as the current or former conditions. Only the day dummies and the temperature variables signal the conditions of the given hours, as a result of which it can be assumed that there are some latent variables which are embedded into the error terms and cause systematic deviation in the prediction, which leads to serial correlation in the residuals. My argumentation also explained why the demeaned net load term, which stands for capturing the systematic difference from the expected value of the net load, does not include the missed information. In my point of view, the solution for the problem was the incorporation of the past errors into the model in order to capture the effect of the latent variables and get rid of the serial correlation. I completed the model with the past errors of the last six hours and with the past errors occurred 23-25 hours ago. In relation to the final specification, the model passed the Durbin-Watson and the LM-test but it had to be estimated with the White heteroscedasticity-consistent standard errors in order to get a theoretically proper model (see on A3.40).

Variable	Coefficient	P-value
C	703.0654	0.0000
MEAN_NET_LOAD	0.750043	0.0000
DEMEAN_NET_LOAD(-1)	0.686545	0.0000
TEMP_V*D1	7.344490	0.0000
TEMP_V*D2	2.648539	0.0008
TUESDAY	-122.9244	0.0288
THURSDAY	-167.6213	0.0272
FRIDAY	-298.1386	0.0000
SATURDAY	-994.1958	0.0000
SUNDAY	-1477.538	0.0000
NET_LOAD(-1)	0.374283	0.0000
NET_LOAD(-2)	-0.207570	0.0001
NET_LOAD(-3)	0.048703	0.0070
MONDAY*NET_LOAD(-23)	0.107999	0.0000
FRIDAY*NET_LOAD(-23)	0.043292	0.0001
TUESDAY*NET_LOAD(-23)	0.041150	0.0000
THURSDAY*NET_LOAD(-23)	0.101347	0.0000
SATURDAY*NET_LOAD(-23)	-0.030992	0.0304
SUNDAY*NET_LOAD(-23)	-0.075839	0.0041
FRIDAY*NET_LOAD(-24)	0.095013	0.0000
SUNDAY*NET_LOAD(-24)	0.144127	0.0000
MONDAY*NET_LOAD(-25)	-0.107287	0.0000
TUESDAY*NET_LOAD(-25)	-0.035280	0.0000
THURSDAY*NET_LOAD(-25)	-0.092116	0.0000
FRIDAY*NET_LOAD(-25)	-0.124653	0.0000
SATURDAY*NET_LOAD(-25)	0.069163	0.0000
MA(1)	0.311662	0.0000
MA(3)	-0.134514	0.0000
MA(4)	-0.111078	0.0000
MA(5)	-0.065439	0.0000
MA(6)	0.052466	0.0000
MA(23)	-0.048861	0.0100
MA(24)	0.375970	0.0000
MA(25)	0.090434	0.0016
R-squared	0.992287	
Adjusted R-squared	0.992256	
Akaike information criterion	13.70645	
Durbin-Watson stat	1.989081	

6. Table - The second correct specification of the model with exact-day matching

After the positive results I also incorporated the past error terms into the seasonally differenced model as well to test which approach can handle the problem of serial correlation in a more efficient way. The final specification of the model was statistically correct only with the White standard error correction. It had a better Durbin-Watson statistics and it was more robust against the LM-test, which could also disproved the presence of autocorrelation (see on A3.41).

Variable	**Coefficient**	**P-value**
D24_MEAN_NET_LOAD	-0.015685	0.0057
D24_NET_LOAD(-1)	1.619028	0.0000
D24_NET_LOAD(-2)	-0.902104	0.0000
D24_NET_LOAD(-3)	0.213492	0.0000
D24_NET_LOAD(-5)	0.048101	0.0000
D24_TEMP_V*D2	2.734653	0.0027
D24_TEMP_V(-1)*D1	3.326240	0.0002
MA(24)	-0.761517	0.0000
MA(23)	-0.077169	0.0000
MA(25)	-0.047631	0.0000
MA(2)	-0.037431	0.0000
MA(4)	0.034521	0.0000
MA(5)	-0.021465	0.0021
MA(6)	0.027168	0.0000
R-squared	0.975396	
Adjusted R-squared	0.975356	
Akaike information criterion	13.65141	
Durbin-Watson stat	1.997509	

7. Table - The third correct specification of the model with exact-day matching

Then I tested the forecasting performance of the two improved versions. As it can be noticed on the charts A3.42 the forecasting efficiency of the model using seasonally differenced data is very weak. This model specification is even outperformed by the naïve benchmark model. The critical days are again Thursdays and Fridays, furthermore the worst predicted months are March, April and October. However, as the charts A3.43 show the forecasting power of the model specification that applies daily dummies instead of differencing is quite strong. The average MAPE of the one-step-ahead forecasting procedure was 0.85%, which is even better than the MAPE of the exponential smoothing with double seasonal cycles technique and this model is statistically correct. In the case of the original model the average day-ahead MAPE was 2.1%. Therefore it can be stated that the latest model specification is almost as robust as the original one.

To sum it up, one of the key experiences of the replication of Ergun and Jun (2011) is that the exact-day matching technique cannot handle the daily and the weekly seasonality. Besides, with respect to the Hungarian dataset it is also relevant to take into consideration the nonlinear relationship of temperature and net load. Furthermore, it seemed that beside temperature and past values of net load, there are still further relevant omitted information which could be included into the model via past errors. However the most essential finding was that the incorporation of the day dummies into the model can handle the intraday periodicity better than seasonal differencing. Finally I tried out another multivariate model specification of Ramanathan et al (1997), which follows the local modelling approach and estimate separate models for each hours.

3.5 EGRV model

In line with the paper of Ramanathan et al (1997), I separated the Hungarian database according to weekdays and weekends. Then I created the necessary daily and monthly dummy variables as well as the special temperature variables. The authors used the square of the defined temperature variables in order to capture the nonlinear effect counter to Ergün and Jun (2011) who applied a V-shape formula. After the necessary data manipulation, I ran the original model specification on the Hungarian dataset. As the estimate was noisy and at some cases the statistical tests failed, I customized each regression separately. First of all, I got rid of all of the insignificant explanatory variables with the help of the Wald-test, then conducted the LM-test and Heteroscedasticity test, finally all of the 48 models were statistically correct (see on A3.44). Generally, there was less problems with serial correlation connected with this local approach and it could be easily eliminated by the drop of the insignificant variables from the equations.

Besides, dummy variables for months and days worked well in the case of these models. Furthermore, their interaction with the weather variables were most of the time statistically significant and influential. Usually the dummy variables of winter and summer months were the most impactful in all regressions, but their interactions with temperature were only statistically significant. However, different days remained significant in the different equations and the most important one was Monday. The "Dayafterholiday" variable was a relevant explanatory variable in each equation. Its coefficient was negative until 4 am then it turned into positive. This fact underpins the spillover effect that the load profile behaves in the dawn of a day, after holiday still like on a holiday, and it begins to catching up its normal pattern only from morning. Beside the calendar variables the temperature of the given hour and former day maximum temperature were important variables as well as the average midnight temperature of the past seven days. Squared variants of the different temperature variables remained significant but their explanatory power were weak. It is important to mention that MA terms in the case of these model specifications were still essential. The difference between the weekday and weekend models were that fewer explanatory variables remained significant, which were the months and the interacted terms with temperature. The key findings of the model specifications were in line with the original one, which means that in both model the months, Monday, temperature and maximum temperature of the former day were the most relevant explanatory variables (see on A3.45).

Finally I tested the forecasting performance of the models. It is important to take into consideration that the forecasting horizon of this specification is one-day-ahead instead of one-step-ahead. As it can be observed on the charts A3.44 the worst prediction occurred in August and May, but otherwise the average daily MAPE on the given days and in the given months were around 2.5-3%. The total average MAPE of the day-ahead forecast was 3.07%, which is even better than in the case of the original model where the average daily MAPE was 4.79%.

After the detailed analyses of the different model specifications, it can be concluded that the univariate models had the worst forecasting performance, with the exception of exponential smoothing techniques, however this model remained serially correlated. The model specification with exact-day matching technique and dummy variables proved to be the best in terms of forecasting accuracy. However, the forecast of this model was only one-step-ahead, therefore it is questionable whether in the case of a 24-step-ahead forecast it could still outperform the EGRV model. The main advantage of the EGRV-type model is that it was homoscedastic as opposed to the other ones. On the other hand, the RESET-test signaled misspecification almost at each equation, however in the case of the other models the null hypothesis of the RESET-test could not be rejected. Hence, there is no ultimate model specification; each of them has its advantages and disadvantages which have to be considered during the modelling and have to be handle with the appropriate econometrics techniques. However it has to be emphasized that according to the interpretation of the Chow forecasting test, there is a change in the set of coefficients, which indicates that the model is not robust and questions the long-lasting predicting power of the model or the linearity of the underlying relationships (see on A.46).

All in all, the main objective of the replication of Taylor et al (2006), Ergün and Jun (2011) and Ramanathan et al (1997) was to find the fittest modelling approach and modelling techniques for the forecasting model of the Hungarian net load. In line with the mentioned studies, I was experimenting with univariate and multivariate models, moreover with local and global modelling approaches, furthermore I tested whether seasonal differencing or dummy variables handle the seasonality pattern of the net load better. According to the result of the replications, it can be concluded that with respect to the Hungarian data both the weekly and the daily periods are strong and dummy variables can handle them. Therefore, in the framework of my own model I will apply dummies for the days of the week and for the months. Besides, it turned out that univariate models are not suitable for precise forecasting because they do not take into account lot of influential factors. Hence in spite of the incorporation of lags and MA terms they usually suffer from serial correlation. As a consequence, my model specification will be multivariate because based on the replication of Ergün and Jun (2011) and Ramanathan et al (1997) the weather and the calendar variables have a huge impact on the net load, as a result of which these variables had significant explanatory power in the replicated models as well. However it is not straightforward which solution can capture the nonlinear impact of temperature on net load better, the V-shape or the incorporation of the squared values. Furthermore, it is also questionable whether the other weather variables such as humidity, wind speed and illumination have a significant effect on net load beside the temperature or not. As far as I see, these questions open the field for development, therefore according to the result of the graphic analyses it is worth to experiment with the incorporation of the further weather variables and test the fit of the two temperature specifications. On the other hand, it is evident that the interaction of the months and temperature is essential, because it can also capture the seasonal cycles and the changing influence of temperature, therefore I am planning to apply this approach with regard to the other weather variables. The other crucial experience of the replication is that on holiday the load profile changes and some spillover effect also can be identified. However I doubt that the "Dayafterholiday" variables defined by Ramanathan et al (1997) can capture all of the alterations in electricity demand which are due to a holiday. Hence I also regard this question as a potential improvement in my own model. Finally, I considered the modelling approach and chose the local one, because separate models for the hours can handle the intraday change in a more parsimonious way. In the case of a monolithic model, extra interacted variables should be introduced to be able to pursue the within day electricity demand change. Besides, regarding a non-monolithic model, there are less fear of serially correlated error terms and accumulation of forecast errors, because instead of a 24-step-ahead forecast it is enough to execute one step to get the day-ahead load prediction. However, the drawback of this local modelling approach is that it cannot take into account the within day impacts, because it prepares the forecast from the yesterday value. Therefore as a solution, it is important to use aggregated variables such daily mean, minimum or maximum of the weather variables like Ramanathan et al (1997). As a result, in the next chapter I will experiment with an own model based on the key findings of the graphic analyses and the replications, trying to answer the open questions and tying to exploit the field of potential improvement. Hence, I will use the EGRV model as a starting point and test the incorporation of the further weather and holiday variables in order to figure out the most precise forecasting model of the Hungarian net load for 2015.

4. My own model

The main objective of this chapter is to figure out the most fitted forecasting model of the Hungarian system net load for 2015, based on the findings of the graphic analyses and the replications. Before the deep analyses I committed myself to some standpoints in advance. First of all, it was a straightforward conclusion of the replication that the dummy variables can handle the seasonal pattern of the load profile better than seasonal differencing. Therefore I incorporated the days of the week and months as dummy variables into my own model. Besides, I decided to employ the local modelling approach and ran separate regressions for each hour of the day. One of the advantages of this solution might be an easier handling of the intraday transition in the load profile, since the value of the coefficients of the given dummy variables can differ by equations in this setup. On the other hand, in the case of a monolithic model interacted day and hour variables should be created for this reason, which would mean less parsimonious and precise estimation procedure. However, it remained an open question of the replications whether the separation of the weekdays and weekends increases the accuracy of the forecast or not. Furthermore, it was also questionable whether the V-shape or the U-shape definition of the temperature can capture the nonlinear nature of the temperature in a more precise way. As a result, during the development of my own model I had to consider these problems as well. As opposed to the introduced studies, I deemed the incorporation of the humidity and the illumination in my own model necessary based on the graphic analyses. I also considered the expansion of the set of the holiday variables important. The studied papers used only one holiday variable, but based on the result of the replication it could not capture the complex and long-lasting impact of a holiday on the net load profile. Therefore I introduced a “national holiday” variable, a simple “holiday” variable, which contained the extra days of the long weekends and a “working Saturday” variable. Moreover, I defined “after holiday” and “before holiday” variables in order to pick up the spillover effects of a holiday, moreover a

"Sunday after working Saturday" variable for the spillover effects of working Saturdays. Besides, I also created a "summer holiday" variable based on the expertise advice of MAVIR. This dummy variable was 1 from the last week of July till August, which is the most favored period of the (family) vacations and in-parallel it is a slowing down period in the business life. After the clarification of the main aspects of my own modelling procedure, I continue the description of the technical parts.

Based on the logic of the EGRV model, I expanded the original setup with the forecasted value of illumination and humidity of the given day and their interaction with the months. I added also the mean, minimum and maximum humidity values of the given and previous days in order to incorporate the cross-hourly impact of the variables as well. I employed the minimum and mean values of the new weather variables beside their maximum because there was no rule-of-thumb in the literature about which measures are the most relevant regarding to these variables such in the case of temperature. Furthermore, I also included the different holiday variables into the regressions. However I used the past seven lags of the net load instead of the net load at 8 am, because the latter was not an essential explanatory variable in the replicated EGRV model.

The first specification of my own model already outperformed the replicated EGRV model, therefore I did not hold it important to separate the weekends and the weekdays. As a next step, I tested which temperature specification is the better. I chose the solution of Ramanathat et al (1997), because that model specification had better forecasting performance which used their temperature variables. Besides, the solution of Ergün and Jun (2011) is less parsimonious in the matter of estimation, because it requires the introduction of several new interacted terms. Finally, I customized each equation of my own model and dropped the insignificant variables (see on A4.1).

Generally, the R-squared was higher, the AIC was lower in the case of the 24 models compare to the replicated EGRV model, which indicated that the new specification had stronger explanatory power regard to the evolution of the load profile. This improvement reflected in the forecasting performance of my own model as well. As it can be observed on the charts A4.2 the average MAPE of the day-ahead-forecast was 2.89%, which means 0.18% point decrease compare to the replication of the EGRV model. The biggest forecasting errors occurred on Sundays and in August, however the MAPE was below 2% in November, June and February.

Although, the most important feature of a forecasting model is the accuracy but it is worth to check whether the models are in line with the intuition and the findings of the graphic analyses. The most relevant variables are the months of summer and winter, the days of the week, temperature and the newly introduced different holiday variables. Saturday, Sunday had significant negative coefficient, while Thursday and Tuesday were not relevant from the economic point of view. Monday and Friday were significant explanatory variables during the transitory period, which means that Monday had negative coefficient between 1am and 4 am, while Friday had negative coefficient from 5 pm. These facts are in line with the discoveries of the graphic analyses, therefore it can be proved that the dawn of Monday still acts like Sunday, while the load profile pursues its weekend pattern after the end of the working hours on Friday. A similar phenomenon characterized the "after holiday" and "before holiday" variables. The former was significant at the beginning of the day, while the latter from the afternoon. "Working Saturday" variable had positive coefficient, which means that the electricity consumption is higher on those Saturdays which are working Saturdays versus it would not be a working Saturday. It was relevant from 5 am, which is the start of the first working shift and the daytime public transport. However, the impact of working Saturdays on Sundays can be neglected based on the result of the regressions. Among the weather variables the temperature and the maximum temperature of the given day remained only significant from economic perspective.

As well as in the replication of the EGRV model, the lags of the net load, interaction of the months and the weather variables, moreover the MA terms contributed only to the statistical correctness of the model specification.

All of the explanatory variables were significant at 5% significance level, furthermore none of the models was serially correlated based on the Durbin-Watson statistics and the LM-test. The heteroscedasticity could not be eliminated, thus I applied the White heteroscedasticity-consistent standard errors. Furthermore, the RESET-test signaled again the misspecification of the regressions, similar to the replication of the EGRV model. However, in the case of my own model the underlying problem of the RESET-test also could have been the heteroscedasticity. Therefore, I carried out the PE-test in order to figure out whether the linear specification is correct or not. As it can be noted in the table A4.3 the results of the PE-test were ambiguous. It could be stated that the logarithmic terms had mostly significant contribution to the linear equations, however the linear terms in the logarithmic specification were not significantly zero. Consequently, I turned to a graphic solution to be able to draw conclusion. As the graphs A4.4 show the evolution of the net load profile of the given hours were far not linear, it rather pursued a U-shape or other complex polynomial form.

Finally I conducted the logarithmic specification of the first version of my own model and then I dropped the insignificant explanatory variables. As it can be detected in the table A4.5 the R-squared and especially the AIC improved in a substantial way. Based on the AIC it can be assumed that the logarithmic version is a better specification than the linear. The forecasting performance of the new setup also proved the superiority of the logarithmic form. As the graphs A4.6 represent, the average MAPE of the day-ahead-forecast was 2.39%, which means 0.51% point improvement compare to the linear version and 0.86% point compare to the replication of the EGRV model. Sunday remained still the critical day and August the critical months with respect to accuracy. It is important to mention that for a correct comparison of the MAPE indices, the forecasted logarithmic values first had to be retransformed to linear ones. The transformation of the forecasted logarithmic values were executed according to the formula of Wooldridge (2013, p211-212). Taking the exponential of the forecasted logarithmic values would be a naïve approach (Wooldridge, 2013) and would consistently overestimate the true values based on Jensen's inequality. In terms of the set of explanatory variable there was no remarkable change between the linear and logarithmic version. The only exceptions are Monday and the lag of net load which were not significant in the logarithmic models.

In summarizing, I experimented with the development of the fittest forecasting model of the Hungarian system net load for 2015. Based on the results of the replications and the graphic analyses I considered the EGRV model as my starting point with the inclusion of humidity, illumination and different holiday variables. The extended version had already outperformed the replicated EGRV model according to the R-square, AIC and MAPE statistics. However the RESET-test indicated that the hourly models are misspecified, and this result was also underpinned by the PE-test. The re-specified models in logarithmic forms proved to be the most accurate in terms of forecasting performance. Hence, the main findings of my own modeling were the discovery of the nonlinear nature of the Hungarian load profile and its handling with a logarithmic form. Besides, it is also useful to mention that illumination and humidity variables were only statistically significant. However the extension of the original EGRV model with further specific holiday variables seemed to increase the accuracy of the forecast. As a result, it is a field of improvement to scrutinize the exact functional form of the Hungarian net profile and the weather variables.

Conclusion

The main objective of my thesis was to develop a handy and accurate load forecasting model for the Hungarian Transmission System Operator. My research question is relevant, because the currently applied load forecasting model of the Hungarian TSO is inaccurate, which risks security of the system and increases the operational cost. Nevertheless, there was no Hungarian study written about this problem. According to the literature the most suitable models for my research topic are the conventional statistical methods and regression-based econometric techniques. Therefore, I tested the forecasting performance of some classical methods: the univariate double seasonal ARIMA and Holt-Winters exponential smoothing with double seasonal cycle models of Taylor et al (2006), the monolithic multivariate model with exact-day matching techniques of Ergün and Jun(2011), as well as the multivariate non-monolithic model specification of Ramanathan et al (1997). Based on the results of the replication of the mentioned model specifications and the graphic analyses I developed an own model, taking into account all of the specificity of the Hungarian net load.

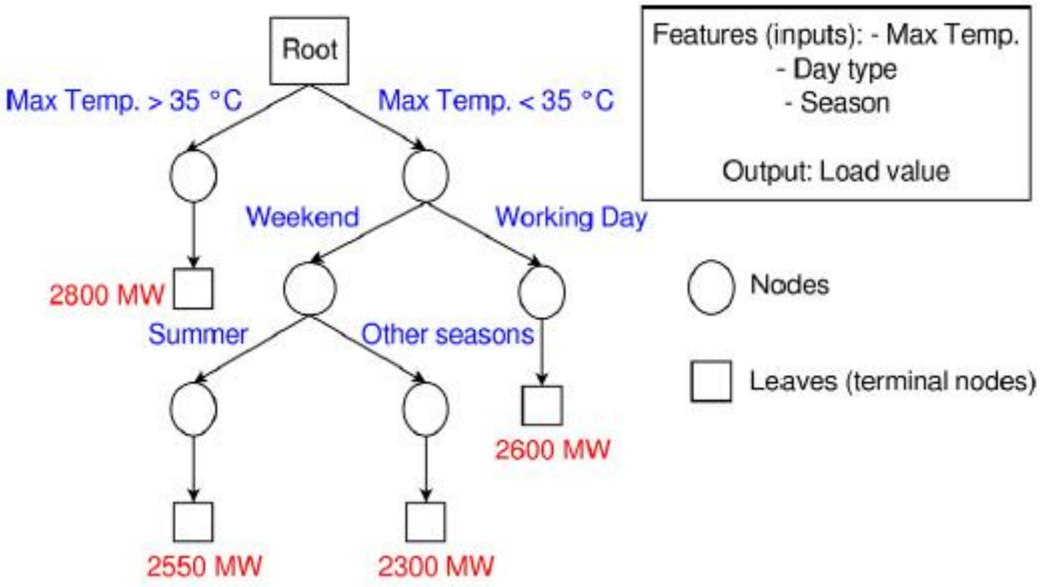
One of the key findings of my modelling procedure was that the Hungarian net load is characterized by several cycles and it means a challenge to handle the intraday, intraweek and yearly periodicity or to capture the special load profile pattern on weekend and on holidays. In the case of the Hungarian dataset the dummy variables were more efficient in the handling of the underlying seasonal cycles than the seasonal differencing processes. Furthermore the holiday, national holiday, before and after holiday variables, as well as the working Saturday and Sunday after working Saturday special dummy variables could effectively explain the deviant behavior of the load profile. Furthermore, it turned out that the log-linear model specification is more precise than the linear models because it can solved the estimation of the non-linear evolution of the system load with a linear regression. However, none of the covered literature applied this logarithmic transformation. The final version of my own model was log-linear transformation of the expanded EGRV model. This final specification had the best forecasting accuracy, in the case of which the average MAPE of a day-ahead forecast was 2.4%.

However there is still field for the improvement of forecasting accuracy. First of all, the MAPE was the highest in the case of the summer months, which could be decreased with the incorporation of the impact of air conditioning. For this extension it would be necessary to know the reaction of the market participants to the heath and their cooling behavior and heuristics. Moreover, the explanatory power of humidity and illumination could be strengthened by the clarification of their exact functional form such in the case of temperature. Furthermore, the determination of the accurate polynomial form of the system load would be also a great progress. Besides, as far as I see, the biggest risk of the Transmission System Operator is the deviation from the balanced schedule during the day, therefore it would be advisable to forecast only the difference instead of the total net load of the hour. The advantage of this new approach would be the exclusion of latent variables and the decrease of the noise of the estimate. For the implementation of this new approach the dataset of the balanced net load would be required.

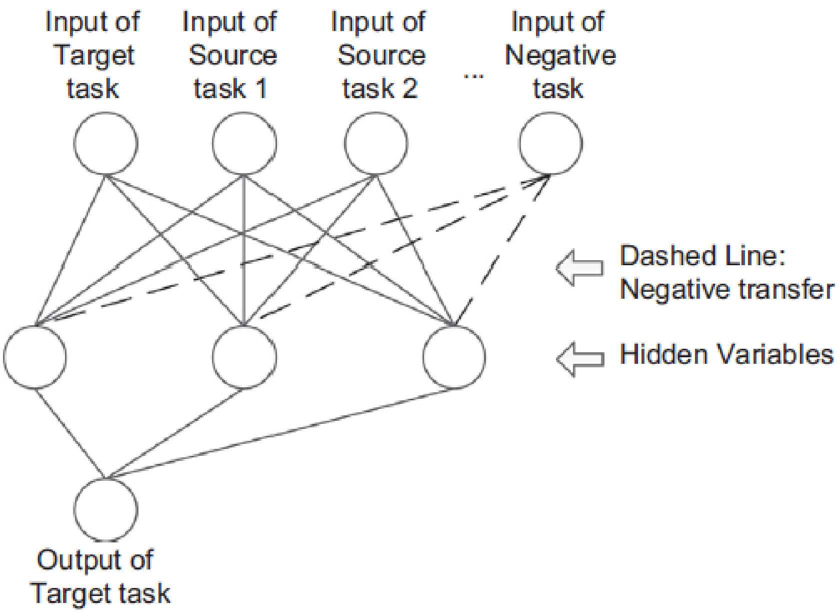
Appendices

A1. Appendix for literature review

1. A1 – Sample Regression Tree for predicting load at midday in 2013, Tunisia (Lahouar and Slama, 2015, p1042.)

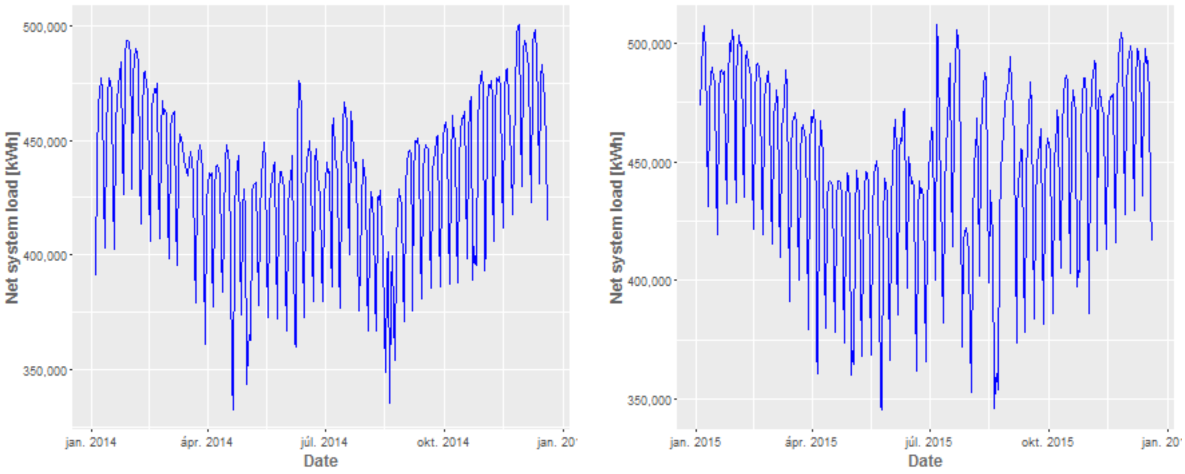


2. A1 – Target task, source task and negative transfer in transfer learning problems (Zhang and Luo, 2015, p163.)

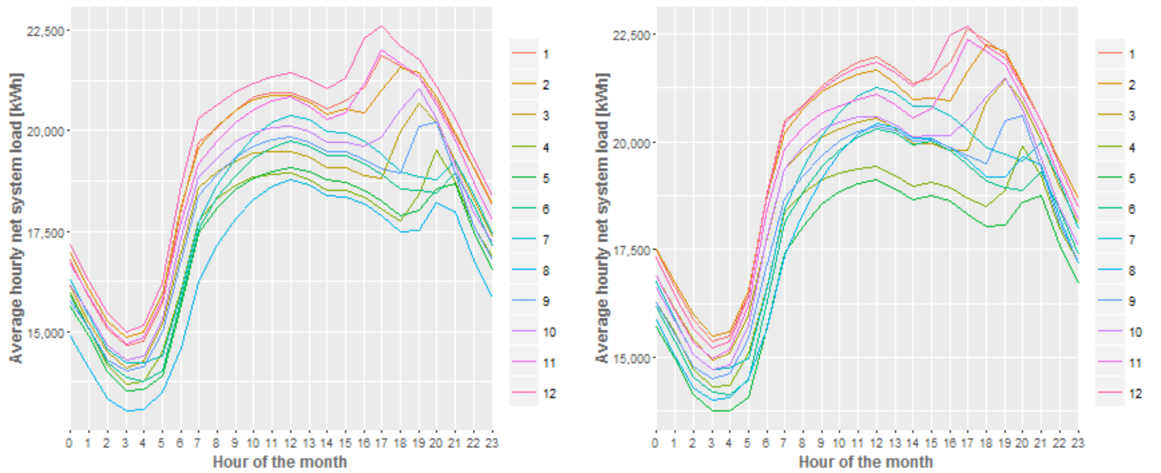


A2. Appendix for Data and Methodology[1]

1. A2– The evolution of net system load in 2014 and 2015

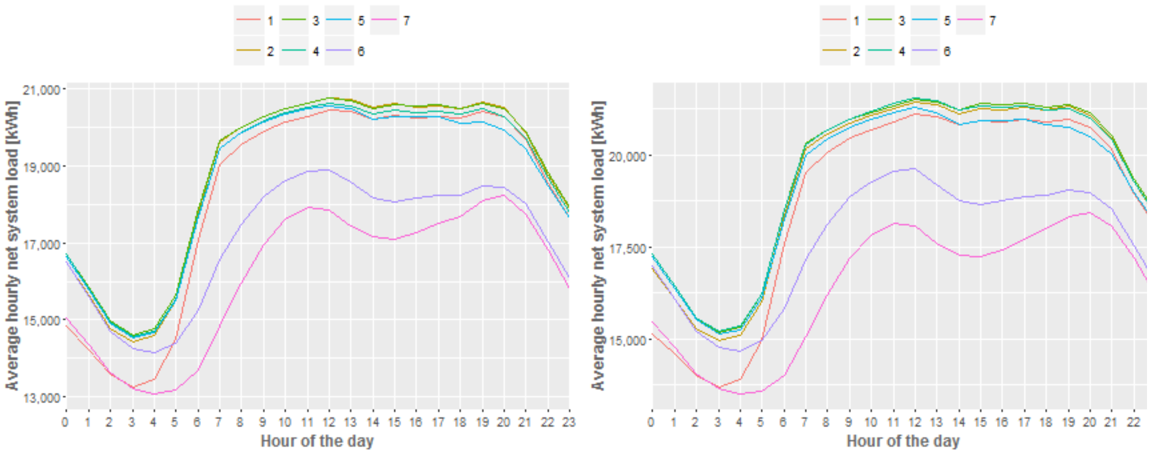


2. A2- Average hourly net system load by months in 2014 and 2015

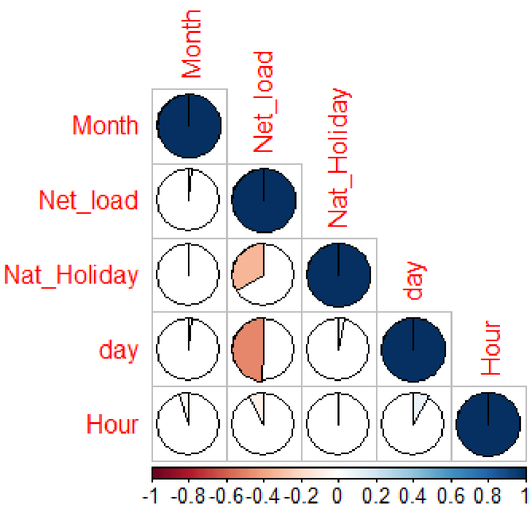


[1] All of the charts are own edition, using the dataset of MAVIR

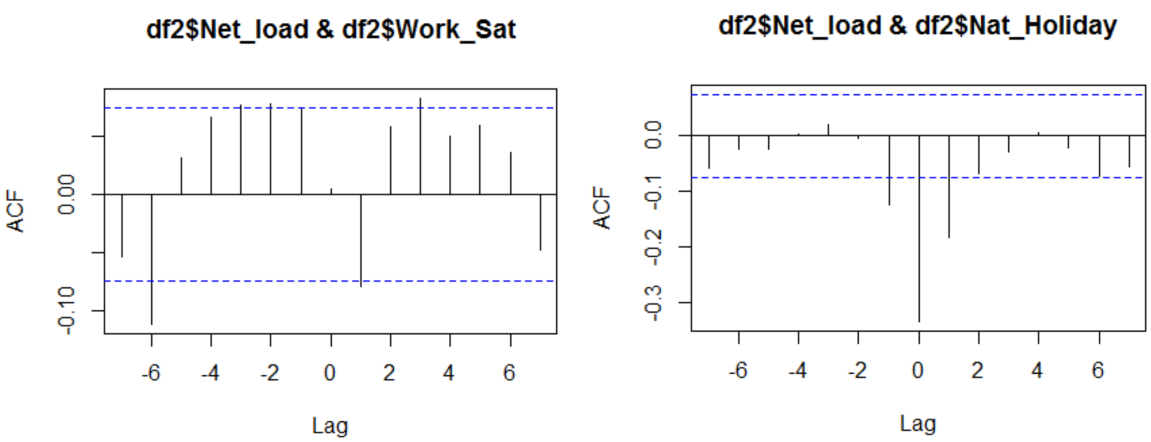
3. A2- Average hourly net system load by days in 2014 and 2015



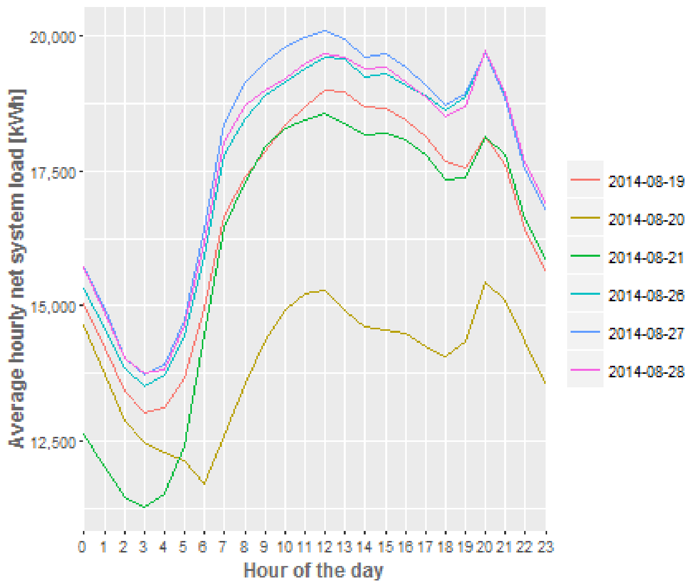
4. A2- Correlation matrix of the net load and the calendar variables



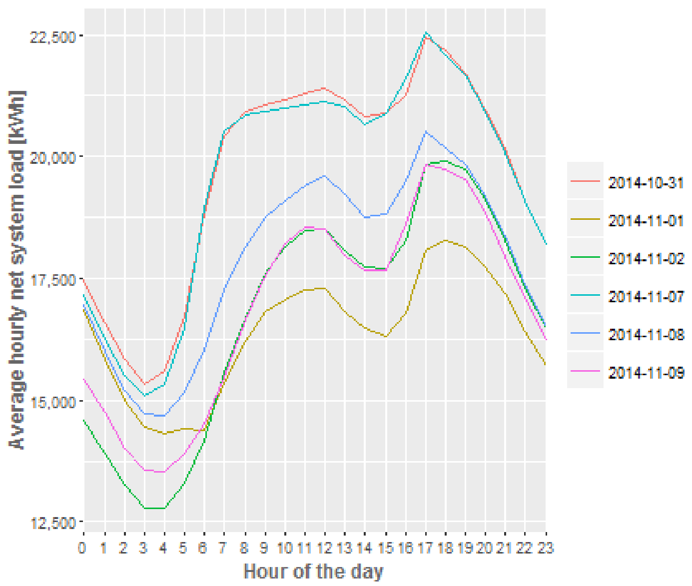
5. A2– Cross-collelograms of daily net load with “working Saturdays” and “national holidays



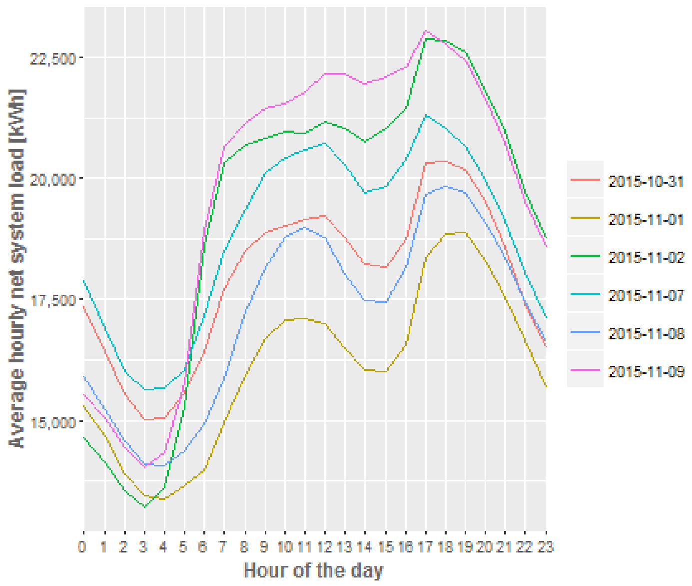
6. *A2*– **Single holiday: 2014.08.20 – Wednesday**



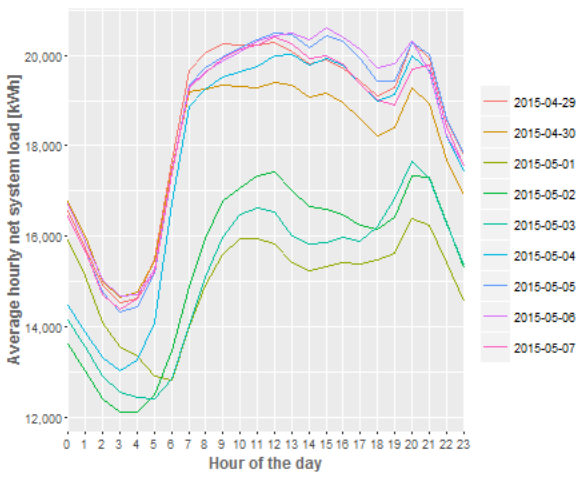
7. *A2*– **Single holiday: 2014.11.01 – Saturday**

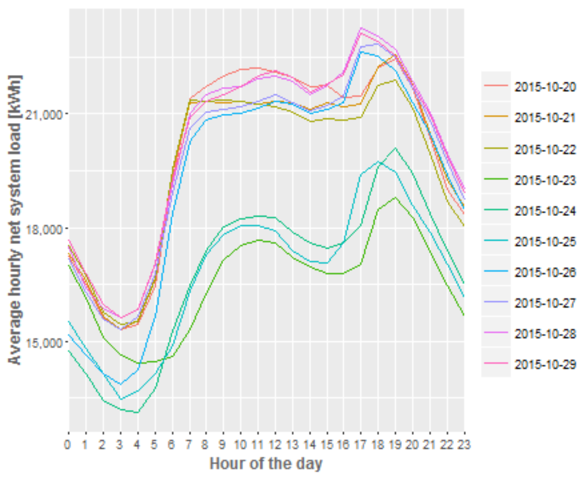


8. *A2* **– Single holiday: 2015.11.01 – Sunday**

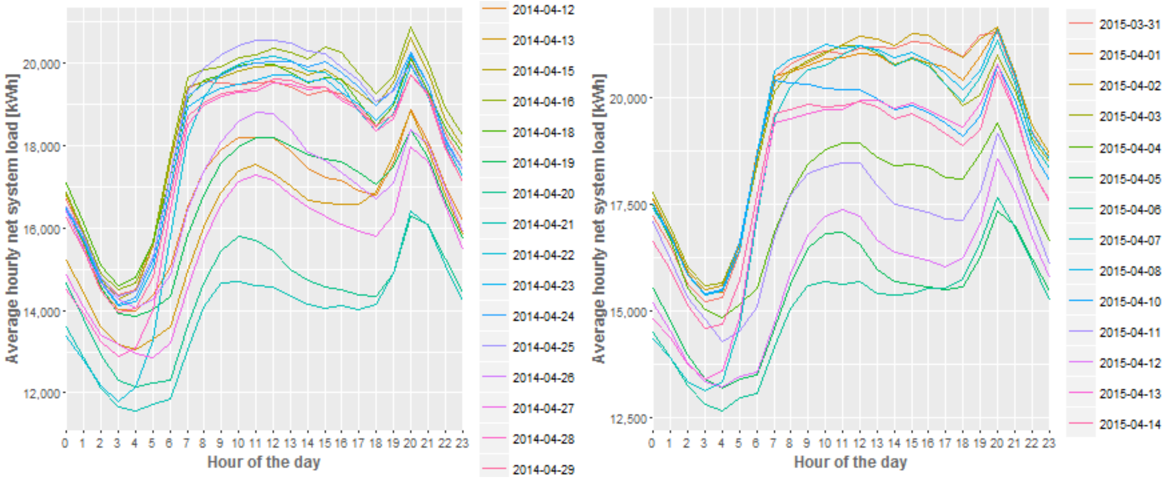


9. *A2***– Long weekends, 3 days with Friday: 2015.05.01 and 2015.10.23**

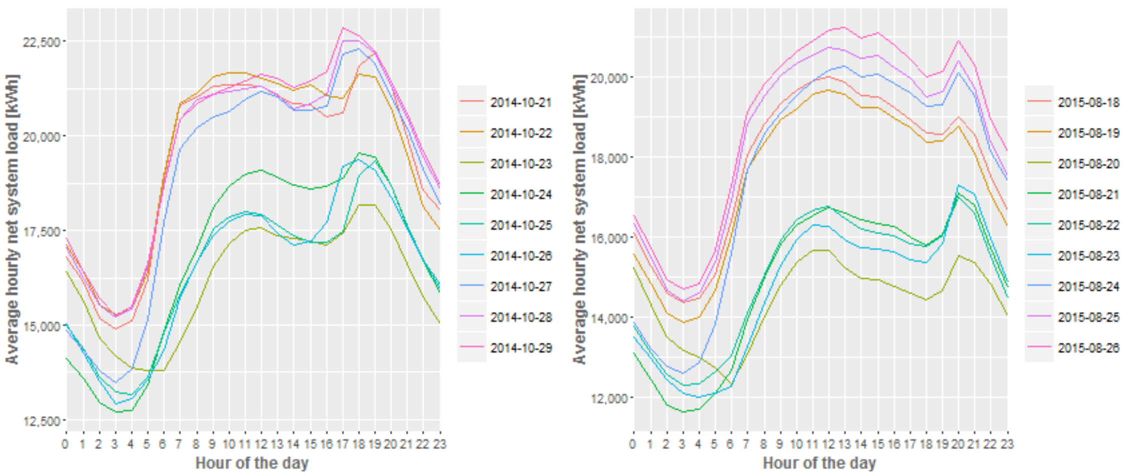




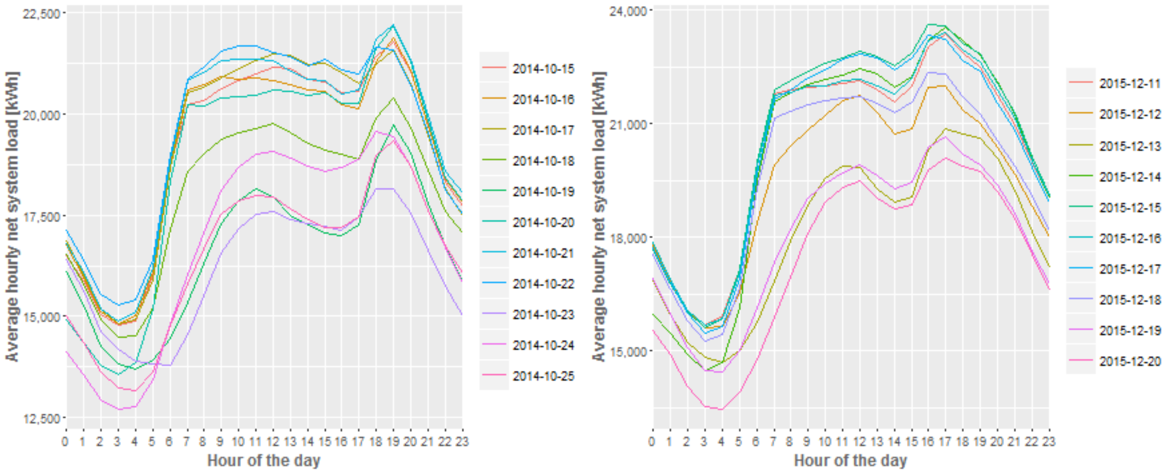
10. *A2* **– Long weekends, 3 days with Monday: 2014.04.20 and 2015.04.05**



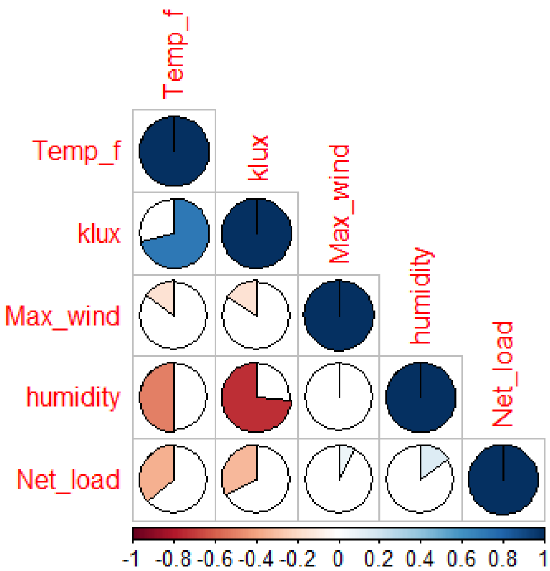
11. *A2* **– Long weekends, 4 days with Thursday and Friday: 2014.10.23 and 2015.08.20**



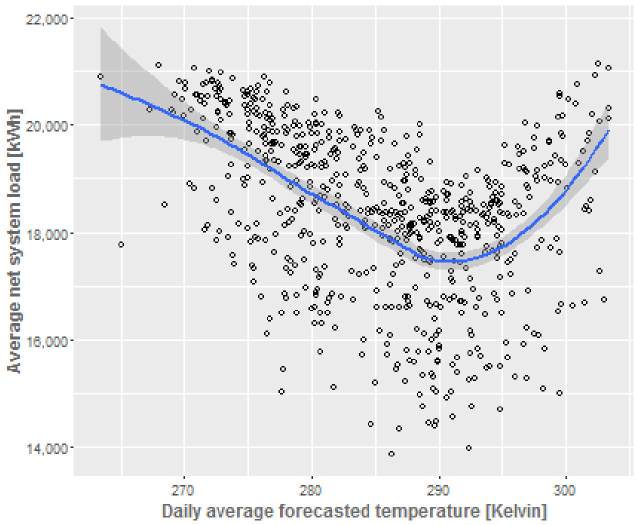
12. *A2*– **Working Saturdays: 2014.10.18 and 2015.10.18**



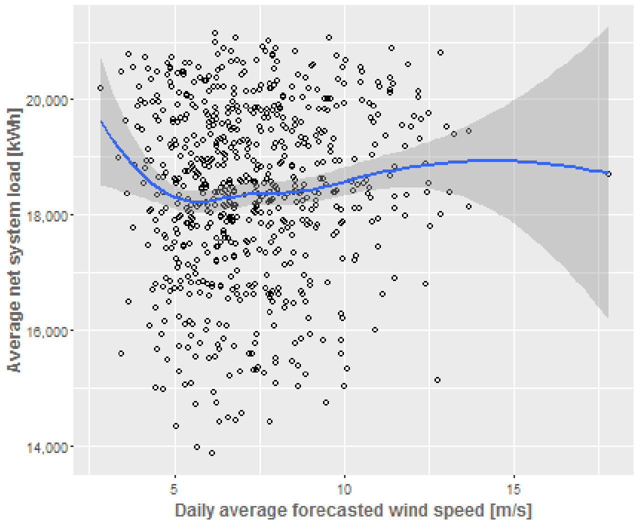
13. *A2*- **Correlation matrix of net load and the weather variables**



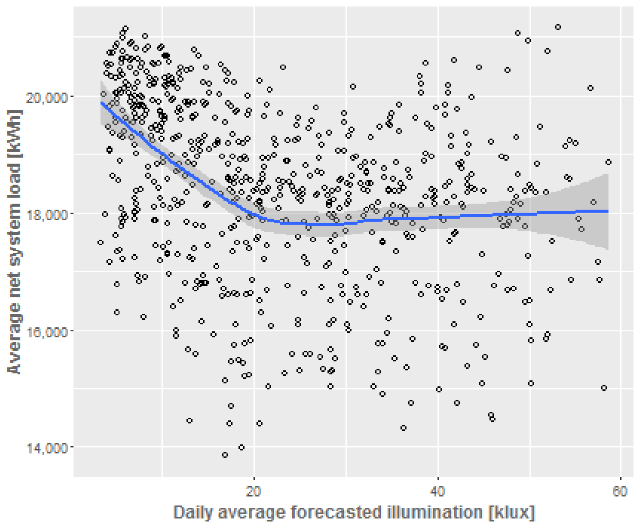
14. *A2*– **Average daily net load and daily average forecasted temperature**



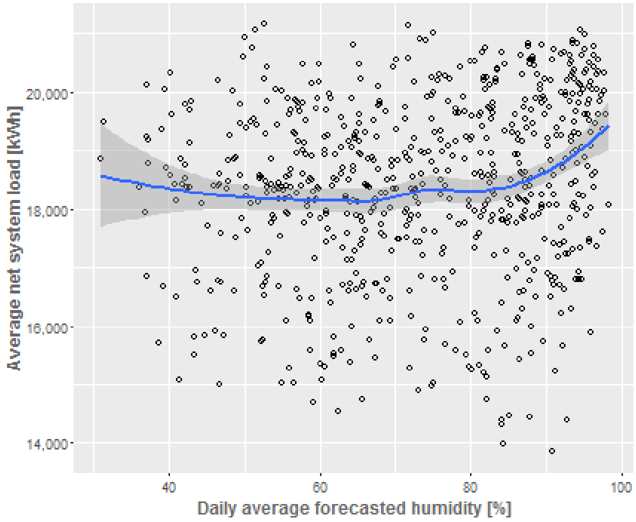
15. A2 - Average daily net load and daily average forecasted wind speed



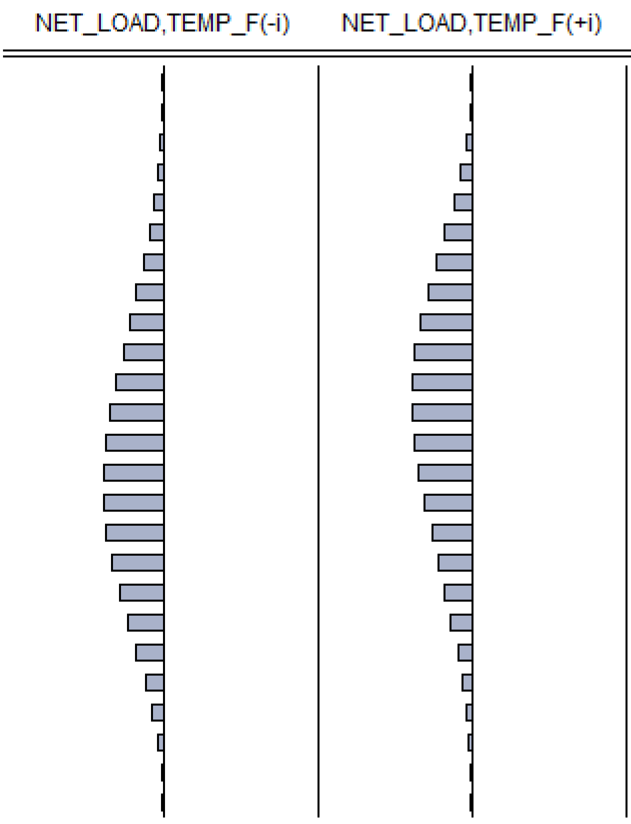
16. *A2* **- Average daily net load and daily average forecasted illumination**



17. *A2*- **Average daily net load and daily average forecasted humidity**

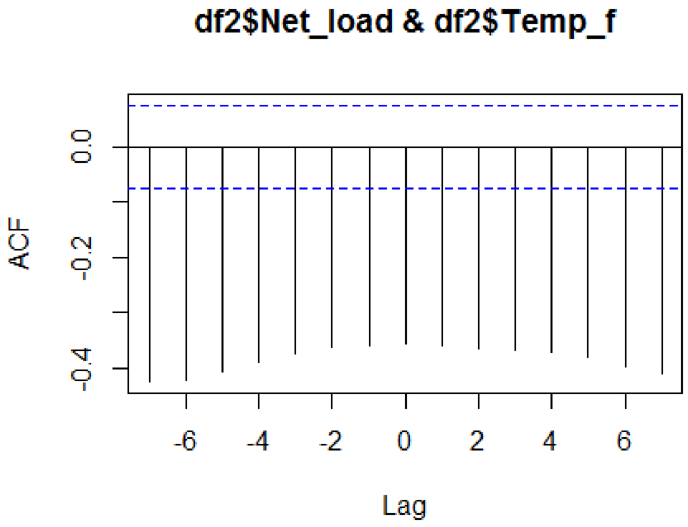


18. *A2*– **Cross-correlogram of hourly net load with hourly temperature**

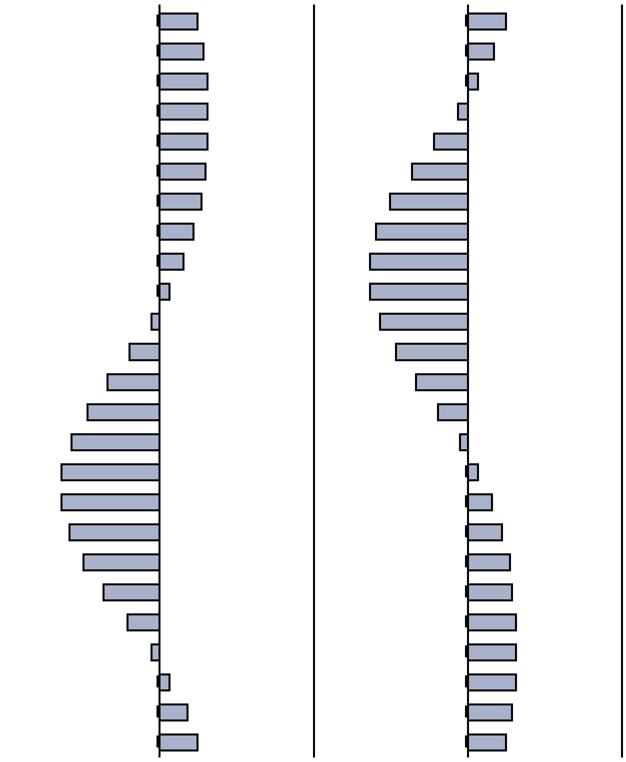


i	lag	lead
0	-0.0027	-0.0027
1	-0.0055	-0.0116
2	-0.0172	-0.0343
3	-0.0358	-0.0710
4	-0.0607	-0.1197
5	-0.0921	-0.1766
6	-0.1298	-0.2368
7	-0.1733	-0.2940
8	-0.2209	-0.3429
9	-0.2695	-0.3789
10	-0.3160	-0.3989
11	-0.3564	-0.4011
12	-0.3861	-0.3861
13	-0.4012	-0.3566
14	-0.3991	-0.3165
15	-0.3792	-0.2703
16	-0.3432	-0.2220
17	-0.2942	-0.1747
18	-0.2368	-0.1315
19	-0.1764	-0.0941
20	-0.1191	-0.0628
21	-0.0702	-0.0378
22	-0.0334	-0.0190
23	-0.0107	-0.0070
24	-0.0020	-0.0040

19. *A2*- **Cross-correlogram of daily net load with daily temperature**

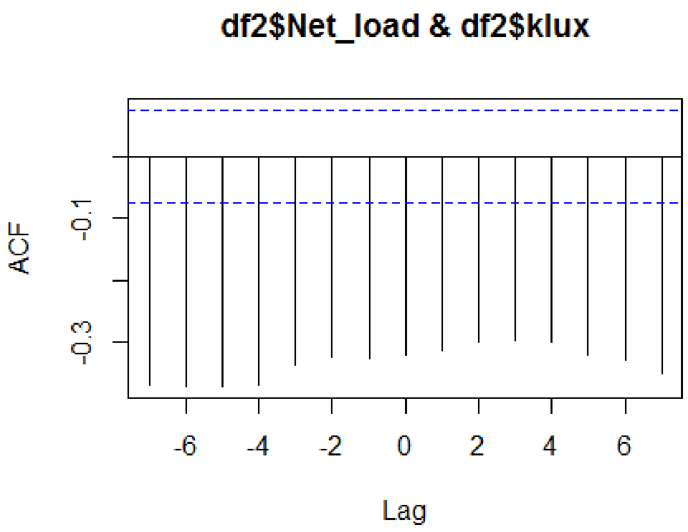


20. *A2*- **Cross-correlogram of hourly net load with hourly illumination**



NET_LOAD,KLUX(-i)	NET_LOAD,KLUX(+i)	i	lag	lead
		0	0.2601	0.2601
		1	0.3021	0.1869
		2	0.3216	0.0791
		3	0.3276	-0.0586
		4	0.3249	-0.2146
		5	0.3128	-0.3712
		6	0.2866	-0.5085
		7	0.2405	-0.6087
		8	0.1701	-0.6585
		9	0.0732	-0.6511
		10	-0.0484	-0.5878
		11	-0.1889	-0.4783
		12	-0.3370	-0.3398
		13	-0.4759	-0.1921
		14	-0.5856	-0.0519
		15	-0.6491	0.0694
		16	-0.6565	0.1663
		17	-0.6065	0.2369
		18	-0.5063	0.2833
		19	-0.3690	0.3103
		20	-0.2123	0.3233
		21	-0.0561	0.3268
		22	0.0818	0.3214
		23	0.1899	0.3022
		24	0.2635	0.2603

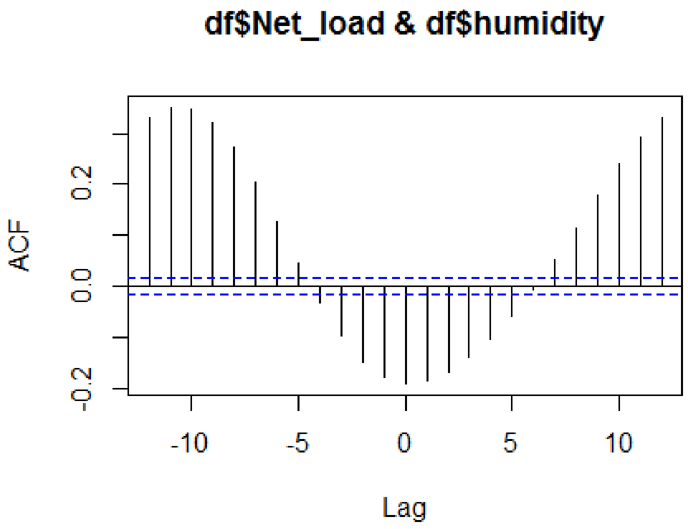
21. *A2*- **Cross-correlogram of daily net load with daily illumination**



22. *A2*- **Cross-correlogram of hourly net load with hourly humidity**

NET_LOAD,HUMIDITY(-i)	NET_LOAD,HUMIDITY(+i)	i	lag	lead
		0	-0.1903	-0.1903
		1	-0.1852	-0.1784
		2	-0.1673	-0.1475
		3	-0.1395	-0.0972
		4	-0.1035	-0.0306
		5	-0.0592	0.0471
		6	-0.0069	0.1286
		7	0.0520	0.2061
		8	0.1153	0.2720
		9	0.1795	0.3205
		10	0.2405	0.3476
		11	0.2932	0.3510
		12	0.3321	0.3317
		13	0.3520	0.2931
		14	0.3490	0.2406
		15	0.3222	0.1797
		16	0.2736	0.1158
		17	0.2072	0.0528
		18	0.1290	-0.0059
		19	0.0468	-0.0581
		20	-0.0316	-0.1023
		21	-0.0985	-0.1385
		22	-0.1486	-0.1664
		23	-0.1792	-0.1847
		24	-0.1905	-0.1902

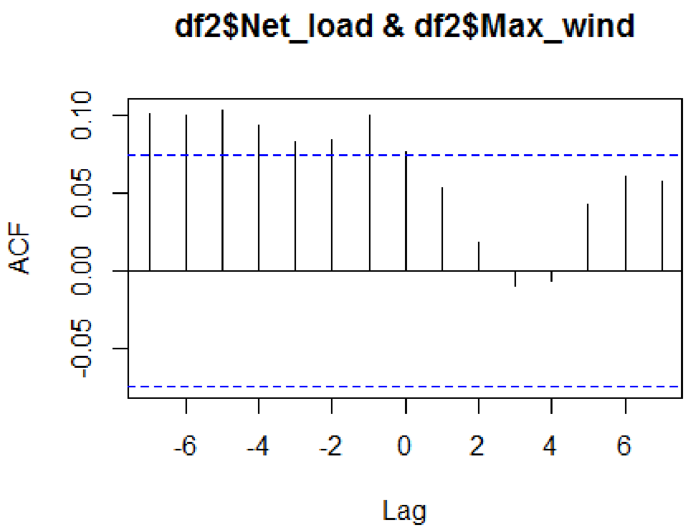
23. *A2*- **Cross-correlogram of daily net load with daily humidity**



24. *A2*- **Cross-correlogram of hourly net load with hourly wind speed**

NET_LOAD,MAX_WIND(-i)	NET_LOAD,MAX_WIND(+i)	i	lag	lead
		0	0.0006	0.0006
		1	-0.0227	0.0255
		2	-0.0429	0.0511
		3	-0.0582	0.0775
		4	-0.0671	0.1022
		5	-0.0693	0.1223
		6	-0.0646	0.1355
		7	-0.0532	0.1405
		8	-0.0356	0.1379
		9	-0.0132	0.1302
		10	0.0131	0.1172
		11	0.0408	0.0990
		12	0.0672	0.0765
		13	0.0901	0.0502
		14	0.1088	0.0226
		15	0.1227	-0.0034
		16	0.1311	-0.0251
		17	0.1338	-0.0421
		18	0.1282	-0.0534
		19	0.1139	-0.0583
		20	0.0923	-0.0565
		21	0.0660	-0.0480
		22	0.0384	-0.0330
		23	0.0121	-0.0128
		24	-0.0133	0.0112

25. *A2*- **Cross-correlogram of daily net load with daily wind speed**



26. *A2*– **ADF test of net load**

Null Hypothesis: NET_LOAD has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 42 (Automatic - based on SIC, maxlag=42)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.51928	0.0000
Test critical values:	1% level	-3.958593	
	5% level	-3.410076	
	10% level	-3.126765	

A3-Appendix for Replication

1. *A3*– **ADF test of net load**

Null Hypothesis: NET_LOAD has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 42 (Automatic - based on SIC, maxlag=42)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.51928	0.0000
Test critical values:	1% level	-3.958593	
	5% level	-3.410076	
	10% level	-3.126765	

2. *A3*– **First specification of the naïve benchmark model**

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1786.841	229.5242	7.784977	0.0000
NET_LOAD(-168)	0.901512	0.012204	73.87201	0.0000

R-squared	0.814114	Mean dependent var	18086.59
Adjusted R-squared	0.814090	S.D. dependent var	2595.480
S.E. of regression	1119.100	Akaike info criterion	16.87869
Sum squared resid	9.98E+09	Schwarz criterion	16.88044
Log likelihood	-67242.69	Hannan-Quinn criter.	16.87929
F-statistic	34888.11	Durbin-Watson stat	0.047691
Prob(F-statistic)	0.000000		

3. *A3* – **LM test of the first specification of the naïve benchmark model**

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	161075.5	Prob. F(1,7965)	0.0000
Obs*R-squared	7592.557	Prob. Chi-Square(1)	0.0000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-34.92383	19.13850	-1.824795	0.0681
NET_LOAD(-168)	0.001919	0.001048	1.831658	0.0670
RESID(-1)	0.976331	0.002433	401.3421	0.0000

R-squared	0.952881	Mean dependent var	2.90E-12
Adjusted R-squared	0.952869	S.D. dependent var	1119.029
S.E. of regression	242.9370	Akaike info criterion	13.82386
Sum squared resid	4.70E+08	Schwarz criterion	13.82649
Log likelihood	-55071.25	Hannan-Quinn criter.	13.82476
F-statistic	80537.73	Durbin-Watson stat	1.196582
Prob(F-statistic)	0.000000		

4. *A3*- **Second specification of the naïve benchmark model**

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)
MA Backcast: 145 168

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1910.595	227.5018	8.398154	0.0000
NET_LOAD(-168)	0.894647	0.012082	74.05018	0.0000
MA(1)	0.906963	0.006835	132.6860	0.0000
MA(24)	0.077954	0.004457	17.49144	0.0000

R-squared	0.946327	Mean dependent var	18086.59
Adjusted R-squared	0.946307	S.D. dependent var	2595.480
S.E. of regression	601.4182	Akaike info criterion	15.63696
Sum squared resid	2.88E+09	Schwarz criterion	15.64047
Log likelihood	-62293.65	Hannan-Quinn criter.	15.63816
F-statistic	46805.51	Durbin-Watson stat	0.332108
Prob(F-statistic)	0.000000		

5. *A3*– **LM test of the second specification of the naïve benchmark model**

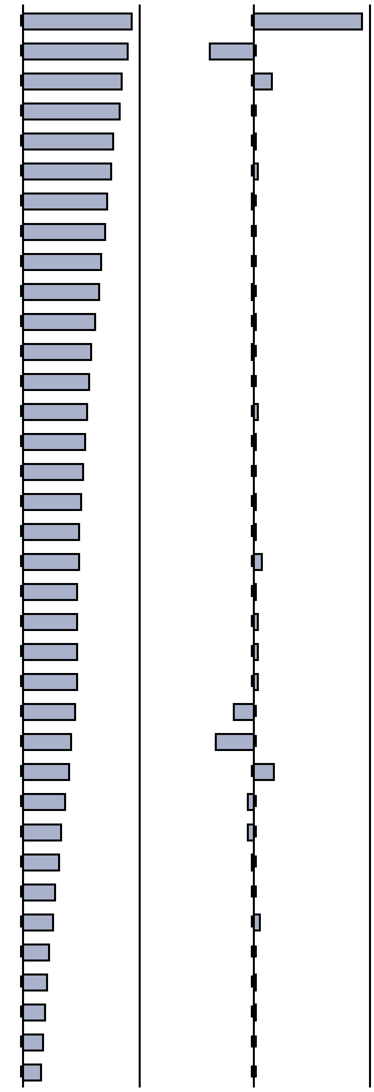
Breusch-Godfrey Serial Correlation LM Test:

F-statistic	28760.08	Prob. F(1,7963)	0.0000
Obs*R-squared	6240.226	Prob. Chi-Square(1)	0.0000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	38.95376	42.47568	0.917084	0.3591
NET_LOAD(-168)	-0.002167	0.002324	-0.932350	0.3512
MA(1)	-0.104858	0.002015	-52.04325	0.0000
MA(24)	-0.000230	0.001909	-0.120518	0.9041
RESID(-1)	0.939211	0.005538	169.5879	0.0000

R-squared	0.783161	Mean dependent var	-0.008923
Adjusted R-squared	0.783052	S.D. dependent var	601.3050
S.E. of regression	280.0739	Akaike info criterion	14.10861
Sum squared resid	6.25E+08	Schwarz criterion	14.11299
Log likelihood	-56203.71	Hannan-Quinn criter.	14.11011
F-statistic	7190.019	Durbin-Watson stat	3.024818
Prob(F-statistic)	0.000000		

6. *A3*– **ACF and PACF of the residuals after the second specification of the naïve benchmark model**



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.976	0.976	7592.7	0.000
		2	0.933	-0.408	14535.	0.000
		3	0.891	0.167	20860.	0.000
		4	0.853	0.007	26662.	0.000
		5	0.819	0.011	32012.	0.000
		6	0.788	0.027	36968.	0.000
		7	0.759	-0.008	41567.	0.000
		8	0.731	-0.001	45832.	0.000
		9	0.704	-0.004	49783.	0.000
		10	0.676	-0.021	53430.	0.000
		11	0.649	0.008	56791.	0.000
		12	0.623	-0.008	59885.	0.000
		13	0.598	0.008	62736.	0.000
		14	0.575	0.027	65375.	0.000
		15	0.555	0.020	67836.	0.000
		16	0.537	0.001	70141.	0.000
		17	0.521	0.020	72306.	0.000
		18	0.506	0.022	74352.	0.000
		19	0.495	0.075	76312.	0.000
		20	0.488	0.012	78215.	0.000
		21	0.483	0.031	80078.	0.000
		22	0.479	0.027	81914.	0.000
		23	0.477	0.031	83734.	0.000
		24	0.469	-0.184	85489.	0.000
		25	0.440	-0.350	87040.	0.000
		26	0.405	0.188	88354.	0.000
		27	0.373	-0.052	89466.	0.000
		28	0.343	-0.053	90406.	0.000
		29	0.314	-0.014	91195.	0.000
		30	0.287	0.001	91854.	0.000
		31	0.262	0.045	92405.	0.000
		32	0.241	0.001	92869.	0.000
		33	0.221	0.017	93258.	0.000
		34	0.201	0.012	93583.	0.000
		35	0.183	-0.001	93850.	0.000
		36	0.164	-0.005	94066.	0.000

7. *A3*– **The third specifications of the naïve benchmark model**

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.437129	37.27824	-0.119027	0.9053
D24_NET_LOAD(-168)	0.641108	0.037419	17.13340	0.0000

R-squared	0.411883	Mean dependent var	1.472728
Adjusted R-squared	0.411808	S.D. dependent var	1419.711
S.E. of regression	1088.828	Akaike info criterion	16.82384
Sum squared resid	9.42E+09	Schwarz criterion	16.82560
Log likelihood	-66822.31	Hannan-Quinn criter.	16.82445
F-statistic	5562.105	Durbin-Watson stat	0.052462
Prob(F-statistic)	0.000000		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_NET_LOAD(-168)	0.737957	0.007412	99.56104	0.0000
MA(1)	0.934840	0.004108	227.5426	0.0000
MA(24)	-0.049821	0.004112	-12.11595	0.0000

R-squared	0.871439	Mean dependent var	0.237995
Adjusted R-squared	0.871408	S.D. dependent var	1501.567
S.E. of regression	538.4573	Akaike info criterion	15.41565
Sum squared resid	2.41E+09	Schwarz criterion	15.41819
Log likelihood	-64187.78	Hannan-Quinn criter.	15.41652
Durbin-Watson stat	0.370209		

8. *A3*– **The fourth specification of the naïve benchmark model**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_NET_LOAD(-168)	0.584046	0.031278	18.67267	0.0000
MA(1)	1.569711	0.028223	55.61878	0.0000
MA(2)	1.658692	0.043520	38.11354	0.0000
MA(3)	1.517071	0.047816	31.72730	0.0000
MA(4)	1.365331	0.046945	29.08377	0.0000
MA(5)	1.223452	0.046860	26.10873	0.0000
MA(6)	1.152322	0.045861	25.12632	0.0000
MA(7)	1.075753	0.046814	22.97926	0.0000
MA(8)	1.006205	0.046123	21.81588	0.0000
MA(9)	0.974011	0.046140	21.10994	0.0000
MA(10)	0.921432	0.047842	19.25977	0.0000
MA(11)	0.750372	0.044599	16.82472	0.0000
MA(12)	0.464541	0.031701	14.65388	0.0000
MA(24)	-0.452062	0.022770	-19.85313	0.0000
MA(25)	-0.673809	0.036511	-18.45502	0.0000
MA(26)	-0.646493	0.041486	-15.58357	0.0000
MA(27)	-0.522839	0.041892	-12.48057	0.0000
MA(28)	-0.399069	0.039600	-10.07759	0.0000
MA(29)	-0.258546	0.036741	-7.037014	0.0000
MA(30)	-0.179422	0.031362	-5.721069	0.0000
MA(31)	-0.114115	0.022350	-5.105803	0.0000
MA(32)	-0.058420	0.011916	-4.902475	0.0000
MA(13)	0.167330	0.014516	11.52696	0.0000
MA(23)	0.074196	0.011914	6.227368	0.0000

R-squared	0.980861	Mean dependent var	1.472728
Adjusted R-squared	0.980806	S.D. dependent var	1419.711
S.E. of regression	196.6921	Akaike info criterion	13.40417
Sum squared resid	3.06E+08	Schwarz criterion	13.42526
Log likelihood	-53217.38	Hannan-Quinn criter.	13.41139
Durbin-Watson stat	1.938987		

9. *A3*– **The LM and Heteroscedasticity test of the fourth specification of the naïve benchmark model**

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	2.316585	Prob. F(3,7913)	0.0736
Obs*R-squared	6.958738	Prob. Chi-Square(3)	0.0732

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_NET_LOAD(-167)	0.004657	0.006550	0.711101	0.4770
D24_NET_LOAD(-168)	-0.003669	0.006354	-0.577382	0.5637
D24_NET_LOAD(-1)	-0.026636	0.033148	-0.803531	0.4217
D24_NET_LOAD(-2)	0.068873	0.096516	0.713596	0.4755
D24_NET_LOAD(-3)	-0.086750	0.138889	-0.624603	0.5322
D24_NET_LOAD(-4)	0.069439	0.129317	0.536972	0.5913
D24_NET_LOAD(-5)	-0.031821	0.078154	-0.407151	0.6839
D24_NET_LOAD(-6)	0.005664	0.023992	0.236095	0.8134
D24_NET_LOAD(-23)	0.000150	0.002495	0.060113	0.9521
D24_NET_LOAD(-26)	-0.000574	0.013806	-0.041587	0.9668
D24_NET_LOAD(-27)	0.001756	0.031809	0.055204	0.9560
D24_NET_LOAD(-28)	-0.003210	0.038824	-0.082688	0.9341
D24_NET_LOAD(-29)	0.002416	0.029572	0.081714	0.9349
D24_NET_LOAD(-30)	-0.000592	0.011180	-0.052916	0.9578
MA(1)	-0.146815	0.090789	-1.617097	0.1059
MA(2)	0.096376	0.115831	0.832037	0.4054
MA(3)	-0.068672	0.084909	-0.808767	0.4187
MA(4)	0.037780	0.047971	0.787561	0.4310
MA(6)	-0.014092	0.023044	-0.611534	0.5409
MA(7)	0.001174	0.010822	0.108444	0.9136
MA(8)	0.002354	0.006212	0.378872	0.7048
MA(11)	-0.006406	0.005607	-1.142506	0.2533
MA(24)	0.003846	0.007564	0.508530	0.6111
MA(25)	0.116289	0.072663	1.600387	0.1096
MA(26)	-0.065826	0.087367	-0.753448	0.4512
MA(27)	0.053105	0.062625	0.847983	0.3965
MA(28)	-0.029881	0.036697	-0.814266	0.4155
MA(30)	0.009226	0.015965	0.577867	0.5634
RESID(-1)	0.176482	0.097037	1.818709	0.0690
RESID(-2)	0.067206	0.050316	1.335679	0.1817
RESID(-3)	-0.018870	0.042109	-0.448107	0.6541

R-squared	0.000876	Mean dependent var	-0.236990
Adjusted R-squared	-0.002912	S.D. dependent var	191.7724
S.E. of regression	192.0514	Akaike info criterion	13.35730
Sum squared resid	2.92E+08	Schwarz criterion	13.38454
Log likelihood	-53024.19	Hannan-Quinn criter.	13.36662
Durbin-Watson stat	2.000746		

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	6.238248	Prob. F(14,7929)	0.0000
Obs*R-squared	86.54740	Prob. Chi-Square(14)	0.0000
Scaled explained SS	826.5737	Prob. Chi-Square(14)	0.0000

10. *A3*– **Wald test of the fourth specification of the naïve benchmark model**

Wald Test:
Equation: FINALRW

Test Statistic	Value	df	Probability
F-statistic	55219.49	(28, 7916)	0.0000
Chi-square	1546146.	28	0.0000

Null Hypothesis: C(1)=0,C(2)=0,C(3)=0,C(4)=0,
C(5)=0,C(6)=0,C(7)=0,C(8)=0,C(9)=0,
C(10)=0,C(11)=0,C(12)=0,C(13)=0,
C(14)=0,C(15)=0,C(16)=0,C(17)=0,
C(18)=0,C(19)=0,C(20)=0,C(21)=0,
C(22)=0,C(23)=0,C(24)=0,C(25)=0,
C(26)=0,C(27)=0,C(28)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	0.226744	0.010111
C(2)	-0.216492	0.009553
C(3)	2.508853	0.036754
C(4)	-3.140097	0.105360
C(5)	2.842575	0.149462
C(6)	-1.876450	0.136549
C(7)	0.829308	0.080402
C(8)	-0.180802	0.024045
C(9)	0.023930	0.004847
C(10)	-0.092504	0.015592
C(11)	0.164791	0.031080
C(12)	-0.172303	0.035719
C(13)	0.122430	0.026442
C(14)	-0.048349	0.010027
C(15)	-1.060635	0.045687
C(16)	0.957771	0.062979
C(17)	-0.618984	0.059273
C(18)	0.196859	0.033105
C(19)	0.062944	0.018052
C(20)	-0.062221	0.010219
C(21)	0.047652	0.006977
C(22)	0.058931	0.006304
C(23)	-0.772072	0.012533
C(24)	0.753567	0.040455
C(25)	-0.695775	0.050780
C(26)	0.450985	0.047075
C(27)	-0.136039	0.026476
C(28)	-0.044159	0.013733

Restrictions are linear in coefficients.

11. *A3*– **RESET test of the fourth specification of the naïve benchmark model**

Ramsey RESET Test
Equation: FINALRW
Specification: D24_NET_LOAD D24_NET_LOAD(-167 TO -168)
D24_NET_LOAD(-1 TO -6) D24_NET_LOAD(-23) D24_NET_LOAD(
-26TO -30) MA(1) MA(2) MA(3) MA(4) MA(6) MA(7) MA(8) MA(11)
MA(24) MA(25) MA(26) MA(27) MA(28) MA(30)
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	0.972811	7915	0.3307
F-statistic	0.946361	(1, 7915)	0.3307
Likelihood ratio	0.949771	1	0.3298

WARNING: the MA backcasts differ for the original and test equation.
Under the null hypothesis, the impact of this difference vanishes asymptotically.

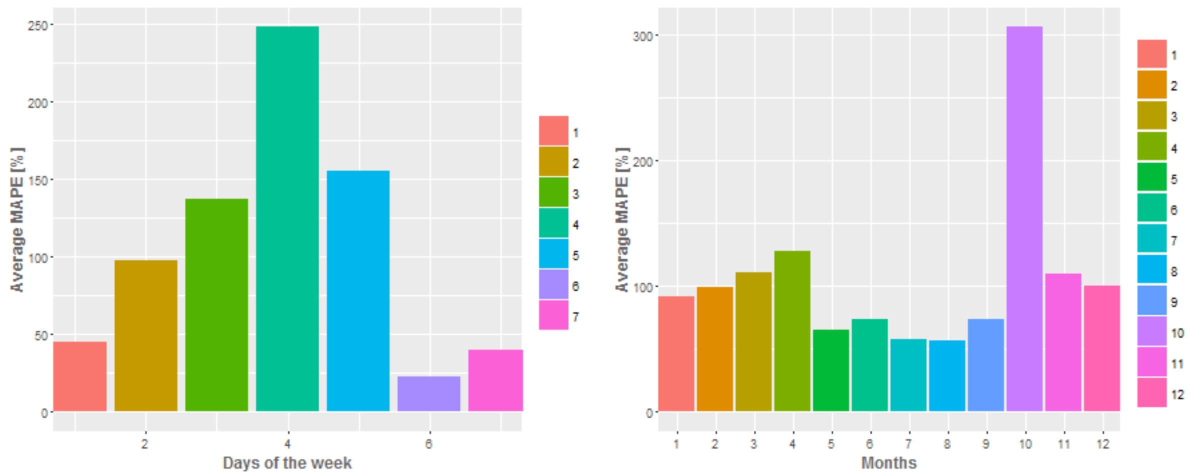
F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	34922.98	1	34922.98
Restricted SSR	2.92E+08	7916	36902.16
Unrestricted SSR	2.92E+08	7915	36902.41
Unrestricted SSR	2.92E+08	7915	36902.41

LR test summary:

	Value	df
Restricted LogL	-53027.68	7916
Unrestricted LogL	-53027.20	7915

12. *A3*– **Daily and Monthly average MAPE of the one-step-ahead forecast of the final naïve benchmark model, 2015**



13. *A3*– **Chow forecast test of the day-ahead forecast of the final naïve benchmark model for 2015**

Chow Forecast Test
Equation: FINALRW
Specification: D24_NET_LOAD D24_NET_LOAD(-167 TO -168)
D24_NET_LOAD(-1 TO -6) D24_NET_LOAD(-23) D24_NET_LOAD(
-26TO -30) MA(1) MA(2) MA(3) MA(4) MA(6) MA(7) MA(8) MA(11)
MA(24) MA(25) MA(26) MA(27) MA(28) MA(30)
Test predictions for observations from 2400 to 8136

	Value	df	Probability
F-statistic	0.984543	(5737, 2179)	0.6713
Likelihood ratio	10158.43	5737	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	2.11E+08	5737	36743.37
Restricted SSR	2.92E+08	7916	36902.16
Unrestricted SSR	81320776	2179	37320.23
Unrestricted SSR	81320776	2179	37320.23

LR test summary:

	Value	df
Restricted LogL	-53027.68	7916
Unrestricted LogL	-47948.46	2179

Unrestricted log likelihood adjusts test equation results to account for observations in forecast sample

Chow Forecast Test
Equation: FINALRW
Specification: D24_NET_LOAD D24_NET_LOAD(-167 TO -168)
D24_NET_LOAD(-1 TO -6) D24_NET_LOAD(-23) D24_NET_LOAD(
-26TO -30) MA(1) MA(2) MA(3) MA(4) MA(6) MA(7) MA(8) MA(11)
MA(24) MA(25) MA(26) MA(27) MA(28) MA(30)
Test predictions for observations from 5300 to 8136

	Value	df	Probability
F-statistic	1.165170	(2837, 5079)	0.0000
Likelihood ratio	3982.174	2837	0.0000

F-test summary:

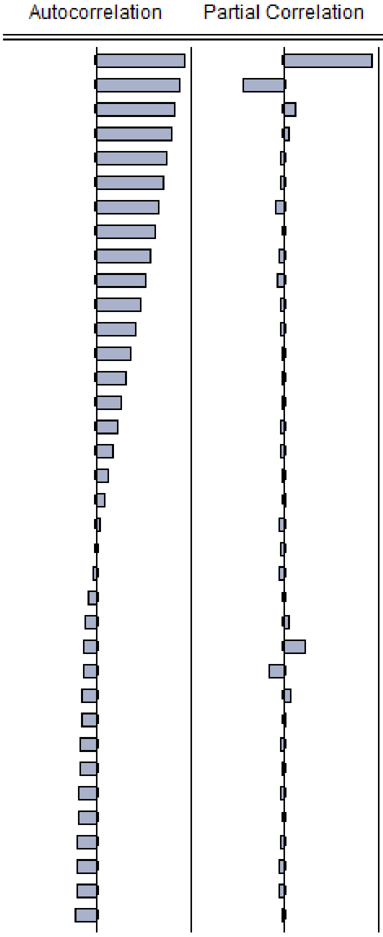
	Sum of Sq.	df	Mean Squares
Test SSR	1.15E+08	2837	40594.30
Restricted SSR	2.92E+08	7916	36902.16
Unrestricted SSR	1.77E+08	5079	34839.82
Unrestricted SSR	1.77E+08	5079	34839.82

LR test summary:

	Value	df
Restricted LogL	-53027.68	7916
Unrestricted LogL	-51036.59	5079

Unrestricted log likelihood adjusts test equation results to account for observations in forecast sample

14. *A3*– **The ACF and PACF of the seasonally double differenced net load**



	AC	PAC	Q-Stat	Prob
1	0.973	0.973	7530.2	0.000
2	0.924	-0.452	14313.	0.000
3	0.871	0.129	20345.	0.000
4	0.823	0.042	25733.	0.000
5	0.779	-0.033	30562.	0.000
6	0.737	-0.027	34875.	0.000
7	0.691	-0.096	38669.	0.000
8	0.642	-0.004	41951.	0.000
9	0.592	-0.062	44740.	0.000
10	0.540	-0.074	47056.	0.000
11	0.485	-0.038	48928.	0.000
12	0.430	-0.044	50399.	0.000
13	0.376	-0.012	51523.	0.000
14	0.324	-0.020	52357.	0.000
15	0.274	-0.013	52955.	0.000
16	0.226	-0.039	53360.	0.000
17	0.177	-0.047	53609.	0.000
18	0.129	-0.012	53741.	0.000
19	0.083	0.024	53796.	0.000
20	0.041	-0.053	53809.	0.000
21	-0.001	-0.037	53809.	0.000
22	-0.043	-0.054	53824.	0.000
23	-0.084	-0.001	53881.	0.000
24	-0.120	0.063	53995.	0.000
25	-0.140	0.228	54151.	0.000
26	-0.152	-0.164	54336.	0.000
27	-0.162	0.068	54546.	0.000
28	-0.171	0.015	54778.	0.000
29	-0.179	-0.032	55034.	0.000
30	-0.188	-0.023	55314.	0.000
31	-0.196	-0.047	55620.	0.000
32	-0.203	-0.006	55949.	0.000
33	-0.210	-0.044	56301.	0.000
34	-0.217	-0.050	56676.	0.000
35	-0.225	-0.064	57079.	0.000
36	-0.233	-0.015	57512.	0.000

15. *A3*– **The first specification of the double-seasonal ARMA model**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.485998	0.029151	50.97554	0.0000
AR(2)	-0.702646	0.044610	-15.75080	0.0000
AR(3)	0.190783	0.019262	9.904628	0.0000
AR(23)	0.158549	0.030482	5.201475	0.0000
AR(24)	-0.155917	0.028053	-5.558018	0.0000
AR(168)	-0.182411	0.029999	-6.080530	0.0000
AR(169)	0.177037	0.028889	6.128096	0.0000
MA(23)	-0.175031	0.014629	-11.96469	0.0000
MA(24)	-0.700021	0.015767	-44.39881	0.0000
MA(25)	-0.129187	0.010347	-12.48557	0.0000
MA(167)	0.156121	0.012840	12.15894	0.0000
MA(168)	-0.257109	0.013626	-18.86855	0.0000
MA(169)	0.112231	0.009900	11.33600	0.0000

R-squared	0.974733	Mean dependent var	-6.246654
Adjusted R-squared	0.974694	S.D. dependent var	1214.988
S.E. of regression	193.2781	Akaike info criterion	13.36781
Sum squared resid	2.90E+08	Schwarz criterion	13.37944
Log likelihood	-51954.35	Hannan-Quinn criter.	13.37180
Durbin-Watson stat	2.015921		

16. *A3*- **The LM test of the first specification of the double-seasonal ARMA model**

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	5.013366	Prob. F(2,7760)	0.0067
Obs*R-squared	9.580650	Prob. Chi-Square(2)	0.0083

17. *A3*- **The RESET test of the first specification of the double-seasonal ARMA model**

Ramsey RESET Test
Equation: DTOT_ARMA
Specification: DTOT_NET_LOAD AR(1) AR(2) AR(3) AR(23) AR(24) AR(168) AR(169) MA(23) MA(24) MA(25) MA(167) MA(168) MA(169)
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	1.320740	7761	0.1866
F-statistic	1.744354	(1, 7761)	0.1866
Likelihood ratio	1.747305	1	0.1862

18. *A3*- **The Chow breakpoint test of the first specification of the double-seasonal ARMA model**

Chow Breakpoint Test: 2400
Null Hypothesis: No breaks at specified breakpoints

Equation Sample: 362 5000

F-statistic	8.828362	Prob. F(13,4613)	0.0000
Log likelihood ratio	114.0032	Prob. Chi-Square(13)	0.0000

Chow Breakpoint Test: 5300
Null Hypothesis: No breaks at specified breakpoints

Equation Sample: 362 8136

F-statistic	5.937069	Prob. F(13,7749)	0.0000
Log likelihood ratio	77.05774	Prob. Chi-Square(13)	0.0000

19. *A3*– **The RESET, Heteroscedasticity and Wald test of the final specification of the "double-seasonal ARMA model"**

Ramsey RESET Test
Equation: DTOT_MA_LAGS
Specification: DTOT_NET_LOAD DTOT_NET_LOAD(-1 TO -4)
DTOT_NET_LOAD(-168 TO-170) MA(23) MA(167) MA(168) MA(1)
DTOT_NET_LOAD(-23) DTOT_NET_LOAD(-11) DTOT_NET_LOAD(
-18 TO -19) DTOT_NET_LOAD(-16) DTOT_NET_LOAD(-25)
DTOT_NET_LOAD(-27)MA(2)MA(3)MA(6)MA(7)MA(24)
MA(26)MA(27)MA(30) MA(9) MA(12)
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	1.352387	7745	0.1763
F-statistic	1.828951	(1, 7745)	0.1763
Likelihood ratio	1.835583	1	0.1755

WARNING: the MA backcasts differ for the original and test equation. Under the null hypothesis, the impact of this difference vanishes asymptotically.

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	66240.47	1	66240.47
Restricted SSR	2.81E+08	7746	36221.61
Unrestricted SSR	2.81E+08	7745	36217.74
Unrestricted SSR	2.81E+08	7745	36217.74

LR test summary:

	Value	df
Restricted LogL	-51820.24	7746
Unrestricted LogL	-51819.32	7745

Heteroskedasticity Test: Harvey

F-statistic	1.819462	Prob. F(15,7758)	0.0266
Obs*R-squared	27.25235	Prob. Chi-Square(15)	0.0267
Scaled explained SS	31.84796	Prob. Chi-Square(15)	0.0068

Wald Test:
Equation: DTOT_MA_LAGS

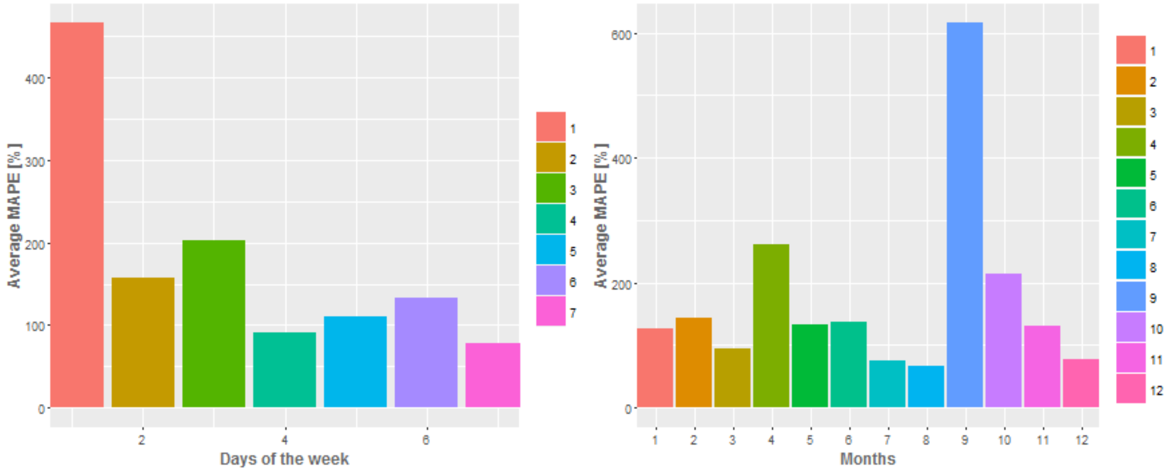
Test Statistic	Value	df	Probability
F-statistic	31183.46	(28, 7746)	0.0000
Chi-square	873136.9	28	0.0000

Null Hypothesis: C(1)=0,C(2)=0,C(3)=0,C(4)=0,
C(5)=0,C(6)=0,C(7)=0,C(8)=0,C(9)=0,
C(10)=0,C(11)=0,C(12)=0,C(13)=0,
C(14)=0,C(15)=0,C(16)=0,C(17)=0,
C(18)=0,C(19)=0,C(20)=0,C(21)=0,
C(22)=0,C(23)=0,C(24)=0,C(25)=0,
C(26)=0,C(27)=0,C(28)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	1.517865	0.030551
C(2)	-0.446246	0.046482
C(3)	-0.155827	0.039967
C(4)	0.077533	0.017348
C(5)	-0.305443	0.027766
C(6)	0.482028	0.041755
C(7)	-0.180587	0.019074
C(8)	0.114546	0.019808
C(9)	-0.014418	0.004365
C(10)	-0.068195	0.022665
C(11)	0.053116	0.020219
C(12)	0.021063	0.007917
C(13)	-0.177737	0.022398
C(14)	0.062009	0.010282
C(15)	-0.124965	0.011793
C(16)	0.086158	0.010373
C(17)	-0.103925	0.012440
C(18)	-0.046213	0.008817
C(19)	-0.291605	0.033298
C(20)	-0.102289	0.024955
C(21)	0.056405	0.024866
C(22)	-0.012679	0.005980
C(23)	-0.797394	0.014263
C(24)	0.260387	0.032245
C(25)	0.124648	0.024064
C(26)	-0.049571	0.024497
C(27)	0.011507	0.003959
C(28)	-0.005468	0.002525

Restrictions are linear in coefficients.

20. *A3*- **Daily and Monthly average MAPE of one-step-ahead forecast of the final specification of the "double-seasonal ARMA model" for 2015**



21. *A3*– **The Chow forecast test of the final specification of the "double-seasonal ARMA model"**

Chow Forecast Test
Equation: DTOT_MA_LAGS
Specification: DTOT_NET_LOAD DTOT_NET_LOAD(-1 TO -4)
DTOT_NET_LOAD(-168 TO-170) MA(23) MA(167) MA(168) MA(1)
DTOT_NET_LOAD(-23) DTOT_NET_LOAD(-11) DTOT_NET_LOAD(
-18 TO -19) DTOT_NET_LOAD(-16) DTOT_NET_LOAD(-25)
DTOT_NET_LOAD(-27)MA(2)MA(3)MA(6)MA(7)MA(24)
MA(26)MA(27)MA(30) MA(9) MA(12)
Test predictions for observations from 2400 to 8136

	Value	df	Probability
F-statistic	0.987351	(5737, 2009)	0.6381
Likelihood ratio	10418.15	5737	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	2.07E+08	5737	36101.66
Restricted SSR	2.81E+08	7746	36221.61
Unrestricted SSR	73457398	2009	36564.16
Unrestricted SSR	73457398	2009	36564.16

LR test summary:

	Value	df
Restricted LogL	-51820.24	7746
Unrestricted LogL	-46611.17	2009

Unrestricted log likelihood adjusts test equation results to account for observations in forecast sample

Chow Forecast Test
Equation: DTOT_MA_LAGS
Specification: DTOT_NET_LOAD DTOT_NET_LOAD(-1 TO -4)
DTOT_NET_LOAD(-168 TO-170) MA(23) MA(167) MA(168) MA(1)
DTOT_NET_LOAD(-23) DTOT_NET_LOAD(-11) DTOT_NET_LOAD(
-18 TO -19) DTOT_NET_LOAD(-16) DTOT_NET_LOAD(-25)
DTOT_NET_LOAD(-27)MA(2)MA(3)MA(6)MA(7)MA(24)
MA(26)MA(27)MA(30) MA(9) MA(12)
Test predictions for observations from 5300 to 8136

	Value	df	Probability
F-statistic	1.151100	(2837, 4909)	0.0000
Likelihood ratio	3964.508	2837	0.0000

F-test summary:

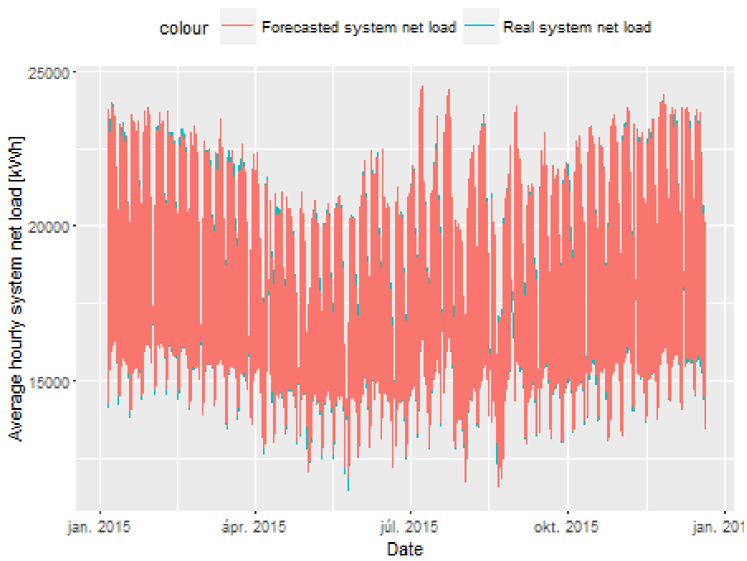
	Sum of Sq.	df	Mean Squares
Test SSR	1.12E+08	2837	39508.27
Restricted SSR	2.81E+08	7746	36221.61
Unrestricted SSR	1.68E+08	4909	34322.19
Unrestricted SSR	1.68E+08	4909	34322.19

LR test summary:

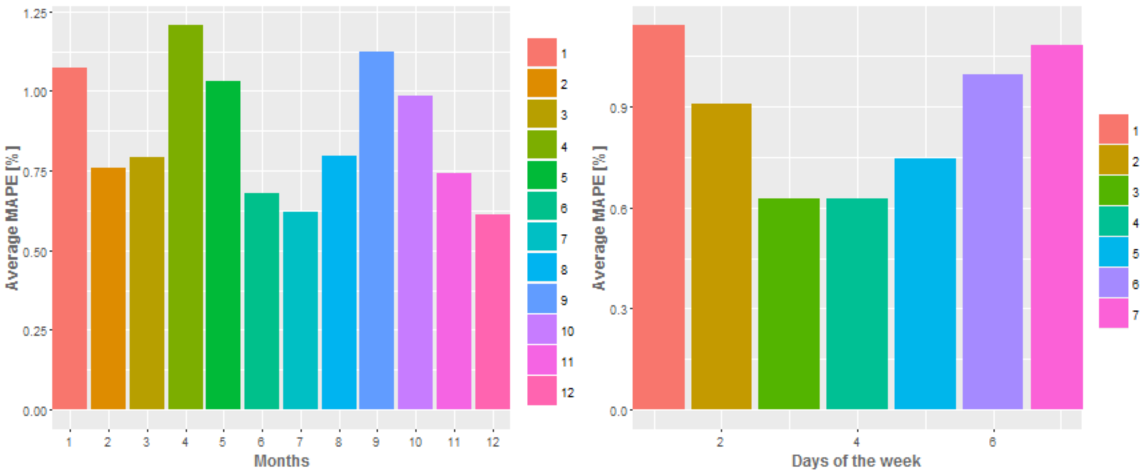
	Value	df
Restricted LogL	-51820.24	7746
Unrestricted LogL	-49837.99	4909

Unrestricted log likelihood adjusts test equation results to account for observations in forecast sample

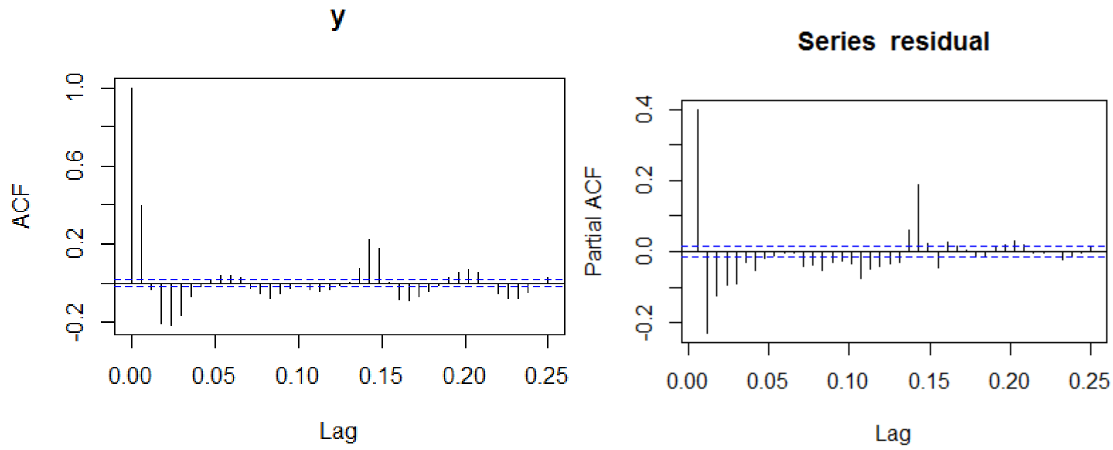
22. *A3* **– Forecasted system net load for 2015 with the Taylor (2003) exponential smoothing**



23. *A3*- **Daily and Monthly average MAPE of the one-step-ahead forecast of the Taylor (2003) exponential smoothing for 2015**



24. *A3*– **The ACF and PACF of the residuals of the of the day-ahead forecast of the Taylor (2003) exponential smoothing for 2015**



25. *A3*– **The first specification of the model with the exact-day matching**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	299.0817	27.83755	10.74382	0.0000
MEAN_NET_LOAD	0.982088	0.001586	619.0711	0.0000
DEMEAN_NET_LOAD(-1)	0.972511	0.002387	407.3882	0.0000
TEMP_V*D1	-0.256117	5.914901	-0.043300	0.9655
TEMP_V*D2	10.12413	5.062075	1.999995	0.0455
TEMP_V(-1)*D1	6.541399	5.923027	1.104401	0.2695
TEMP_V(-1)*D2	-7.164025	5.064662	-1.414512	0.1572

R-squared	0.984404	Mean dependent var	18100.76
Adjusted R-squared	0.984392	S.D. dependent var	2598.780
S.E. of regression	324.6656	Akaike info criterion	14.40433
Sum squared resid	8.57E+08	Schwarz criterion	14.41036
Log likelihood	-58575.40	Hannan-Quinn criter.	14.40639
F-statistic	85494.75	Durbin-Watson stat	1.148821
Prob(F-statistic)	0.000000		

26. *A3*– **Wald test of the Temperature variables of the first specification of the model with the exact-day matching**

Wald Test:
Equation: ERG1

Test Statistic	Value	df	Probability
F-statistic	0.917869	(2, 8127)	0.3994
Chi-square	1.835738	2	0.3994

Null Hypothesis: C(6)=0, C(7)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(6)	6.541399	7.959555
C(7)	-7.164025	6.751411

27. *A3*- **The second specification of the model with the exact-day matching**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	300.8572	27.80024	10.82211	0.0000
MEAN_NET_LOAD	0.981997	0.001583	620.4272	0.0000
DEMEAN_NET_LOAD(-1)	0.972440	0.002385	407.6834	0.0000
TEMP_V*D1	6.178001	1.345430	4.591842	0.0000
TEMP_V*D2	2.977166	0.666043	4.469931	0.0000

R-squared	0.984398	Mean dependent var	18100.76
Adjusted R-squared	0.984390	S.D. dependent var	2598.780
S.E. of regression	324.6903	Akaike info criterion	14.40424
Sum squared resid	8.57E+08	Schwarz criterion	14.40854
Log likelihood	-58577.02	Hannan-Quinn criter.	14.40571
F-statistic	128221.8	Durbin-Watson stat	1.148449
Prob(F-statistic)	0.000000		

28. *A3*– **The LM test of the second specification of the model with the exact-day matching**

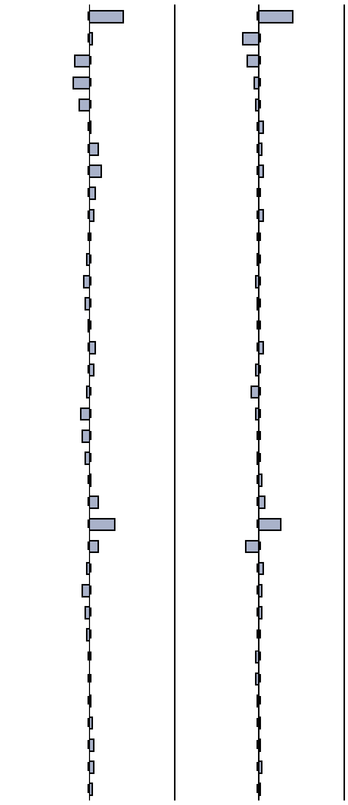
Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1103.710	Prob. F(2,8127)	0.0000
Obs*R-squared	1737.413	Prob. Chi-Square(2)	0.0000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.30200	24.66249	0.458267	0.6468
MEAN_NET_LOAD	-0.000523	0.001404	-0.372378	0.7096
DEMEAN_NET_LOAD(-1)	-0.012193	0.002227	-5.475506	0.0000
TEMP_V*D1	0.386384	1.193961	0.323615	0.7462
TEMP_V*D2	0.056488	0.590743	0.095622	0.9238
RESID(-1)	0.512428	0.010926	46.90121	0.0000
RESID(-2)	-0.175101	0.011188	-15.65123	0.0000

R-squared	0.213599	Mean dependent var	-3.67E-13
Adjusted R-squared	0.213018	S.D. dependent var	324.6104
S.E. of regression	287.9684	Akaike info criterion	14.16444
Sum squared resid	6.74E+08	Schwarz criterion	14.17047
Log likelihood	-57599.77	Hannan-Quinn criter.	14.16650
F-statistic	367.9033	Durbin-Watson stat	2.048500

29. *A3*– **The ACF and PACF of the residuals of the second specification of the model with the exact-day matching and the third specification**



Lag	AC	PAC	Q-Stat	Prob
1	0.426	0.426	1474.8	0.000
2	0.026	-0.190	1480.3	0.000
3	-0.188	-0.153	1769.5	0.000
4	-0.204	-0.059	2107.2	0.000
5	-0.134	-0.043	2253.7	0.000
6	0.016	0.062	2255.7	0.000
7	0.099	0.026	2335.1	0.000
8	0.133	0.055	2480.2	0.000
9	0.074	-0.004	2524.6	0.000
10	0.045	0.056	2541.1	0.000
11	-0.006	-0.000	2541.4	0.000
12	-0.040	-0.010	2554.7	0.000
13	-0.080	-0.048	2607.3	0.000
14	-0.062	-0.012	2638.2	0.000
15	-0.019	-0.000	2641.2	0.000
16	0.072	0.063	2683.6	0.000
17	0.064	-0.027	2716.7	0.000
18	-0.028	-0.086	2723.3	0.000
19	-0.100	-0.046	2804.5	0.000
20	-0.084	0.002	2862.3	0.000
21	-0.056	-0.023	2888.0	0.000
22	0.020	0.027	2891.2	0.000
23	0.106	0.069	2982.0	0.000
24	0.299	0.269	3710.8	0.000
25	0.099	-0.166	3791.2	0.000
26	-0.029	0.041	3798.1	0.000
27	-0.084	0.034	3855.8	0.000
28	-0.058	0.028	3883.2	0.000
29	-0.038	-0.003	3894.7	0.000
30	0.002	-0.031	3894.7	0.000
31	-0.002	-0.043	3894.7	0.000
32	0.024	-0.010	3899.6	0.000
33	0.031	0.012	3907.4	0.000
34	0.057	0.022	3934.0	0.000
35	0.049	0.025	3953.7	0.000
36	0.029	0.015	3960.8	0.000

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	369.7019	37.31385	9.907899	0.0000
MEAN_NET_LOAD	0.757293	0.020402	37.11839	0.0000
DEMEAN_NET_LOAD(-1)	0.752438	0.020183	37.28157	0.0000
TEMP_V*D1	6.197113	1.094614	5.661462	0.0000
TEMP_V*D2	3.169993	0.719054	4.408558	0.0000
NET_LOAD(-1)	0.515380	0.036098	14.27722	0.0000
NET_LOAD(-2)	-0.413421	0.019563	-21.13292	0.0000
NET_LOAD(-3)	0.118815	0.005750	20.66196	0.0000

R-squared	0.986986	Mean dependent var	18101.81
Adjusted R-squared	0.986975	S.D. dependent var	2598.238
S.E. of regression	296.5282	Akaike info criterion	14.22314
Sum squared resid	7.14E+08	Schwarz criterion	14.23003
Log likelihood	-57823.31	Hannan-Quinn criter.	14.22550
F-statistic	88020.36	Durbin-Watson stat	1.544496
Prob(F-statistic)	0.000000		

30. *A3*– **RESET test of the third specification of the model with the exact-day matching**

Ramsey RESET Test
Equation: ERG1
Specification: NET_LOAD C MEAN_NET_LOAD DEMEAN_NET_LOAD(-1)
TEMP_V*D1 TEMP_V*D2 NET_LOAD(-1 TO -3)
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	7.041718	8123	0.0000
F-statistic	49.58580	(1, 8123)	0.0000
Likelihood ratio	49.48984	1	0.0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	4334107.	1	4334107.
Restricted SSR	7.14E+08	8124	87928.95
Unrestricted SSR	7.10E+08	8123	87406.22
Unrestricted SSR	7.10E+08	8123	87406.22

LR test summary:

	Value	df
Restricted LogL	-57823.31	8124
Unrestricted LogL	-57798.56	8123

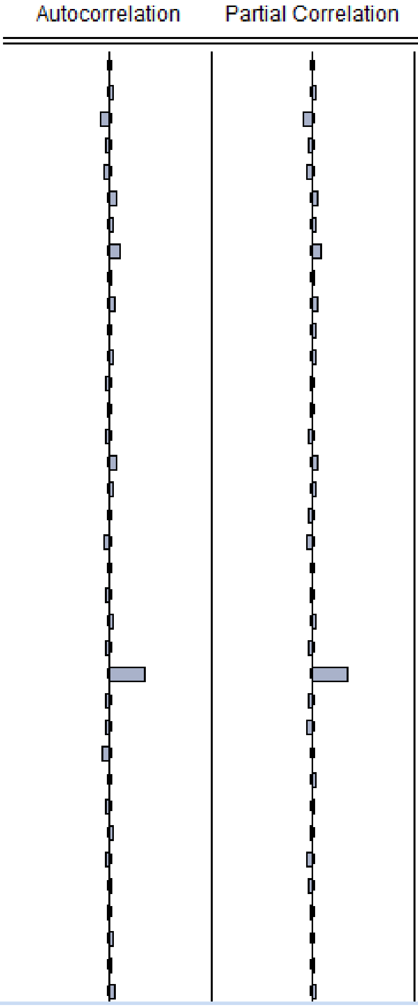
31. *A3*– **The fourth specification of the model with the exact-day matching**

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed
bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	358.4734	72.37289	4.953144	0.0000
MEAN_NET_LOAD	0.849455	0.037174	22.85071	0.0000
MEAN_NET_LOAD(-1)	0.241972	0.067391	3.590546	0.0003
MEAN_NET_LOAD(-2)	-0.112508	0.033939	-3.314980	0.0009
TEMP_V*D1	4.804909	1.011843	4.748669	0.0000
TEMP_V*D2	2.999902	0.759125	3.951788	0.0001
DEMEAN_NET_LOAD(-1)	1.493487	0.057824	25.82810	0.0000
DEMEAN_NET_LOAD(-2)	-0.697800	0.102245	-6.824776	0.0000
DEMEAN_NET_LOAD(-3)	0.174174	0.047290	3.683116	0.0002

R-squared	0.988322	Mean dependent var	18101.81
Adjusted R-squared	0.988311	S.D. dependent var	2598.238
S.E. of regression	280.9109	Akaike info criterion	14.11506
Sum squared resid	6.41E+08	Schwarz criterion	14.12281
Log likelihood	-57382.83	Hannan-Quinn criter.	14.11771
F-statistic	85935.64	Durbin-Watson stat	2.017164
Prob(F-statistic)	0.000000		

32. *A3*– **The ACF and PACF of the fourth specification of the model with the exact-day matching**



	AC	PAC	Q-Stat	Prob
1	-0.002	-0.002	0.0211	0.884
2	0.034	0.034	9.2419	0.010
3	-0.090	-0.090	74.322	0.000
4	-0.038	-0.039	85.879	0.000
5	-0.062	-0.056	116.84	0.000
6	0.070	0.065	156.77	0.000
7	0.038	0.036	168.32	0.000
8	0.107	0.093	260.88	0.000
9	0.009	0.014	261.48	0.000
10	0.045	0.048	278.21	0.000
11	0.007	0.035	278.65	0.000
12	0.031	0.039	286.48	0.000
13	-0.037	-0.024	297.59	0.000
14	-0.009	-0.019	298.32	0.000
15	-0.042	-0.039	312.63	0.000
16	0.071	0.056	353.80	0.000
17	0.041	0.035	367.33	0.000
18	0.000	-0.029	367.33	0.000
19	-0.054	-0.058	391.22	0.000
20	0.003	0.004	391.29	0.000
21	-0.033	-0.014	400.41	0.000
22	0.039	0.029	412.72	0.000
23	-0.039	-0.046	425.05	0.000
24	0.362	0.356	1493.1	0.000
25	-0.034	-0.026	1502.5	0.000
26	-0.029	-0.054	1509.4	0.000
27	-0.068	-0.007	1546.7	0.000
28	0.007	0.027	1547.1	0.000
29	-0.025	0.012	1552.1	0.000
30	0.040	-0.022	1565.0	0.000
31	-0.035	-0.059	1575.1	0.000
32	0.023	-0.045	1579.3	0.000
33	-0.014	-0.015	1580.9	0.000
34	0.028	-0.001	1587.5	0.000
35	0.011	0.009	1588.4	0.000
36	0.046	0.026	1605.9	0.000

33. *A3*– **The fifth specification of the model with the exact-day matching and its LM test**

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed
bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	351.9691	70.76167	4.974008	0.0000
MEAN_NET_LOAD	0.840093	0.039874	21.06866	0.0000
MEAN_NET_LOAD(-1)	0.239830	0.067000	3.579537	0.0003
MEAN_NET_LOAD(-2)	-0.111516	0.033729	-3.306264	0.0009
TEMP_V*D1	4.983267	1.080641	4.611398	0.0000
TEMP_V*D2	3.129881	0.788280	3.970520	0.0001
DEMEAN_NET_LOAD(-1)	1.486768	0.058710	25.32389	0.0000
DEMEAN_NET_LOAD(-2)	-0.690439	0.102580	-6.730711	0.0000
DEMEAN_NET_LOAD(-3)	0.167850	0.048092	3.490173	0.0005
NET_LOAD(-24)	0.010762	0.004222	2.549177	0.0108

R-squared	0.988359	Mean dependent var	18105.74
Adjusted R-squared	0.988346	S.D. dependent var	2598.335
S.E. of regression	280.5012	Akaike info criterion	14.11227
Sum squared resid	6.37E+08	Schwarz criterion	14.12090
Log likelihood	-57222.29	Hannan-Quinn criter.	14.11522
F-statistic	76421.18	Durbin-Watson stat	2.003059
Prob(F-statistic)	0.000000		

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.366927	Prob. F(1,8100)	0.5447
Obs*R-squared	0.367408	Prob. Chi-Square(1)	0.5444

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.304379	28.41425	-0.151487	0.8796
MEAN_NET_LOAD	0.000979	0.007280	0.134417	0.8931
MEAN_NET_LOAD(-1)	0.000706	0.012025	0.058722	0.9532
MEAN_NET_LOAD(-2)	-0.000998	0.006904	-0.144532	0.8851
TEMP_V*D1	-0.144862	1.188542	-0.121882	0.9030
TEMP_V*D2	-0.049893	0.587082	-0.084985	0.9323
DEMEAN_NET_LOAD(-1)	0.026405	0.044922	0.587790	0.5567
DEMEAN_NET_LOAD(-2)	-0.036641	0.063068	-0.580978	0.5613
DEMEAN_NET_LOAD(-3)	0.011339	0.021558	0.525998	0.5989
NET_LOAD(-24)	-0.000426	0.002772	-0.153670	0.8779
RESID(-1)	-0.028077	0.046352	-0.605745	0.5447

R-squared	0.000045	Mean dependent var	4.27E-12
Adjusted R-squared	-0.001189	S.D. dependent var	280.3455
S.E. of regression	280.5122	Akaike info criterion	14.11247
Sum squared resid	6.37E+08	Schwarz criterion	14.12196
Log likelihood	-57222.11	Hannan-Quinn criter.	14.11571
F-statistic	0.036693	Durbin-Watson stat	2.001129
Prob(F-statistic)	0.999999		

34. *A3* **- The sixth specification of the model with the exact-day matching on differenced net load**

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	0.848907	0.021131	40.17340	0.0000
D24_DEMEAN_NET_LOAD(-1)	0.940689	0.005897	159.5202	0.0000
D24_TEMP_V*D1	7.271383	7.911724	0.919064	0.3581
D24_TEMP_V*D2	30.44331	8.820174	3.451554	0.0006
D24_TEMP_V(-1)*D1	5.097942	8.126116	0.627353	0.5304
D24_TEMP_V(-1)*D2	-25.87723	8.942083	-2.893871	0.0038

R-squared	0.929943	Mean dependent var	2.956045
Adjusted R-squared	0.929900	S.D. dependent var	1418.399
S.E. of regression	375.5407	Akaike info criterion	14.69535
Sum squared resid	1.14E+09	Schwarz criterion	14.70053
Log likelihood	-59583.65	Hannan-Quinn criter.	14.69712
Durbin-Watson stat	1.040204		

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	0.848844	0.021124	40.18416	0.0000
D24_DEMEAN_NET_LOAD(-1)	0.940740	0.005896	159.5464	0.0000
D24_TEMP_V*D1	11.93046	2.693984	4.428555	0.0000
D24_TEMP_V*D2	30.44377	8.820009	3.451671	0.0006
D24_TEMP_V(-1)*D2	-25.88007	8.941643	-2.894330	0.0038

R-squared	0.929939	Mean dependent var	2.956045
Adjusted R-squared	0.929905	S.D. dependent var	1418.399
S.E. of regression	375.5285	Akaike info criterion	14.69516
Sum squared resid	1.14E+09	Schwarz criterion	14.69948
Log likelihood	-59583.88	Hannan-Quinn criter.	14.69664
Durbin-Watson stat	1.040204		

35. *A3*– **The ACF and PACF of the residuals of the sixth specification of the model with the exact-day matching and further re-specifications**

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.480	0.480	1867.8	0.000
		2	0.069	-0.210	1906.1	0.000
		3	-0.141	-0.109	2066.9	0.000
		4	-0.159	-0.026	2273.1	0.000
		5	-0.068	0.024	2311.1	0.000
		6	0.060	0.073	2340.4	0.000
		7	0.109	0.020	2436.2	0.000
		8	0.130	0.073	2572.9	0.000
		9	0.107	0.042	2665.7	0.000
		10	0.074	0.044	2710.5	0.000
		11	0.015	-0.008	2712.4	0.000
		12	-0.054	-0.044	2736.1	0.000
		13	-0.090	-0.035	2801.5	0.000
		14	-0.074	-0.022	2845.9	0.000
		15	-0.024	-0.004	2850.5	0.000
		16	0.029	0.003	2857.4	0.000
		17	0.024	-0.037	2862.0	0.000
		18	-0.060	-0.096	2890.9	0.000
		19	-0.076	0.007	2937.7	0.000
		20	-0.035	0.023	2947.9	0.000
		21	-0.001	-0.002	2947.9	0.000
		22	-0.016	-0.041	2949.9	0.000
		23	-0.061	-0.051	2980.3	0.000
		24	-0.148	-0.112	3159.1	0.000
		25	-0.035	0.134	3168.9	0.000
		26	-0.022	-0.099	3172.7	0.000
		27	-0.003	0.008	3172.7	0.000
		28	0.019	0.029	3175.7	0.000
		29	0.026	0.019	3181.3	0.000
		30	0.011	0.012	3182.3	0.000
		31	-0.020	-0.037	3185.6	0.000
		32	-0.013	0.045	3187.0	0.000
		33	0.028	0.048	3193.6	0.000
		34	0.062	0.052	3225.3	0.000
		35	0.077	0.032	3273.9	0.000
		36	0.062	-0.002	3305.2	0.000

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.143638	2.846612	-0.050459	0.9598
D24_MEAN_NET_LOAD	-0.175437	0.057338	-3.059714	0.0022
D24_MEAN_NET_LOAD(-1)	1.749462	0.059528	29.38870	0.0000
D24_MEAN_NET_LOAD(-2)	-0.640078	0.024235	-26.41176	0.0000
D24_DEMEAN_NET_LOAD(-1)	1.599249	0.021313	75.03519	0.0000
D24_DEMEAN_NET_LOAD(-2)	-0.830124	0.035274	-23.53328	0.0000
D24_DEMEAN_NET_LOAD(-3)	0.198007	0.017600	11.25073	0.0000
D24_TEMP_V*D1	5.372039	1.292557	4.156132	0.0000
D24_TEMP_V*D2	19.07988	5.637895	3.384220	0.0007
D24_TEMP_V(-1)*D2	-17.56207	5.731883	-3.063926	0.0022

R-squared	0.967731	Mean dependent var	2.965853
Adjusted R-squared	0.967695	S.D. dependent var	1418.574
S.E. of regression	254.9681	Akaike info criterion	13.92139
Sum squared resid	5.26E+08	Schwarz criterion	13.93002
Log likelihood	-56427.30	Hannan-Quinn criter.	13.92434
F-statistic	26983.90	Durbin-Watson stat	1.940911
Prob(F-statistic)	0.000000		

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	-0.175437	0.057335	-3.059887	0.0022
D24_MEAN_NET_LOAD(-1)	1.749463	0.059521	29.39230	0.0000
D24_MEAN_NET_LOAD(-2)	-0.640078	0.024225	-26.42171	0.0000
D24_DEMEAN_NET_LOAD(-1)	1.599249	0.021309	75.05146	0.0000
D24_DEMEAN_NET_LOAD(-2)	-0.830124	0.035269	-23.53702	0.0000
D24_DEMEAN_NET_LOAD(-3)	0.198007	0.017599	11.25079	0.0000
D24_TEMP_V*D1	5.369597	1.289157	4.165200	0.0000
D24_TEMP_V*D2	19.07976	5.637718	3.384306	0.0007
D24_TEMP_V(-1)*D2	-17.56093	5.729995	-3.064738	0.0022

R-squared	0.967731	Mean dependent var	2.965853
Adjusted R-squared	0.967699	S.D. dependent var	1418.574
S.E. of regression	254.9524	Akaike info criterion	13.92114
Sum squared resid	5.26E+08	Schwarz criterion	13.92891
Log likelihood	-56427.30	Hannan-Quinn criter.	13.92380
Durbin-Watson stat	1.940910		

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	-0.156932	0.055161	-2.844959	0.0045
D24_DEMEAN_NET_LOAD(-1)	-0.133546	0.050625	-2.637977	0.0084
D24_TEMP_V*D2	19.30455	5.448045	3.543390	0.0004
D24_TEMP_V(-1)*D2	-17.96129	5.583816	-3.216670	0.0013
D24_TEMP_V(-1)*D1	5.008207	1.424248	3.516388	0.0004
D24_NET_LOAD(-1)	1.761574	0.049484	35.59896	0.0000
D24_NET_LOAD(-2)	-0.919797	0.045201	-20.34907	0.0000

D24_NET_LOAD(-3)	0.240234	0.026673	9.006681	0.0000
D24_NET_LOAD(-5)	0.044787	0.011986	3.736773	0.0002
D24_NET_LOAD(-6)	-0.020762	0.008430	-2.462820	0.0138

R-squared	0.968470	Mean dependent var	2.673242
Adjusted R-squared	0.968435	S.D. dependent var	1418.710
S.E. of regression	252.0559	Akaike info criterion	13.89841
Sum squared resid	5.14E+08	Schwarz criterion	13.90705
Log likelihood	-56313.31	Hannan-Quinn criter.	13.90137
Durbin-Watson stat	2.000854		

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.010	-0.010	0.7425	0.389
		2	0.019	0.019	3.6525	0.161
		3	-0.013	-0.013	5.0369	0.169
		4	0.007	0.006	5.4071	0.248
		5	-0.048	-0.048	24.301	0.000
		6	0.110	0.109	121.69	0.000
		7	0.005	0.009	121.90	0.000
		8	0.065	0.061	155.92	0.000
		9	0.060	0.065	185.11	0.000
		10	0.020	0.016	188.40	0.000
		11	0.058	0.070	215.76	0.000
		12	-0.020	-0.031	219.15	0.000
		13	-0.025	-0.023	224.11	0.000
		14	-0.024	-0.031	228.61	0.000
		15	-0.030	-0.046	236.03	0.000
		16	0.012	0.009	237.28	0.000
		17	0.046	0.022	254.23	0.000
		18	-0.091	-0.098	322.26	0.000
		19	-0.014	-0.023	323.81	0.000
		20	-0.016	-0.017	325.81	0.000
		21	0.008	0.019	326.34	0.000
		22	-0.010	-0.004	327.17	0.000
		23	-0.005	-0.011	327.37	0.000
		24	-0.233	-0.216	770.51	0.000
		25	0.072	0.075	812.73	0.000
		26	-0.052	-0.035	834.45	0.000
		27	-0.015	-0.019	836.34	0.000
		28	-0.012	-0.005	837.44	0.000
		29	0.005	-0.007	837.67	0.000
		30	0.005	0.066	837.90	0.000
		31	-0.026	-0.040	843.19	0.000
		32	-0.029	0.000	850.24	0.000
		33	0.007	0.026	850.68	0.000
		34	0.013	0.022	852.16	0.000
		35	0.002	0.049	852.18	0.000
		36	0.036	0.007	862.57	0.000

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	-0.157018	0.055147	-2.847273	0.0044
D24_DEMEAN_NET_LOAD(-1)	-0.133616	0.050608	-2.640242	0.0083
D24_TEMP_V*D2	19.31576	5.450732	3.543700	0.0004
D24_TEMP_V(-1)*D2	-17.97342	5.587749	-3.216576	0.0013
D24_TEMP_V(-1)*D1	5.012959	1.424519	3.519053	0.0004
D24_NET_LOAD(-1)	1.760863	0.049851	35.32229	0.0000
D24_NET_LOAD(-2)	-0.915376	0.050088	-18.27520	0.0000
D24_NET_LOAD(-3)	0.228777	0.042490	5.384303	0.0000
D24_NET_LOAD(-4)	0.015408	0.027624	0.557777	0.5770
D24_NET_LOAD(-5)	0.034074	0.018341	1.857820	0.0632
D24_NET_LOAD(-6)	-0.017612	0.008778	-2.006355	0.0449

R-squared	0.968472	Mean dependent var	2.673242
Adjusted R-squared	0.968433	S.D. dependent var	1418.710
S.E. of regression	252.0648	Akaike info criterion	13.89861
Sum squared resid	5.14E+08	Schwarz criterion	13.90811
Log likelihood	-56313.10	Hannan-Quinn criter.	13.90186
Durbin-Watson stat	1.998580		

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	2.675872	Prob. F(2,8093)	0.0689
Obs*R-squared	5.343443	Prob. Chi-Square(2)	0.0691

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	0.005576	0.015902	0.350644	0.7259
D24_DEMEAN_NET_LOAD(-1)	-0.001627	0.015763	-0.103233	0.9178
D24_TEMP_V*D2	-1.802994	4.644213	-0.388224	0.6979
D24_TEMP_V(-1)*D2	1.180916	4.588104	0.257387	0.7969
D24_TEMP_V(-1)*D1	-1.152921	2.058426	-0.560098	0.5754
D24_NET_LOAD(-1)	0.211661	0.100383	2.108540	0.0350
D24_NET_LOAD(-2)	-0.308750	0.138638	-2.227022	0.0260
D24_NET_LOAD(-3)	0.130871	0.058810	2.225312	0.0261
D24_NET_LOAD(-5)	-0.025811	0.018742	-1.377154	0.1685
D24_NET_LOAD(-6)	0.000532	0.011546	0.046092	0.9632
RESID(-1)	-0.213348	0.098744	-2.160628	0.0308
RESID(-2)	-0.030417	0.034198	-0.889448	0.3738

R-squared	0.000659	Mean dependent var	-0.315354
Adjusted R-squared	-0.000699	S.D. dependent var	251.9157
S.E. of regression	252.0037	Akaike info criterion	13.89824
Sum squared resid	5.14E+08	Schwarz criterion	13.90861
Log likelihood	-56310.63	Hannan-Quinn criter.	13.90179
Durbin-Watson stat	1.993357		

Heteroskedasticity Test: White

F-statistic	22.35826	Prob. F(53,8051)	0.0000
Obs*R-squared	1039.881	Prob. Chi-Square(53)	0.0000
Scaled explained SS	11816.74	Prob. Chi-Square(53)	0.0000

White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	-0.156932	0.043114	-3.639957	0.0003
D24_DEMEAN_NET_LOAD(-1)	-0.133546	0.040034	-3.335842	0.0009
D24_TEMP_V*D2	19.30455	5.636429	3.424961	0.0006
D24_TEMP_V(-1)*D2	-17.96129	5.698225	-3.152085	0.0016
D24_TEMP_V(-1)*D1	5.008207	1.374023	3.644923	0.0003
D24_NET_LOAD(-1)	1.761574	0.046113	38.20146	0.0000
D24_NET_LOAD(-2)	-0.919797	0.042351	-21.71838	0.0000
D24_NET_LOAD(-3)	0.240234	0.024058	9.985736	0.0000
D24_NET_LOAD(-5)	0.044787	0.012509	3.580371	0.0003
D24_NET_LOAD(-6)	-0.020762	0.009328	-2.225884	0.0260

R-squared	0.968470	Mean dependent var	2.673242
Adjusted R-squared	0.968435	S.D. dependent var	1418.710
S.E. of regression	252.0559	Akaike info criterion	13.89841
Sum squared resid	5.14E+08	Schwarz criterion	13.90705
Log likelihood	-56313.31	Hannan-Quinn criter.	13.90137
Durbin-Watson stat	2.000854		

Wald Test:
Equation: D24_WHITE_SEASON_BEF_FIN

Test Statistic	Value	df	Probability
F-statistic	2887.470	(8, 8094)	0.0000
Chi-square	23099.76	8	0.0000

Null Hypothesis: C(1)=0,C(2)=0,C(3)=0,C(4)=0,C(5)=0,C(6)
=0,C(7)=0,C(8)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	-0.157018	0.043106
C(2)	-0.133616	0.040027
C(3)	19.31576	5.636267
C(4)	-17.97342	5.698234
C(5)	5.012959	1.374057
C(6)	1.760863	0.046213
C(7)	-0.915376	0.044419
C(8)	0.228777	0.033739

Restrictions are linear in coefficients.

36. *A3*– **Wald test for the dummies of the seventh specification of the model with exact-day matching**

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	0.932630	(4, 16448)	0.4437
Chi-square	3.730522	4	0.4437

Null Hypothesis: C(6)=0, C(7)=0, C(9)=0, C(10)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(6)	-3.037145	3.736342
C(7)	-4.633860	3.295732
C(9)	7.321488	8.922574
C(10)	8.698754	8.974031

Restrictions are linear in coefficients.

37. *A3*– **The eighth specification, LM-test of the model with exact-day matching**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	393.8356	27.69775	14.21905	0.0000
MEAN_NET_LOAD	0.981938	0.001534	639.9296	0.0000
DEMEAN_NET_LOAD(-1)	0.914302	0.003669	249.2263	0.0000
TEMP_V*D1	7.593717	1.306562	5.811983	0.0000
TEMP_V*D2	2.783018	0.644673	4.316945	0.0000
MONDAY	32.44513	11.00001	2.949555	0.0032
FRIDAY	-45.73352	10.62051	-4.306153	0.0000
SATURDAY	-271.4916	13.49651	-20.11569	0.0000
SUNDAY	-262.0624	15.54966	-16.85325	0.0000

R-squared	0.985413	Mean dependent var	18100.76
Adjusted R-squared	0.985399	S.D. dependent var	2598.780
S.E. of regression	314.0278	Akaike info criterion	14.33795
Sum squared resid	8.01E+08	Schwarz criterion	14.34570
Log likelihood	-58303.43	Hannan-Quinn criter.	14.34060
F-statistic	68609.11	Durbin-Watson stat	1.153065
Prob(F-statistic)	0.000000		

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1149.373	Prob. F(2,8123)	0.0000
Obs*R-squared	1794.133	Prob. Chi-Square(2)	0.0000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	57.73433	24.63739	2.343362	0.0191
MEAN_NET_LOAD	-0.001157	0.001356	-0.852969	0.3937
DEMEAN_NET_LOAD(-1)	-0.034974	0.003681	-9.502071	0.0000
TEMP_V*D1	1.282403	1.156593	1.108776	0.2676

TEMP_V*D2	0.127672	0.569309	0.224258	0.8226
MONDAY	-19.02611	9.785603	-1.944296	0.0519
FRIDAY	-8.412660	9.384567	-0.896436	0.3700
SATURDAY	-81.41253	12.57782	-6.472706	0.0000
SUNDAY	-110.0965	14.77847	-7.449788	0.0000
RESID(-1)	0.526896	0.010997	47.91181	0.0000
RESID(-2)	-0.160183	0.011575	-13.83904	0.0000

R-squared	0.220572	Mean dependent var	1.41E-11
Adjusted R-squared	0.219613	S.D. dependent var	313.8734
S.E. of regression	277.2743	Akaike info criterion	14.08924
Sum squared resid	6.25E+08	Schwarz criterion	14.09871
Log likelihood	-57289.95	Hannan-Quinn criter.	14.09248
F-statistic	229.8746	Durbin-Watson stat	2.045263
Prob(F-statistic)	0.000000		

38. *A3*– **ACF and PACF of the eighth specification of the model with exact-day matching**

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.423	0.423	1458.7	0.000
		2	0.016	-0.199	1460.7	0.000
		3	-0.207	-0.166	1809.6	0.000
		4	-0.233	-0.082	2252.9	0.000
		5	-0.170	-0.068	2488.8	0.000
		6	-0.015	0.043	2490.7	0.000
		7	0.097	0.034	2568.1	0.000
		8	0.142	0.043	2733.3	0.000
		9	0.078	-0.017	2782.4	0.000
		10	0.041	0.051	2796.3	0.000
		11	-0.009	0.004	2797.0	0.000
		12	-0.044	-0.007	2813.1	0.000
		13	-0.086	-0.050	2873.0	0.000
		14	-0.070	-0.017	2912.9	0.000
		15	-0.036	-0.016	2923.2	0.000
		16	0.052	0.053	2944.8	0.000
		17	0.056	-0.024	2970.7	0.000
		18	-0.004	-0.063	2970.8	0.000
		19	-0.066	-0.044	3006.8	0.000
		20	-0.050	0.019	3027.0	0.000
		21	-0.023	0.002	3031.3	0.000
		22	0.045	0.048	3047.8	0.000
		23	0.124	0.087	3173.4	0.000
		24	0.316	0.291	3988.5	0.000
		25	0.111	-0.146	4089.5	0.000
		26	-0.026	0.055	4095.1	0.000
		27	-0.092	0.036	4164.3	0.000
		28	-0.070	0.031	4204.1	0.000
		29	-0.050	-0.003	4224.9	0.000
		30	-0.004	-0.025	4225.0	0.000
		31	0.019	-0.031	4228.0	0.000
		32	0.056	0.003	4253.6	0.000
		33	0.057	0.027	4280.2	0.000
		34	0.077	0.041	4328.4	0.000
		35	0.067	0.048	4365.4	0.000
		36	0.046	0.046	4382.4	0.000

39. *A3*– **The ninth specification of the model with exact-day matching**

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 11.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	726.1602	55.63285	13.05272	0.0000
MEAN_NET_LOAD	0.671669	0.023810	28.20917	0.0000
DEMEAN_NET_LOAD(-1)	0.624460	0.025955	24.05898	0.0000
TEMP_V*D1	6.022807	1.411466	4.267057	0.0000
TEMP_V*D2	1.846939	0.545978	3.382809	0.0007
TUESDAY	-354.0742	74.43751	-4.756664	0.0000
THURSDAY	-250.7364	71.69008	-3.497504	0.0005
FRIDAY	-412.3337	77.80744	-5.299413	0.0000
SATURDAY	-1306.597	92.36943	-14.14534	0.0000
SUNDAY	-1532.115	101.1931	-15.14051	0.0000
NET_LOAD(-1)	0.515648	0.034066	15.13690	0.0000
NET_LOAD(-2)	-0.333561	0.015947	-20.91638	0.0000
NET_LOAD(-3)	0.109771	0.006044	18.16106	0.0000
MONDAY*NET_LOAD(-23)	0.143262	0.011992	11.94631	0.0000
FRIDAY*NET_LOAD(-23)	0.034715	0.009605	3.614372	0.0003
TUESDAY*NET_LOAD(-23)	0.036430	0.008721	4.177443	0.0000
THURSDAY*NET_LOAD(-23)	0.043196	0.011138	3.878448	0.0001
SUNDAY*NET_LOAD(-23)	-0.090895	0.029705	-3.059901	0.0022
FRIDAY*NET_LOAD(-24)	0.176352	0.019886	8.867964	0.0000
TUESDAY*NET_LOAD(-24)	0.179995	0.017754	10.13827	0.0000
THURSDAY*NET_LOAD(-24)	0.217774	0.022803	9.550228	0.0000
SUNDAY*NET_LOAD(-24)	0.268926	0.049636	5.417984	0.0000
MONDAY*NET_LOAD(-25)	-0.141891	0.011851	-11.97337	0.0000
TUESDAY*NET_LOAD(-25)	-0.198432	0.014768	-13.43642	0.0000
THURSDAY*NET_LOAD(-25)	-0.247482	0.018196	-13.60112	0.0000
FRIDAY*NET_LOAD(-25)	-0.191127	0.017108	-11.17157	0.0000
SATURDAY*NET_LOAD(-25)	0.056854	0.004989	11.39561	0.0000
SUNDAY*NET_LOAD(-25)	-0.103108	0.029139	-3.538561	0.0004

R-squared	0.990383	Mean dependent var	18106.13
Adjusted R-squared	0.990350	S.D. dependent var	2598.253
S.E. of regression	255.2320	Akaike info criterion	13.92567
Sum squared resid	5.26E+08	Schwarz criterion	13.94984
Log likelihood	-56440.59	Hannan-Quinn criter.	13.93394
F-statistic	30824.73	Durbin-Watson stat	1.553554
Prob(F-statistic)	0.000000		

40. *A3*– **The LM and Wald test of the second correct specification of the model with exact-day matching**

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3.824160	Prob. F(1,8075)	0.0506
Obs*R-squared	3.838649	Prob. Chi-Square(1)	0.0501

Wald Test:
Equation: FINAL

Test Statistic	Value	df	Probability
F-statistic	1225429.	(34, 8076)	0.0000
Chi-square	41664596	34	0.0000

Null Hypothesis: C(1)=0,C(2)=0,C(3)=0,C(4)=0,
C(5)=0,C(6)=0,C(7)=0,C(8)=0,C(9)=0,
C(10)=0,C(11)=0,C(12)=0,C(13)=0,
C(14)=0,C(15)=0,C(16)=0,C(17)=0,
C(18)=0,C(19)=0,C(20)=0,C(21)=0,
C(22)=0,C(23)=0,C(24)=0,C(25)=0,
C(26)=0,C(27)=0,C(28)=0,C(29)=0,C(30)=0,C(31
)=0,C(32)=0,C(33)=0,C(34)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	703.0654	57.05434
C(2)	0.750043	0.038011
C(3)	0.686545	0.038017
C(4)	7.344490	1.160118
C(5)	2.648539	0.793201
C(6)	-122.9244	56.20736
C(7)	-167.6213	75.87193
C(8)	-298.1386	68.20254
C(9)	-994.1958	71.98901
C(10)	-1477.538	83.28664
C(11)	0.374283	0.071318
C(12)	-0.207570	0.052683
C(13)	0.048703	0.018068
C(14)	0.107999	0.014797
C(15)	0.043292	0.011181
C(16)	0.041150	0.009729
C(17)	0.101347	0.008699
C(18)	-0.030992	0.014310
C(19)	-0.075839	0.026390
C(20)	0.095013	0.021327
C(21)	0.144127	0.026870
C(22)	-0.107287	0.014681
C(23)	-0.035280	0.008591
C(24)	-0.092116	0.008063
C(25)	-0.124653	0.018337
C(26)	0.069163	0.014699
C(27)	0.311662	0.058145
C(28)	-0.134514	0.015729
C(29)	-0.111078	0.014940
C(30)	-0.065439	0.011578
C(31)	0.052466	0.011291
C(32)	-0.048861	0.018960
C(33)	0.375970	0.024028
C(34)	0.090434	0.028577

Restrictions are linear in coefficients.

41. *A3*- **The LM and Wald test of the third correct specification of the model with exact-day matching**

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.683945	Prob. F(3,8089)	0.1681
Obs*R-squared	5.058120	Prob. Chi-Square(3)	0.1676

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D24_MEAN_NET_LOAD	-0.012991	0.008319	-1.561616	0.1184
D24_NET_LOAD(-1)	-0.888154	0.554133	-1.602781	0.1090
D24_NET_LOAD(-2)	1.391113	0.908323	1.531517	0.1257
D24_NET_LOAD(-3)	-0.643135	0.442121	-1.454659	0.1458
D24_NET_LOAD(-5)	0.121761	0.077228	1.576650	0.1149
D24_TEMP_V*D2	2.309507	1.519074	1.520339	0.1285
D24_TEMP_V(-1)*D1	2.632082	1.832359	1.436444	0.1509
MA(24)	0.001938	0.007189	0.269573	0.7875
MA(23)	-0.001112	0.006883	-0.161573	0.8716
MA(25)	0.000407	0.007234	0.056296	0.9551
MA(2)	-0.003171	0.008936	-0.354866	0.7227
MA(4)	0.007296	0.008379	0.870688	0.3840
MA(5)	-0.005820	0.007713	-0.754587	0.4505
MA(6)	-0.003415	0.007496	-0.455629	0.6487
RESID(-1)	0.888494	0.553952	1.603918	0.1088
RESID(-2)	0.051141	0.046268	1.105333	0.2690
RESID(-3)	-0.077200	0.081571	-0.946420	0.3440

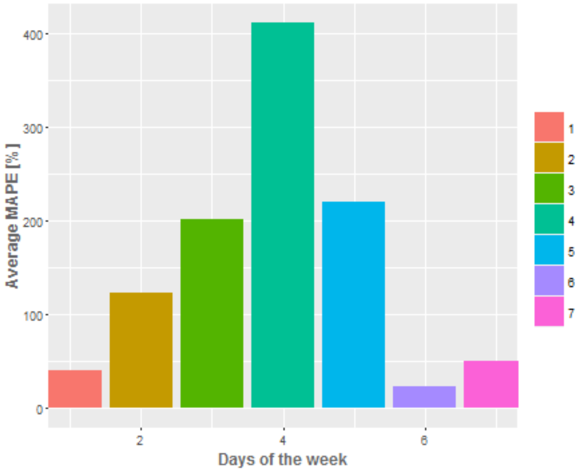
R-squared	0.000624	Mean dependent var	-0.084641
Adjusted R-squared	-0.001353	S.D. dependent var	222.5386
S.E. of regression	222.6890	Akaike info criterion	13.65152
Sum squared resid	4.01E+08	Schwarz criterion	13.66621
Log likelihood	-55312.63	Hannan-Quinn criter.	13.65655
Durbin-Watson stat	2.002033		

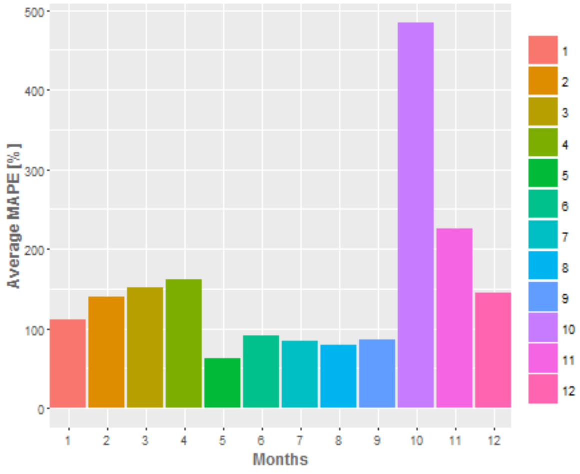
Wald Test:
Equation: D24_6LAG_MA

Test Statistic	Value	df	Probability
F-statistic	22463.18	(14, 8092)	0.0000
Chi-square	314484.5	14	0.0000

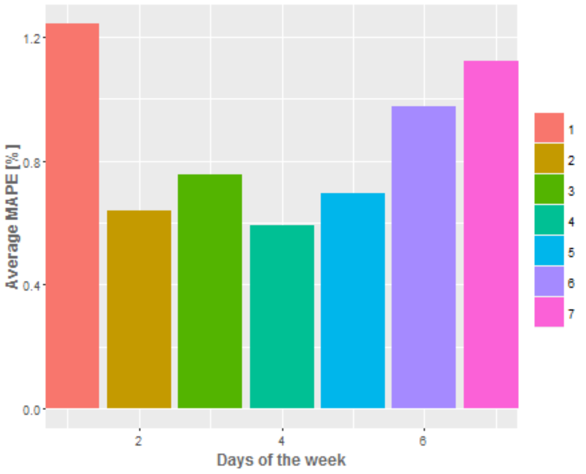
Normalized Restriction (= 0)	Value	Std. Err.
C(1)	-0.015685	0.005669
C(2)	1.619028	0.022281
C(3)	-0.902104	0.035351
C(4)	0.213492	0.021124
C(5)	0.048101	0.006677
C(6)	2.734653	0.912262
C(7)	3.326240	0.882131
C(8)	-0.761517	0.013836
C(9)	-0.077169	0.009841
C(10)	-0.047631	0.008433
C(11)	-0.037431	0.008023
C(12)	0.034521	0.006689
C(13)	-0.021465	0.006975
C(14)	0.027168	0.006258

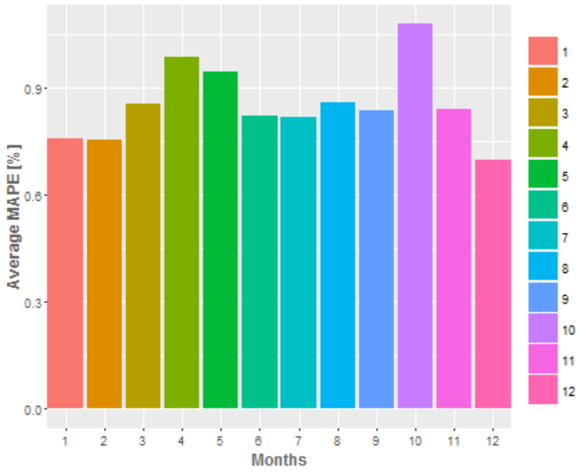
42. *A3*– **Daily and Monthly average MAPE of the one-step-ahead forecast of seasonally differenced model with exact-day matching for 2015**





43. *A3*– **Daily and Monthly average MAPE of the one-step-ahead forecast of the model with exact-day matching and dummies for 2015**





44. *A3* – Replication of the EGRV model (weekday and weekend versions)

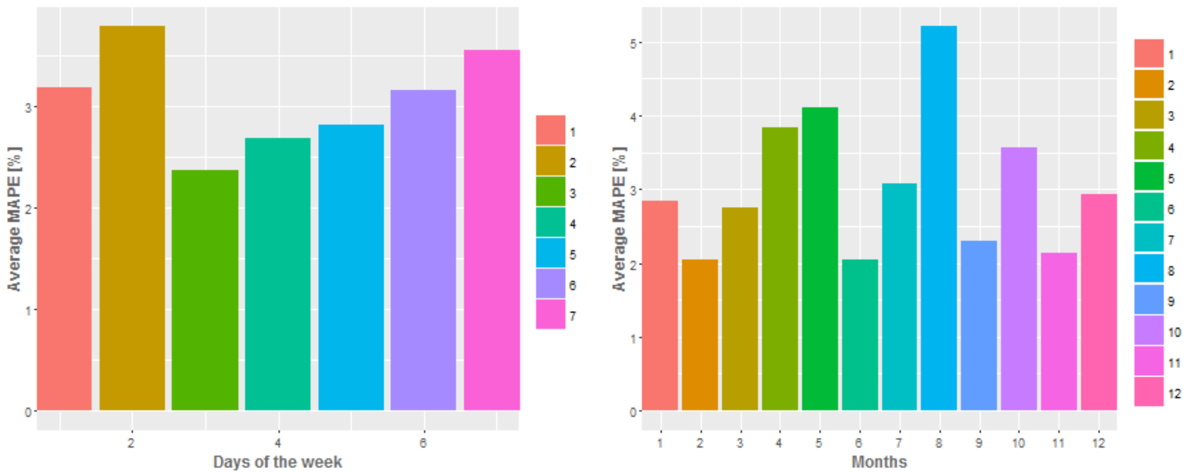
Homoscedastic weekday models	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	421 532,50	388 028,10	359 967,30	371 030,60	323 982,40	304 007,60	211 048,70	340 698,80	585 078,50	223 196,60	511 075,20	569 116,90	646 248,10	1 094 588,00	477 645,20	728 759,50	943 867,20	1 014 898,00	776 649,10	545 929,20	522 914,90	600 379,90	531 500,00	513 834,20
MONDAY	-	-	-	-	-	-	-	-17 997,89	-16 185,37	-	-16 169,98	-19 732,70	-19 243,04	-15 613,95	-17 254,78	-12 407,06	-17 150,28	-20 535,89	-	-16 083,65	-10 633,53	-14 694,54	-11 226,41	-
APRIL	20 900,68	-	-	-	-	-	-880,33	-	-	-	-	-	-	-	-824,34	-	-	-1 710,16	-2 088,65	-28 166,17	-	-	-	-
AUGUST	8 188,53	8 828,59	-447,76	-	-422,41	-	-	-1 134,47	-572,29	-	-26 768,96	-44 140,55	-38 730,30	-	-29 986,62	-590,95	-	-35 384,77	-1 725,08	-	-1 305,58	-805,91	-	-573,75
DAYAFTERHOLIDAY	-1 418,29	-1 591,32	-1 360,80	-1 300,47	-	-	-	19 734,56	17 582,45	8 545,86	17 118,04	20 707,10	19 857,19	16 730,17	19 901,49	14 755,52	19 758,80	18 889,45	-	14 985,46	10 843,54	16 889,14	13 904,31	2 991,99
FEBRUARY	-10 042,45	-11 202,39	-12 091,48	-12 127,53	-10 653,26	-9 381,15	-	-16 231,49	-40 725,01	-14 418,34	-39 863,74	-41 558,73	-41 549,66	-59 128,88	-16 392,67	-38 954,08	-60 830,87	-54 809,11	-54 919,93	-28 855,53	-27 686,63	-24 172,90	-23 278,01	-21 056,93
FRIDAY	-	-	-	51,24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-305,56	-350,09	-	-	-
JANUARY	-221,74	-366,19	-4 507,91	-7 329,25	-312,08	-381,71	-	-23 630,94	-38 175,00	-	-29 576,49	-44 545,15	-42 559,55	-64 898,88	-27 484,07	-40 647,66	-75 557,71	-86 895,99	-52 756,98	-25 921,80	-24 505,80	-19 756,33	-14 925,55	-15 009,75
JULY	-	-	-13 050,76	-11 980,05	-13 231,61	-18 925,68	-17 055,03	-28 416,70	-	-	-	-41 226,95	-28 508,02	-	-22 885,06	-	-	-	-	-33 652,31	-30 799,67	-	916,60	588,20
JUNE	-	-	283,74	337,84	-	-754,28	-929,65	-	-	-	274,37	594,90	358,02	-	-	-	-	-397,68	-975,00	-2 055,25	-1 105,49	-	643,29	481,35
MA(1)	-	-	-0,12	-0,12	-0,27	-	-	-0,25	-0,27	-0,38	-0,21	0,34	-	-	-	-	-	0,26	-	0,72	0,59	0,65	0,58	-
MA(2)	-0,38	-	-0,37	-0,34	-0,34	-0,32	-0,18	-0,28	-0,28	-0,37	-0,32	-	-0,36	-0,35	-0,45	-	-0,32	-0,34	-0,32	0,19	-	-	-	-0,35
MA(3)	-0,27	-	-	-	-	-0,18	-	-	-	-	-	-	-0,13	-	-	-	-	-0,41	-0,21	-	-	-	-	-0,19
MA(4)	-0,35	-	-0,34	-0,32	-0,23	-0,28	-0,08	-0,18	-0,21	-0,23	-0,21	-	-0,23	-0,27	-0,22	-	-	-0,32	-0,24	-	-	-	-	-0,24
MA(5)	-	0,22	-0,16	-0,22	-0,13	-0,23	-0,15	-0,28	-0,24	-	-0,24	-	-0,26	-0,38	-0,31	-0,14	-0,21	-0,19	-0,23	-	-	-	-	-0,22
MARCH	8 201,89	-	-	128,30	-	-	-	-	208,72	-	-	-	-	502,98	-	-	-	-899,72	-	-	-	-	-	-
MAX_TEMP_F(-1)	-554,95	-670,47	-998,50	-1 135,05	-365,51	-914,91	-1 360,23	-2 197,89	-1 633,79	-	-1 404,52	-1 477,69	-1 291,99	-2 191,86	-	-1 503,29	-1 329,92	-1 920,58	-	-	-	-	-	-532,12
MAX_TEMP_F_S(-1)	0,93	1,13	1,70	1,93	0,61	1,55	2,30	3,75	2,79	-0,06	2,43	2,56	2,24	3,73	-	2,55	2,24	3,27	-	-	-	-0,04	-	0,89
NET_LOAD_8(-1)	0,27	0,21	0,20	0,20	0,28	0,24	0,41	0,37	0,44	0,89	0,43	0,23	0,33	0,30	0,60	0,40	0,70	-	0,23	-	-	0,16	0,26	0,33
NET_LOAD_8(-1)*DAYAFTE	-	-	-	-	-0,06	-0,06	-	-1,51	-1,29	-0,44	-1,19	-1,43	-1,36	-1,19	-1,25	-1,01	-1,26	-1,36	-	-1,03	-0,73	-1,13	-0,90	-0,15
NET_LOAD_8(-1)*MONDAY	-0,02	-	-	-	-	-	-0,03	1,40	1,21	-	1,13	1,36	1,33	1,12	1,11	0,89	1,13	1,43	-	1,07	0,71	1,03	0,77	-
NOVEMBER	-6 086,54	-	-11 498,75	-11 670,25	-9 966,89	-9 970,41	-	-15 182,26	-20 244,52	-	774,05	-	-11 950,74	-33 174,46	-	-16 883,16	-31 677,86	-52 108,57	-36 841,92	-	-19 961,91	-17 396,80	-	-11 843,96
OCTOBER	9 419,68	-	-	-	197,25	-	-	-21 365,67	-18 962,70	-	432,28	-	393,51	684,58	-	-	-	-	-	-	-19 625,23	-	-	-
SEPTEMBER	15 099,16	13 206,82	155,71	313,04	105,59	-	-	458,62	608,27	-	569,08	672,42	464,67	448,63	-	-	-	-	-648,77	-	690,66	-	-	-
SEVENDAYMIDTEMP_F	-14,13	-20,36	-20,17	-12,71	-12,86	-	-	-	-	56,37	-	-	26,84	99,77	71,36	64,33	87,40	-	-	-	-49,36	-	-51,13	-
TEMP_F*APRIL	-73,55	-	-	-	-0,86	-1,97	-	-1,71	-	-1,24	-1,35	-	-2,05	-	-	-1,61	-1,36	-	-	92,15	-	-	-	-
TEMP_F*AUGUST	-31,11	-32,97	-	-1,02	-	-3,75	-5,68	-	-	-	90,05	148,67	130,03	-2,25	100,01	-	-	114,76	-	-9,75	-	-	-	-
TEMP_F*FEBRUARY	36,88	40,59	44,09	44,58	38,77	34,32	-	59,01	148,10	52,39	143,91	148,35	148,75	212,80	59,24	139,49	216,05	195,20	199,14	104,77	100,37	88,29	86,02	77,63
TEMP_F*JANUARY	-	-	15,18	25,78	-	-	-	85,64	138,42	-	106,75	159,16	152,12	231,98	98,69	144,92	269,93	313,07	190,14	93,29	87,94	71,30	54,71	54,60
TEMP_F*JULY	-	-	46,42	43,27	46,70	63,10	55,20	96,46	1,03	1,19	2,81	142,02	98,57	1,02	78,09	0,98	-	-	-2,13	107,51	101,97	-	-	-
TEMP_F*MARCH	-29,13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,10	1,75	1,69
TEMP_F*MAY	-32,47	-	-	-	-0,83	-3,20	-3,39	-0,91	-	-	-	-	-	-	-	-	-	-3,50	-5,11	-9,18	-4,26	-	-	-
TEMP_F*NOVEMBER	22,51	-	42,75	43,65	37,08	37,40	2,07	56,55	75,51	1,57	-	-	45,29	121,06	2,35	62,29	116,15	191,71	135,68	3,00	73,78	64,60	2,68	44,02
TEMP_F*OCTOBER	-33,29	0,44	1,18	1,64	-	0,35	-	76,67	68,48	-	-	1,79	-	-	-	-	-	-	1,50	-	70,34	-	-	-
TEMP_F	-2 311,77	-1 936,86	-1 387,76	-1 341,70	-1 811,93	-1 103,13	-	-	-2 324,04	-1 623,37	-2 027,90	-2 278,07	-3 082,68	-5 404,22	-3 352,30	-3 521,39	-5 265,50	-4 877,15	-5 249,28	-3 639,48	-3 400,90	-4 084,85	-3 546,00	-2 981,25
TEMP_F_S	4,13	3,45	2,45	2,36	3,20	1,92	-	-0,08	4,02	2,84	3,46	3,81	5,23	9,27	5,74	6,04	9,05	8,34	9,05	6,31	5,92	7,18	6,24	5,25
TUESDAY	-	-	-	-	81,25	-	257,50	271,61	281,06	607,44	295,38	-	288,51	307,27	524,78	390,45	432,85	-	394,48	-	-	399,64	350,24	382,49
WEDNESDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	210,31	182,05	202,31
R-squared	0,97	0,95	0,958	0,95	0,94	0,9	0,79	0,75	0,72	0,68	0,71	0,625	0,69	0,63	0,65	0,61	0,73	0,85	0,87	0,86	0,8	0,72	0,78	0,76
AIC	13,65	13,78	13,53	13,67	13,94	14,75	16,08	16,36	16,20	16,11	16,05	16,23	16,13	16,29	16,2	16,4	16,31	16,09	11,99	15,93	15,67	15,42	15,3	15,31
DW	2,15	1,73	1,97	1,9985	1,98	2,18	1,97	1,95	1,96	1,92	1,94	1,996	2,00	1,96	1,989	1,9	1,914	1,91	2,09	1,96	1,85	2	2,01	1,91
LM-test	0,083	0,024	0,03	0,014	0,80	0,034	0,61	0,01	0,06	0,27	0,25	0,9905	1,00	0,12	0,95	0,33	0,3055	0,06	0,075	0,36	0,075	0,734	0,71	0,06
RESET (p-value of F-stat)	0,00	square root	0,09	0,04	0,00	0,03	0,17	0,00	0,00	0,00	0,00	square root of negative	0,00	0,00	0,00	square root of negative	0,72	0,00	0,00	0,00	0,01	0,04	0,15	0,00

Homoscedastic weekend models	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	13 224,75	296 359,40	113 730,10	160 801,60	266 026,40	278 513,10	154 145,70	33 490,25	146 295,60	-	430 539,00	41 704,75	41 046,50	20 014,99	19 590,88	19 732,38	20 860,17	20 598,36	31 880,88	34 970,48	177 781,70	384 436,00	391 874,60	259 561,70
APRIL	-681,15	-482,80	-	-	-665,63	-	-47 668,90	-	-	-42 872,41	-	-	-	-	-	-2 531,52	-	-	-	-	-	-1 273,87	-	20 226,42
AUGUST	-	-1 286,82	-621,81	-492,99	-	-	-2 446,15	-	-956,66	-988,86	-	-24 748,29	-840,99	-32 625,99	-44 401,72	-45 313,27	-39 226,15	-38 624,21	-41 470,93	-25 452,39	-1 323,68	-2 484,52	-2 204,72	-2 366,13
DAYAFTERHOLIDAY	-	-	-	-	-	-960,76	56 444,22	-32 007,47	-767,23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-10 231,50
FEBRUARY	-	-	178,29	-	-	-	-	-	-	21 847,60	-	-	-	27 218,22	25 786,18	28 380,98	36 705,60	45 656,86	-	-	-	-	-639,84	-517,35
JANUARY	-	-	-	-	-309,37	-	-	11 662,52	-	20 552,83	-35 923,47	-	-	15 467,04	13 461,02	10 602,38	-	-	-	13 804,36	-	-1 005,28	-17 554,63	-958,62
JULY	-336,87	-475,43	-52 072,17	-41 845,06	-512,38	-	-2 465,85	-	-	-	-	-	-	-61 073,38	-	-1 423,05	-37 913,00	-43 892,77	-44 256,62	-41 619,54	-	-1 469,29	-804,81	-985,26
JUNE	-643,41	-544,51	-	-	-753,75	-614,94	-2 521,89	-	-347,68	-	-	-	-	-1 930,09	-1 920,70	-2 166,55	-3 394,25	-3 457,91	-2 672,69	-1 997,90	-1 710,17	-1 925,37	-1 167,53	-1 434,66
MA(1)	-	-	-0,63	-	-	0,55	1,91	-	-0,66	1,08	0,33	0,44	0,46	-0,49	-0,46	-0,50	-0,55	-	-	-	0,34	-	-	-
MA(2)	-	-	-0,34	-	-	0,75	0,92	-	-0,34	0,95	-	0,28	0,30	-0,48	-0,52	-0,47	-0,42	-	-	-	-	-0,98	-	-
MARCH	25 945,04	19 888,13	-198,03	-	-442,81	-	-474,48	-	-	-	-	-782,95	-895,90	-1 910,67	-1 963,88	-2 336,61	-3 436,33	-3 086,80	-1 101,98	-	-298,68	-	-	-
MAX_TEMP_F(-1)	-	-	-679,21	-1 011,00	-	-	-837,23	-	-889,83	-	-	-	-	-	-	-	-	-	-41,40	-	-	-789,89	-	-
MAX_TEMP_F_S(-1)	-	-	1,12	1,70	-	-	1,40	-	1,48	-0,08	-	-	-	-	-	-	-	-	-	-	-	1,35	-	-
MAY	-862,74	-641,66	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-1 544,58	-1 493,14	-1 713,65
NET_LOAD_8(-1)	0,26	0,22	0,24	0,24	0,12	-	-0,88	0,19	0,33	-0,51	-	-	-	-	-	-	-	-	-	0,21	-	-0,11	-	-
NET_LOAD_8(-1)*DAYAFTERHOLIDAY	-0,10	-0,09	-0,08	-	-0,08	-	-3,79	2,05	-	-0,04	-	-	-	-0,08	-0,06	-0,06	-0,07	-	-	-	-	-	-	0,65
NOVEMBER	-	-	212,10	-	-	-16 996,98	-	-	321,78	-	-	-	-	47 925,10	47 416,70	47 419,12	47 114,85	-	-	-	-	15 847,79	-	-
OCTOBER	-	-291,13	-	-	-313,08	-	44 115,67	-	-	-	-	-	-	-	-1 405,53	-	-	16 993,70	-889,03	-	-	-1 182,86	-1 200,53	-1 115,11
SEPTEMBER	-	18 994,14	-	-	-442,91	-	-1 619,26	-	-	-	-	-	-	-1 622,62	-1 608,74	-1 935,86	-3 206,14	-3 251,94	-2 270,16	-	-	-1 632,95	-1 568,16	-1 740,10
SEVENDAYMIDTEMP_F	-	-	-	-	-	-56,98	-	-69,01	-	-54,60	-	-79,68	-77,21	-	-	-	-	-	-	-66,25	-51,67	-	-	-
SUNDAY	-1 886,18	-1 595,64	-1 479,63	-1 431,86	-1 286,37	-1 206,31	-	-2 057,22	-2 029,78	-329,88	-936,71	-910,87	-1 018,24	-1 078,01	-916,08	-874,47	-816,52	-595,31	-439,30	-648,47	-150,60	-	-190,47	-282,25
TEMP_F	-	-2 004,98	-	-	-1 785,28	-1 749,51	-	-	-	323,76	-2 798,54	-	-	-	-	-	-	-	-	-	-1 009,45	-1 773,08	-2 630,09	-1 705,96
TEMP_F*APRIL	-	-	-1,20	-	-	-	170,40	-1,26	-1,69	149,92	-	-	-	-7,47	-7,68	-	-13,05	-12,80	-10,72	-6,43	-1,11	-	-4,12	-75,76
TEMP_F*AUGUST	-4,70	-	-	-	-4,35	-3,85	-	-3,11	-	-	-2,88	80,02	-	101,11	141,08	143,67	119,07	117,04	129,90	79,54	-	-	-	-
TEMP_F*FEBRUARY	-	-	-	-	-	-	-	-	-	-76,53	-	-	-	-98,44	-93,25	-103,55	-137,16	-166,95	-	-	-	-2,05	-	-
TEMP_F*JANUARY	-	-	-	-	-	-	-	-43,01	-	-73,87	128,45	-	-	-57,28	-50,09	-40,57	-4,83	-	-	-50,25	-1,03	-	59,60	-
TEMP_F*JULY	-	-	179,55	144,94	-	-	-	-	-	-	-	-	-	198,59	-4,04	-	117,52	137,32	142,14	136,19	-3,63	-	-	-
TEMP_F*MARCH	-95,06	-72,79	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-4,24	-4,56	-4,72
TEMP_F*MAY	-	-	-0,90	-	-2,87	-3,45	-7,39	-	-1,28	-	-	-3,15	-3,25	-6,73	-6,53	-7,80	-12,30	-12,73	-10,68	-7,75	-4,63	-	-	-
TEMP_F*NOVEMBER	-	-	-	-	-	62,36	-2,14	-	-	-	-	-	-	-171,42	-170,09	-170,92	-171,45	-	-	-	-	-59,63	-2,74	-2,42
TEMP_F*OCTOBER	-1,99	-	-	-	-	-	-159,81	-	-	4,79	-	-	-	-4,98	-	-6,10	-10,48	-68,42	-	-	-	-	-	-
TEMP_F*SEPTEMBER	-2,70	-67,92	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TEMP_F_S	-	3,54	-	-	3,15	3,09	-	-	-	-0,54	4,75	-	-	-	-	-	-	-	-	-	1,77	3,16	4,63	3,00
R-squared	0,97	0,95	0,96	0,95	0,94	0,90	0,79	0,75	0,72	0,68	0,71	0,62	0,69	0,63	0,64	0,61	0,73	0,85	0,87	0,86	0,80	0,72	0,78	0,76
AIC	13,65	13,78	13,53	13,67	13,94	14,75	16,08	16,36	16,20	16,11	16,05	16,23	16,13	16,30	16,20	16,40	16,30	16,10	15,99	15,93	15,67	15,40	15,30	15,30
DW	2,15	1,73	1,97	2,00	1,98	2,18	1,97	1,95	1,96	1,92	1,94	2,00	2,00	1,96	1,99	1,92	1,91	1,90	2,10	1,96	1,85	2,00	2,01	1,91
LM-test	0,08	0,02	0,03	0,01	0,80	0,03	0,61	0,01	0,06	0,27	0,25	1,00	1,00	0,12	0,95	0,33	0,31	0,01	0,08	0,36	0,08	0,73	0,71	0,06
RESET (p-value of F-stat)	0,91	0,32	0,01	1,00	0,44	0,04	0,00	0,30	0,00	0,00	square root of	0,24	0,23	0,00	0,00	0,00	0,00	0,78	0,98	0,08	0,90	0,00	0,53	0,15

P-values of the weekday model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MONDAY	-	-	-	-	-	-	-	0,031	0,043	-	0,035	0,001	0,009	0,037	0,022	0,046	0,018	0,007	-	0,001	0,005	0,000	0,003	-
APRIL	0,001	-	-	-	-	-	0,000	-	-	-	-	-	-	-	0,000	-	-	0,000	0,000	0,028	-	-	-	-
AUGUST	0,080	0,014	0,000	-	0,000	-	-	0,000	0,003	-	0,032	0,037	0,003	-	0,004	0,044	-	0,004	0,000	-	0,012	0,014	-	0,000
DAYAFTERHOLIDAY	0,000	0,000	0,000	0,000	-	-	-	0,028	0,046	0,000	0,024	0,000	0,004	0,031	0,004	0,007	0,003	0,010	-	0,000	0,000	0,000	0,000	0,036
FEBRUARY	0,003	0,000	0,000	0,000	0,000	0,019	-	0,035	0,000	0,001	0,000	0,001	0,000	0,000	0,020	0,001	0,000	0,000	0,000	0,001	0,002	0,001	0,001	0,001
FRIDAY	-	-	-	0,054	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	-	-	-
JANUARY	0,000	0,000	0,056	0,005	0,000	0,000	-	0,024	0,001	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,006	0,013	0,014	0,018	0,005
JULY	-	-	0,001	0,017	0,001	0,000	0,001	0,000	-	-	-	0,009	0,092	-	0,012	-	-	-	-	0,017	0,016	-	0,000	0,000
JUNE	-	-	0,000	0,000	-	0,000	0,000	-	-	-	0,052	0,036	0,066	-	-	-	-	0,005	0,000	0,000	0,000	-	0,001	0,000
MA(1)	-	-	0,134	0,191	0,002	-	-	0,043	0,013	0,000	0,043	0,001	-	-	-	-	-	0,026	-	0,000	0,000	0,000	0,000	-
MA(2)	0,000	-	0,000	0,000	0,000	0,000	0,045	0,000	0,000	0,000	0,000	-	0,000	0,000	0,000	-	0,000	0,004	0,000	0,022	-	-	-	0,000
MA(3)	0,000	-	-	-	-	0,002	-	-	-	-	-	-	0,034	-	-	-	-	0,000	0,000	-	-	-	-	0,000
MA(4)	0,000	-	0,000	0,000	0,000	0,000	0,080	0,001	0,000	0,000	0,001	-	0,001	0,000	0,002	-	-	0,000	0,000	-	-	-	-	0,000
MA(5)	-	0,004	0,025	0,005	0,039	0,001	0,040	0,011	0,045	-	0,005	-	0,000	0,004	0,000	0,066	0,029	0,010	0,001	-	-	-	-	0,003
MARCH	0,040	-	-	0,000	-	-	-	-	0,055	-	-	-	-	0,001	-	-	-	0,000	-	-	-	-	-	-
MAX_TEMP_F(-1)	0,026	0,000	0,000	0,000	0,015	0,001	0,000	0,002	0,014	-	0,056	0,011	0,088	0,002	-	0,021	0,025	0,002	-	-	-	-	-	0,040
MAX_TEMP_F_S(-1)	0,030	0,000	0,000	0,000	0,017	0,001	0,000	0,003	0,016	0,022	0,054	0,010	0,084	0,002	-	0,022	0,027	0,002	-	-	-	0,006	-	0,048
NET_LOAD_8(-1)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,027	0,011	0,000	0,000	0,033	0,004	0,017	0,000	0,006	0,000	-	0,010	-	-	0,062	0,012	0,000
NET_LOAD_8(-1)*DAYAFTERHOI	-	-	-	-	0,000	0,000	-	0,016	0,029	0,000	0,025	0,000	0,005	0,030	0,010	0,016	0,008	0,010	-	0,001	0,001	0,000	0,000	0,039
NET_LOAD_8(-1)*MONDAY	0,077	-	-	-	-	-	0,000	0,018	0,029	-	0,033	0,001	0,009	0,034	0,028	0,051	0,024	0,008	-	0,001	0,004	0,000	0,002	-
NOVEMBER	0,040	-	0,000	0,000	0,000	0,001	-	0,062	0,042	-	0,000	-	0,081	0,000	-	0,018	0,000	0,000	0,000	-	0,043	0,022	-	0,002
OCTOBER	0,001	-	-	-	0,000	-	-	0,016	0,015	-	0,000	-	0,001	0,000	-	-	-	-	-	-	0,001	-	-	-
SEPTEMBER	0,000	0,000	0,000	0,000	0,002	-	-	0,000	0,000	-	0,000	0,005	0,017	0,009	-	-	-	-	0,000	-	0,010	-	-	-
SEVENDAYMIDTEMP_F	0,064	0,009	0,001	0,047	0,030	-	-	-	-	0,000	-	-	0,266	0,002	0,000	0,000	0,000	-	-	-	0,005	-	0,001	-
TEMP_F*APRIL	0,001	-	-	-	0,000	0,000	-	0,002	-	0,001	0,008	-	0,001	-	-	0,026	0,006	-	-	0,037	-	-	-	-
TEMP_F*AUGUST	0,053	0,008	-	0,000	-	0,000	0,000	-	-	-	0,033	0,036	0,003	0,000	0,004	-	-	0,006	-	0,000	-	-	-	-
TEMP_F*FEBRUARY	0,003	0,000	0,000	0,000	0,000	0,017	-	0,035	0,000	0,001	0,000	0,001	0,000	0,000	0,019	0,001	0,000	0,000	0,000	0,001	0,002	0,001	0,001	0,001
TEMP_F*JANUARY	-	-	0,077	0,007	-	-	-	0,024	0,001	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,006	0,013	0,014	0,017	0,005
TEMP_F*JULY	-	-	0,001	0,013	0,001	0,000	0,002	0,000	0,003	0,000	0,000	0,007	0,083	0,086	0,011	0,103	-	-	0,004	0,025	0,020	-	-	-
TEMP_F*MARCH	0,041	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,018	0,000	0,000
TEMP_F*MAY	0,001	-	-	-	0,000	0,000	0,000	0,044	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,002	-	-	-
TEMP_F*NOVEMBER	0,035	-	0,000	0,000	0,000	0,000	0,000	0,054	0,035	0,000	-	-	0,061	0,000	0,000	0,013	0,000	0,000	0,000	0,010	0,037	0,017	0,000	0,001
TEMP_F*OCTOBER	0,001	0,039	0,000	0,000	-	0,002	-	0,015	0,013	-	-	0,065	-	-	-	-	-	-	0,000	-	0,001	-	-	-
TEMP_F	0,000	0,000	0,000	0,000	0,000	0,017	-	-	0,013	0,000	0,028	0,037	0,012	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-
TEMP_F_S	0,000	0,000	0,000	0,000	0,000	0,018	-	0,082	0,013	0,000	0,031	0,044	0,013	0,000	0,000	0,001	0,000	0,000	0,000	-	-	-	-	0,000
TUESDAY	-	-	-	-	0,018	-	0,001	0,020	0,005	0,000	0,001	-	0,005	0,003	0,000	0,002	0,000	-	0,000	-	-	0,001	0,001	0,000
WEDNESDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,036	0,039	0,016

P-values of the weekend models	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	0,000	0,000	0,000	0,000	0,000	0,000	0,005	0,000	0,000	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000
APRIL	0,000	0,000	-	-	0,000	-	0,000	-	-	0,066	-	-	-	-	-	0,000	-	-	-	-	-	0,000	-	0,000
AUGUST	-	0,000	0,000	0,000	-	-	0,000	-	0,000	0,000	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
DAYAFTERHOLIDAY	-	-	-	-	-	0,000	0,000	0,000	0,002	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,010
FEBRUARY	-	-	0,001	-	-	-	-	-	-	0,012	-	-	-	0,000	0,000	0,000	0,000	0,000	-	-	-	-	0,000	0,012
JANUARY	-	-	-	-	0,000	-	-	0,000	-	0,003	0,000	-	-	0,000	0,000	0,000	-	-	-	0,000	-	0,000	0,023	0,000
JULY	0,036	0,027	0,000	0,000	0,014	-	0,000	-	-	-	-	-	-	0,000	-	0,000	0,001	0,000	0,000	0,000	-	0,000	0,009	0,002
JUNE	0,000	0,001	-	-	0,000	0,000	0,000	-	0,001	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MA(1)	-	-	0,000	-	-	0,000	0,000	-	0,002	0,000	0,010	0,000	0,001	0,000	0,000	0,000	0,000	-	-	-	0,000	-	-	-
MA(2)	-	-	0,007	-	-	0,000	0,000	-	0,041	0,000	-	0,017	0,002	0,000	0,000	0,000	0,000	-	-	-	-	0,000	-	-
MARCH	0,000	0,000	0,000	-	0,000	-	0,025	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,003	-	0,032	-	-	-
MAX_TEMP_F(-1)	-	-	0,000	0,000	-	-	0,025	-	0,000	-	-	-	-	-	-	-	-	-	0,001	-	-	0,007	-	-
MAX_TEMP_F_S(-1)	-	-	0,000	0,000	-	-	0,029	-	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	0,008	-	-
MAY	0,000	0,001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000
NET_LOAD_8(-1)	0,001	0,001	0,000	0,000	0,007	-	0,000	0,005	0,001	0,000	-	-	-	-	-	-	-	-	-	0,000	-	0,000	-	-
NET_LOAD_8(-1)*DAYAFTERHOI	0,001	0,000	0,000	-	0,000	-	0,000	0,000	-	0,000	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	0,016
NOVEMBER	-	-	0,000	-	-	0,000	-	-	0,000	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	0,013	-	-
OCTOBER	-	0,029	-	-	0,005	-	0,012	-	-	-	-	-	-	-	0,000	-	-	0,039	0,018	-	-	0,000	0,000	0,000
SEPTEMBER	-	0,002	-	-	0,007	-	0,000	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	-	-	0,000	0,000	0,000
SEVENDAYMIDTEMP_F	-	-	-	-	-	0,000	-	0,000	-	0,010	-	0,001	0,001	-	-	-	-	-	-	0,001	0,011	-	-	-
SUNDAY	0,000	0,000	0,000	0,000	0,000	0,000	-	0,000	0,000	0,016	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,005	-	0,003	0,000
TEMP_F	-	0,000	-	-	0,000	0,000	-	-	-	0,000	0,000	-	-	-	-	-	-	-	-	-	0,019	0,000	0,000	0,000
TEMP_F*APRIL	-	-	0,000	-	-	-	0,000	0,011	0,000	0,064	-	-	-	0,000	0,000	-	0,000	0,000	0,000	0,000	0,022	-	0,000	0,000
TEMP_F*AUGUST	0,000	-	-	-	0,000	0,000	-	0,000	-	-	0,000	0,000	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-
TEMP_F*FEBRUARY	-	-	-	-	-	-	-	-	-	0,014	-	-	-	0,000	0,000	0,000	0,000	0,000	-	-	-	0,000	-	-
TEMP_F*JANUARY	-	-	-	-	-	-	-	0,000	-	0,003	0,000	-	-	0,000	0,000	0,000	0,000	-	-	0,000	0,019	-	0,032	-
TEMP_F*JULY	-	-	0,000	0,000	-	-	-	-	-	-	-	-	-	0,000	0,000	-	0,001	0,000	0,000	0,000	0,000	-	-	-
TEMP_F*MARCH	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000
TEMP_F*MAY	-	-	0,000	-	0,000	0,000	0,000	-	0,000	-	-	0,005	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-
TEMP_F*NOVEMBER	-	-	-	-	-	0,000	0,028	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	0,010	0,001	0,006
TEMP_F*OCTOBER	0,000	-	-	-	-	-	0,010	-	-	0,001	-	-	-	0,000	-	0,000	0,000	0,016	-	-	-	-	-	-
TEMP_F*SEPTEMBER	0,000	0,002	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TEMP_F_S	-	0,000	-	-	0,000	0,000	-	-	-	0,000	0,000	-	-	-	-	-	-	-	-	-	0,023	0,000	0,000	0,000

45. *A3*- **Daily and Monthly average MAPE of the day-ahead forecast of the replicated EGRV model for 2015**



46. *A3*- **Daily and Monthly average MAPE of the replicated model for 2015**

2015.01.05-2015.12.20	**Naive benchmark with seasonal differencing**	**"SARMA"**	**Exponential smoothing**	**Exact-day matching with daily seasonal differencing**	**Exact-day matching with dummy**	**EGRV**
Type of forecast	*one-step ahead*	*one-step ahead*	*one-step ahead*	*one-step ahead*	*one-step ahead*	*one-day ahead*
Monday	45%	466%	1,14%	41%	1,24%	3,18%
Tuesday	98%	157%	0,91%	123%	0,64%	3,78%
Wednesday	137%	203%	0,63%	201%	0,76%	2,37%
Thursday	248%	91%	0,63%	411%	0,59%	2,68%
Friday	156%	110%	0,75%	221%	0,70%	2,81%
Saturday	22%	132%	0,99%	23%	0,98%	3,16%
Sunday	40%	79%	1,08%	51%	1,12%	3,54%
January	91%	127%	1,07%	111%	0,76%	2,85%
February	99%	143%	0,76%	140%	0,75%	2,04%
March	111%	93%	0,79%	151%	0,85%	2,75%
April	128%	261%	1,21%	161%	0,99%	3,83%
May	66%	133%	1,03%	62%	0,95%	4,11%
June	73%	136%	0,68%	91%	0,82%	2,04%
July	58%	76%	0,62%	85%	0,82%	3,08%
August	57%	66%	0,79%	79%	0,86%	5,21%
Spetember	73%	616%	1,12%	86%	0,84%	2,30%
October	306%	213%	0,99%	484%	1,08%	3,56%
November	109%	131%	0,74%	226%	0,84%	2,13%
December	101%	77%	0,61%	144%	0,70%	2,92%
MEAN of MAPE	**107%**	**173%**	**0,87%**	**151,67%**	**0,85%**	**3,07%**
Chow forecast test - winter time	0,67	0,64	NA	1,00	0,00	NA
Chow forecast test - summer time	0,00	0,00	NA	0,00	0,00	NA
Heteroscedasticity	yes	yes	?	yes	yes	no

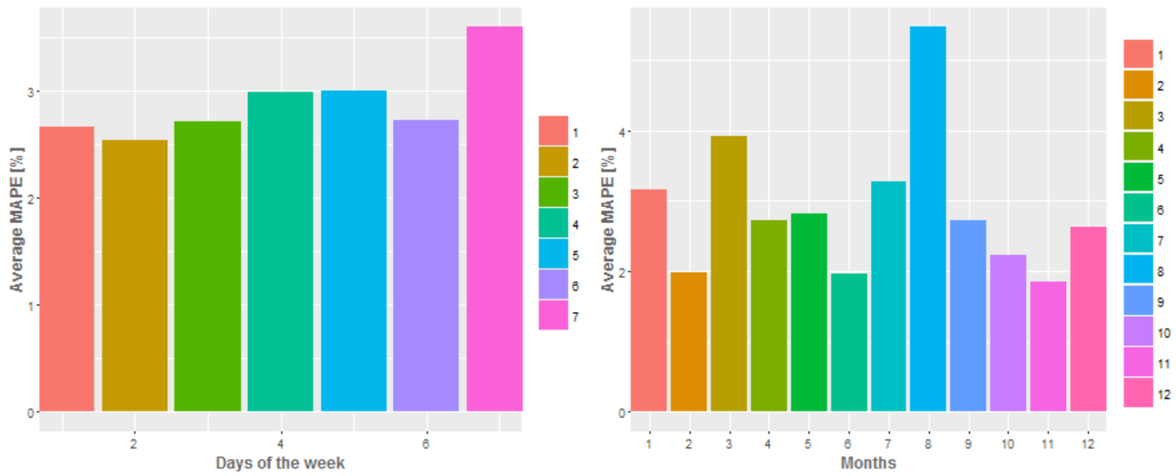
A4-Appendix for My own model

1. *A4*– **The first specification of my own model**

P-values of the heteroscedastic models	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000
_7DAYS_MIDTEMP_F	0,000	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000	0,001	0,001	0,006	0,007	0,012	0,000	0,000	-	-	0,000	0,000	0,001	0,000	0,000	0,000
AFTER_HOLIDAY	0,000	0,000	0,000	0,000	0,037	0,003	0,001	0,000	0,004	0,000	0,000	0,005	-	-	-	-	-	-	-	-	-	-	-	-
APRIL	-	-	-	-	-	-	0,004	-	-	-	-	-	-	-	0,013	0,009	-	-	-	-	-	-	-	-
APRIL*HUMIDITY	-	-	-	0,036	-	-	0,014	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	-	-	-	-
APRIL*TEMP_F	-	-	-	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,011	0,008	-	-	-	-	-	-	-	-
AUGUST	-	-	-	-	-	-	-	0,000	0,001	0,002	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	0,000	-	-	-	-
AUGUST*TEMP_F	-	-	-	-	-	0,000	0,002	-	0,004	0,004	0,000	0,000	-	-	-	-	-	-	-	0,000	-	-	-	-
BEFORE_HOLIDAY	0,038	0,275	0,020	-	-	-	-	-	0,048	0,048	0,016	0,034	0,011	0,015	0,005	0,003	0,003	0,002	0,001	0,001	0,000	0,002	0,000	0,000
DECEMBER	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,007	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*ILLUM	-	-	-	-	-	-	-	-	-	0,015	0,004	0,017	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*TEMP_F	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,009	-	-	-	-	-	-	-	0,003	-	-	-	-
FEBRUARY	0,004	0,012	0,049	0,012	0,000	0,000	0,000	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	0,000	0,000	0,000
FEBRUARY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,003	0,001	-	-	-	-	-	-	-	-
FEBRUARY*HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,029	0,004	-	-	-	-	-	-	-	-
FEBRUARY*TEMP_F	0,004	0,011	0,047	0,012	0,000	0,000	0,000	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	0,000	0,000	0,000
FRIDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000
HOLIDAY	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ILLUM	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY	-	0,011	0,029	-	0,032	0,005	0,001	-	-	0,034	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,016	-	0,000	0,000	0,000
JANUARY*HUMIDITY	-	-	-	0,000	-	0,044	0,043	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY*TEMP_F	-	-	-	-	0,039	0,017	0,002	-	0,012	0,037	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,020	-	0,000	0,000	0,000
JULY	0,021	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,000	-	-	-	0,000	0,002	0,000	0,000	0,000
JULY*HUMIDITY	-	-	-	-	-	-	0,032	0,042	0,005	0,001	0,000	0,000	0,001	0,001	0,000	0,001	-	-	-	0,002	-	0,008	-	0,011
JULY*TEMP_F	0,018	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,000	-	-	-	0,000	0,003	0,000	0,000	0,000
JUNE	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	0,000	-	-	-	-
JUNE*HUMIDITY	-	-	-	0,000	-	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,291	-	0,102
JUNE*TEMP_F	0,005	0,005	0,023	0,000	-	-	0,008	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	0,000	-	0,212	0,000	0,010
MA(1)	0,000	-	0,000	0,002	0,000	0,000	-	-	-	-	-	-	-	-	-	-	0,003	0,000	0,000	0,013	0,000	0,000	0,000	0,028
MA(2)	-	0,013	-	0,020	0,000	0,053	0,000	0,000	0,000	0,000	0,000	0,011	-	-	-	-	0,000	0,001	0,000	-	0,000	-	-	-
MA(3)	0,008	0,085	0,000	-	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,001	0,000	0,000	0,000	0,000	0,003	0,000	0,000	-	0,000	-	-	-
MA(4)	-	-	-	0,000	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	0,000	0,000	0,000	0,000	0,000
MA(5)	0,002	0,000	0,000	0,000	0,000	0,000	0,020	0,001	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	0,000	-	0,000	0,000	0,000
MA(6)	-	-	-	0,071	-	-	0,000	-	-	-	-	0,000	-	-	-	-	0,002	0,011	-	0,000	-	0,000	0,000	0,000
MA(7)	-	-	-	-	-	0,004	-	-	-	-	-	-	-	-	-	-	-	-	-	0,005	-	0,001	0,058	0,000
MARCH	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MARCH*HUMIDITY	-	-	-	0,000	0,000	0,000	0,000	-	0,003	0,003	0,007	0,004	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*ILLUM	-	-	-	-	-	-	-	-	-	-	-	0,019	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*TEMP_F	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	0,001	-	-	-	-
MAX_HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,035	0,007	-	-	-	-	-	-	-	-
MAX_HUMIDITY(-1)	-	-	-	-	-	-	0,010	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MAX_ILLUM	-	-	-	0,003	0,001	0,001	0,000	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MAX_TEMP_F	-	0,003	0,009	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	0,025	0,000	0,000	0,000
MAX_TEMP_F_S	-	0,004	0,011	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	0,026	0,000	0,000	0,000
MAY	-	-	-	-	-	-	-	0,002	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	0,000	-	-	-	-
MAY*HUMIDITY	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-	0,009	-	0,000	0,000	0,000
MAY*TEMP_F	-	-	-	-	-	-	-	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	0,000	0,000	-	-	-	-
MEAN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,001	0,001	-	-	-	-	-	0,003	-	0,011
MEAN_ILLUM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-
MIN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,029
MONDAY	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NAT_HOLIDAY	0,000	0,000	0,000	0,000	-	0,054	0,033	-	0,011	0,011	0,002	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,018
NET_LOAD(-1)	0,000	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
NET_LOAD(-2)	0,000	0,000	0,002	0,006	0,000	0,000	-	-	-	0,000	0,000	0,000	-	-	-	-	0,007	0,000	-	-	0,001	0,000	0,000	0,000
NET_LOAD(-4)	0,001	0,421	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-	0,006	0,000	0,000	0,005	0,000	0,000	0,000	0,000
NET_LOAD(-5)	0,004	0,001	0,016	-	0,005	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	-	-	0,000	-	-	-
NET_LOAD(-3)	-	-	-	-	0,000	0,000	-	-	-	-	-	-	0,010	0,002	-	-	0,005	0,017	-	-	0,045	-	-	-
NET_LOAD(-7)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	-	-	-
NET_LOAD(-6)	-	-	-	-	-	0,013	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NOVEMBER	-	-	-	-	-	0,006	0,004	-	-	-	-	-	0,000	0,000	0,000	0,000	0,002	0,001	0,001	0,000	0,005	0,000	0,000	0,001
NOVEMBER*TEMP_F	0,000	0,000	0,000	0,000	0,000	0,005	0,003	0,000	0,000	0,000	0,000	0,000	-	-	-	-	0,001	0,001	0,000	0,000	0,004	0,000	0,000	0,001
OCTOBER	-	-	-	-	-	0,000	0,000	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	0,000	-	-	-	-
OCTOBER*HUMIDITY	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,006	0,000	0,000	0,000
OCTOBER*TEMP_F	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	0,018	0,000	0,000	-	-	-	-
SATURDAY	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SUMMER_HOL	0,000	0,000	0,000	0,000	0,000	0,001	0,000	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,035	-	0,000	0,000	0,000
SUNDAY	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SUNDAY_AFTER_WORKSAT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,002	0,016	-	-	-	-
TEMP_F	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	0,000	0,000	0,000	0,006	0,004	0,000	0,000	0,000
TEMP_F_S	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	0,000	0,000	0,000	0,009	0,004	0,000	0,000	0,000
THURSDAY	-	0,223	-	-	-	-	-	-	-	-	-	-	-	-	-	0,027	-	-	-	-	-	0,002	0,038	0,000
TUESDAY	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
WORKING_SAT	-	-	-	-	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	-	0,145	0,001	0,000

Heteroscedastic models	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	278 049,00	285 703,00	258 375,90	225 622,90	212 742,00	311 928,40	405 115,70	189 404,70	177 162,80	251 161,50	294 476,30	315 678,60	548 692,20	550 306,70	582 935,10	613 496,50	691 688,50	663 523,10	588 114,70	240 106,20	269 351,50	480 252,90	480 950,00	417 759,90
_7DAYS_MIDTEMP_F	-43,99	-40,82	-32,13	-31,82	-12,72	-19,92	-52,90	-53,92	-42,14	-25,35	-20,31	-16,38	-19,97	-19,34	-32,12	-41,16	-	-	-79,72	-67,33	-45,46	-28,84	-32,10	-35,92
AFTER_HOLIDAY	-2 199,07	-2 292,64	-1 739,58	-1 527,27	-330,34	-397,33	-531,99	-939,74	-623,75	-725,58	-591,90	-467,03	-	-	-	-	-	-	-	-	-	-	-	-
APRIL	-	-	-	-	-	-	20 365,04	-	-	-	-	-	-	-	21 273,29	24 252,48	-	-	-	-	-	-	-	-
APRIL*HUMIDITY	-	-	-	4,81	-	-	-15,40	-	-	-	-	-	-	-	-	-	-	-	-12,47	-22,68	-	-	-	-
APRIL*TEMP_F	-	-	-	-2,54	-0,66	-1,21	-69,61	-1,71	-1,48	-1,47	-1,54	-1,83	-1,37	-1,42	-74,94	-85,39	-	-	-	-	-	-	-	-
AUGUST	-	-	-	-	-	-	-	-1 510,73	-12 343,89	-13 606,60	-17 911,90	-21 084,00	-800,90	-816,37	-919,04	-903,68	-	-	-	-32 463,58	-	-	-	-
AUGUST*TEMP_F	-	-	-	-	-	-1,58	-2,24	-	38,24	42,86	57,98	69,08	-	-	-	-	-	-	-	107,27	-	-	-	-
BEFORE_HOLIDAY	-134,98	-104,91	-152,66	-	-	-	-	-	-392,91	-371,27	-409,38	-379,90	-401,64	-385,43	-436,74	-499,13	-409,39	-492,05	-456,65	-435,46	-527,57	-540,82	-620,72	-710,40
DECEMBER	-	-	-	-	-	-	-	34 075,25	42 893,64	40 374,73	35 423,75	27 836,74	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*ILLUM	-	-	-	-	-	-	-	-	-	-19,95	-15,63	-18,44	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*TEMP_F	-	-	-	-	-	-	-	-122,54	-154,96	-144,29	-126,38	-98,22	-	-	-	-	-	-	-	-2,82	-	-	-	-
FEBRUARY	-10 154,91	-10 230,16	-6 307,03	-6 537,72	-8 509,09	-15 347,13	-18 550,56	-	-	-	-	-	-16 276,13	-18 795,66	-21 561,43	-24 907,20	-42 722,87	-37 903,94	-32 076,56	-24 922,85	-	-20 684,10	-24 238,47	-17 665,94
FEBRUARY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13,00	13,65	-	-	-	-	-	-	-	-
FEBRUARY*HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-7,93	-9,45	-	-	-	-	-	-	-	-
FEBRUARY*TEMP_F	37,10	37,28	23,11	23,66	30,86	55,52	66,63	-	-	-	-	-	58,17	67,16	75,39	87,32	152,09	135,95	116,17	87,17	-	74,67	88,12	64,28
FRIDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-241,56	-329,32	-406,61	-478,91	-589,38	-576,01	-425,51	-352,21
HOLIDAY	-2 913,99	-2 595,22	-2 531,67	-2 276,48	-1 327,73	-1 479,00	-2 968,73	-4 575,48	-3 223,13	-3 133,92	-2 889,79	-2 805,71	-2 376,08	-2 413,33	-2 452,11	-2 558,91	-1 965,21	-1 853,85	-1 557,04	-1 577,57	-1 446,13	-1 627,20	-1 637,89	-1 862,10
ILLUM	-	-	-	-	-	-	-	-6,66	-6,28	-6,03	-6,70	-8,03	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY	-	-184,81	-159,87	-	-4 037,29	-10 465,59	-16 723,46	-	-	-8 596,16	-14 415,07	-15 134,84	-27 058,56	-26 558,83	-25 408,32	-31 753,15	-54 618,80	-50 011,07	-33 571,40	-17 564,49	-	-20 620,79	-25 937,24	-17 599,78
JANUARY*HUMIDITY	-	-	-	-1,67	-	12,52	15,27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY*TEMP_F	-	-	-	-	14,14	32,77	54,28	-	-0,48	30,57	51,30	53,65	96,18	94,47	90,28	113,14	195,17	180,55	120,99	60,03	-	73,40	93,42	63,27
JULY	-10 842,41	-18 095,86	-17 292,46	-21 769,19	-16 914,87	-19 368,29	-33 927,35	-48 279,12	-54 003,17	-51 500,84	-53 734,27	-54 269,80	-25 158,59	-28 866,79	-30 606,42	-29 008,60	-	-	-	-70 696,93	-23 326,39	-32 050,62	-15 716,59	-19 134,17
JULY*HUMIDITY	-	-	-	-	-	-	10,65	10,04	12,91	15,18	16,31	18,77	15,24	14,58	20,92	17,78	-	-	-	16,39	-	9,61	-	6,04
JULY*TEMP_F	38,05	62,93	60,13	75,57	58,30	65,09	110,95	160,07	179,54	170,46	178,06	179,71	81,29	93,63	98,38	93,58	-	-	-	233,82	79,15	106,63	54,61	64,92
JUNE	-	-	-	-	-	-	-	-44 006,09	-50 179,04	-44 855,12	-39 996,91	-37 608,78	-	-	-	-	-	-	-	-43 179,04	-	-	-	-
JUNE*HUMIDITY	-	-	-	-12,20	-	-5,16	-22,64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-5,12	-	-7,60
JUNE*TEMP_F	0,79	0,71	0,63	3,59	-	-	3,88	149,62	170,67	152,00	135,64	127,78	-	-	-	-	-	-	-	144,64	-	1,43	1,26	3,06
MA(1)	0,27	-	0,36	0,19	-0,73	-0,36	-	-	-	-	-	-	-	-	-	-	0,25	0,35	0,40	0,19	0,52	0,30	0,29	0,17
MA(2)	-	0,16	-	-0,16	0,48	0,15	-0,29	-0,20	-0,20	-0,29	-0,23	-0,16	-	-	-	-	0,33	0,28	0,38	-	0,57	-	-	-
MA(3)	0,16	0,10	0,24	-	-0,43	-0,28	-0,24	-0,31	-0,32	-0,19	-0,19	-0,17	-0,26	-0,28	-0,25	-0,26	0,22	0,26	0,32	-	0,59	-	-	-
MA(4)	-	-	-	-0,34	-	-	-	-0,24	-0,29	-0,31	-0,34	-0,22	-0,36	-0,36	-0,40	-0,39	-	-	-	-0,36	0,31	-0,36	-0,33	-0,33
MA(5)	-0,19	-0,23	-0,25	-0,52	-0,30	-0,34	-0,19	-0,22	-0,16	-0,18	-0,22	-0,24	-0,35	-0,33	-0,32	-0,32	-	-	-	-0,38	-	-0,44	-0,48	-0,36
MA(6)	-	-	-	-0,14	-	-	-0,25	-	-	-	-	-0,21	-	-	-	-	0,16	0,16	-	-0,27	-	-0,29	-0,35	-0,23
MA(7)	-	-	-	-	-	-0,15	-	-	-	-	-	-	-	-	-	-	-	-	-	-0,14	-	-0,18	-0,10	-0,21
MARCH	-	-	-	-	-	-	-	-	-	-	-	-	17 166,79	18 821,30	19 878,07	19 781,68	-	-	-	-	-	-	-	-
MARCH*HUMIDITY	-	-	-	-3,22	-2,09	-2,93	-5,26	-	7,63	8,37	7,35	11,11	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*ILLUM	-	-	-	-	-	-	-	-	-	-	-	6,82	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*TEMP_F	-	-	-	-	-	-	-	-1,43	-3,29	-3,49	-2,94	-5,29	-60,85	-66,48	-70,16	-70,24	-	-	-	-1,47	-	-	-	-
MAX_HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-6,14	-7,68	-	-	-	-	-	-	-	-
MAX_HUMIDITY(-1)	-	-	-	-	-	-	8,52	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MAX_ILLUM	-	-	-	-1,28	-0,84	-1,43	-3,10	-	-	-	-	-	-5,92	-6,15	-6,37	-6,48	-	-	-	-	-	-	-	-
MAX_TEMP_F	-	-524,50	-443,32	-	-	-	-	-	-	-	-	-	-3 641,66	-3 659,03	-3 862,45	-4 051,87	-	-	-	-	-793,29	-1 375,48	-1 383,94	-1 294,38
MAX_TEMP_F_S	-	0,88	0,75	-	-	-	-	-	-	-	-	-	6,29	6,33	6,70	7,03	-	-	-	-	1,37	2,37	2,40	2,23
MAY	-	-	-	-	-	-	-	-13 972,63	-15 625,64	-13 090,63	-14 709,01	-15 747,71	-	-	-	-	-	-	-	-27 737,39	-	-	-	-
MAY*HUMIDITY	-	-	-	-2,93	-2,41	-6,95	-9,97	-	-	-	-	-	-	-	-	-	-	-	-	12,12	-	-1,64	-1,44	-3,01
MAY*TEMP_F	-	-	-	-	-	-	-	46,49	52,26	43,55	48,97	52,49	-1,31	-1,22	-1,21	-1,49	-	-	-2,76	86,84	-	-	-	-
MEAN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7,75	7,59	-	-	-	-	-	4,33	-	9,56
MEAN_ILLUM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-19,11	-16,21	-16,47	-10,83	-	-	-	-
MIN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-5,99
MONDAY	-1 869,69	-1 748,31	-1 345,31	-1 292,71	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NAT_HOLIDAY	2 092,58	1 738,33	1 627,08	1 408,28	-	-547,16	-878,91	-	-1 272,66	-989,01	-1 120,76	-1 202,10	-1 417,66	-1 532,21	-1 446,99	-1 484,99	-2 082,18	-2 083,82	-1 998,00	-1 806,65	-1 672,83	-1 119,34	-853,21	-624,49
NET_LOAD(-1)	0,13	-	0,14	0,14	0,86	0,57	0,16	0,10	0,10	-	-	-	0,12	0,11	0,09	0,08	0,13	0,15	0,13	0,12	0,15	0,12	0,14	0,13
NET_LOAD(-2)	0,08	0,13	0,08	0,06	-0,55	-0,26	-	-	-	0,12	0,12	0,11	-	-	-	-	0,06	0,09	-	-	0,08	0,07	0,10	0,08
NET_LOAD(-4)	0,07	0,02	0,07	0,10	-	-	-	-	-	-	-	-	-	-	-	-	0,05	0,09	0,07	0,06	0,12	0,07	0,09	0,06
NET_LOAD(-5)	0,06	0,09	0,05	-	-0,05	-	-	-	-	-	-	-	0,06	0,07	0,07	0,08	0,08	0,09	-	-	0,13	-	-	-
NET_LOAD(-3)	-	-	-	-	0,28	0,10	-	-	-	-	-	-	0,04	0,05	-	-	0,06	0,06	-	-	0,05	-	-	-
NET_LOAD(-7)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,06	-	-	-
NET_LOAD(-6)	-	-	-	-	-	-0,04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NOVEMBER	-	-	-	-	-	-7 969,43	-14 164,38	-	-	-	-	-	568,09	553,85	497,69	552,69	-18 015,23	-24 854,65	-25 910,78	-23 453,72	-14 474,51	-18 991,14	-18 195,63	-13 798,93
NOVEMBER*TEMP_F	1,15	1,19	1,20	0,76	0,68	29,55	52,22	1,67	1,59	1,51	1,41	1,66	-	-	-	-	66,82	92,48	96,51	83,66	53,85	69,23	66,56	50,77
OCTOBER	-	-	-	-	-	-6 570,23	-23 091,65	-	-	-	-	-	15 774,49	14 422,36	12 251,87	13 035,68	-	-	-	-24 446,77	-	-	-	-
OCTOBER*HUMIDITY	3,31	3,98	3,79	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4,31	1,65	1,30	2,00
OCTOBER*TEMP_F	-	-	-	0,62	0,39	23,70	82,78	1,11	0,83	0,99	0,89	0,97	-53,63	-48,92	-41,42	-44,47	-	1,46	2,53	86,81	-	-	-	-
SATURDAY	-246,28	-196,59	-301,15	-400,57	-682,06	-1 263,93	-2 569,68	-2 976,17	-2 453,33	-2 029,08	-1 830,10	-1 779,26	-1 945,67	-2 248,18	-2 402,35	-2 610,65	-2 698,21	-2 747,84	-2 319,07	-2 112,51	-2 195,78	-2 038,46	-1 943,80	-1 991,47
SUMMER_HOL	-511,73	-583,58	-449,30	-596,26	-427,50	-348,48	-828,08	-	-	-	-	-	-	-	-	-	-525,38	-505,34	-795,82	-624,29	-	-686,38	-580,02	-620,06
SUNDAY	-1 763,24	-1 585,69	-1 436,51	-1 486,74	-1 225,62	-1 734,87	-3 668,39	-4 326,49	-3 710,66	-3 246,47	-2 804,17	-2 678,12	-2 762,48	-3 124,58	-3 157,43	-3 329,41	-3 235,44	-3 083,60	-2 574,40	-2 280,34	-2 125,40	-2 098,49	-1 924,80	-2 023,99
SUNDAY_AFTER_WORKSAT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	528,78	469,04	-	-	-	-
TEMP_F	-1 799,14	-1 313,37	-1 235,93	-1 454,91	-1 421,11	-2 073,74	-2 625,94	-1 087,18	-1 008,45	-1 548,40	-1 847,66	-1 995,45	-	-	-	-	-4 686,39	-4 510,19	-3 816,24	-1 368,62	-940,83	-1 814,25	-1 807,07	-1 448,01
TEMP_F_S	3,18	2,34	2,20	2,57	2,50	3,63	4,59	1,89	1,73	2,65	3,16	3,40	-	-	-	-	8,09	7,79	6,63	2,30	1,67	3,19	3,15	2,53
THURSDAY	-	-63,76	-	-	-	-	-	-	-	-	-	-	-	-	-	109,37	-	-	-	-	-	-123,90	-96,43	-136,86
TUESDAY	-	-	-	-	-	-	-	-	-	364,12	329,37	286,22	-	-	-	-	-	-	-	-	-	-	-	-
WORKING_SAT	-	-	-	-	328,61	409,87	1 030,01	1 247,93	1 033,92	852,99	856,90	945,20	925,87	986,38	974,65	1 104,28	1 301,80	1 109,93	885,18	669,10	-	334,36	673,01	964,84
R-squared	0,96	0,95	0,96	0,96	0,95	0,97	0,98	0,97	0,97	0,97	0,96	0,96	0,96	0,96	0,96	0,97	0,966	0,97	0,98	0,98	0,96	0,96	0,96	0,97
AIC	13,8	13,88	13,5	13,3	13,88	13,88	14,52	14,9	14,63	14,50	14,50	14,49	14,55	14,58	14,57	14,57	14,67	14,7	14,56	14,34	14,3	13,90	13,90	13,60
DW	1,9	1,74	1,89	1,9	1,9	1,89	2,04	2,19	2,15	2,05	1,99	1,96	1,94	1,93	1,95	1,93	2	1,96	1,98	2	2,01	1,95	2,00	1,98
LM-test (p-value of F-stat)	0,1	0,015	0,14	0,17	0,05	0,02	0,7	0,07	0,17	0,67	0,93	0,33	0,66	0,57	0,68	0,55	0,68	0,36	0,83	0,98	0,43	0,25	0,75	0,57
RESET (p-value of F-stat)	square root o	square root o	square root o	0.2330	0.0000	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,44	square root of negative number	square root of negative number	0,35	0,00	0,00	0,00	0,00

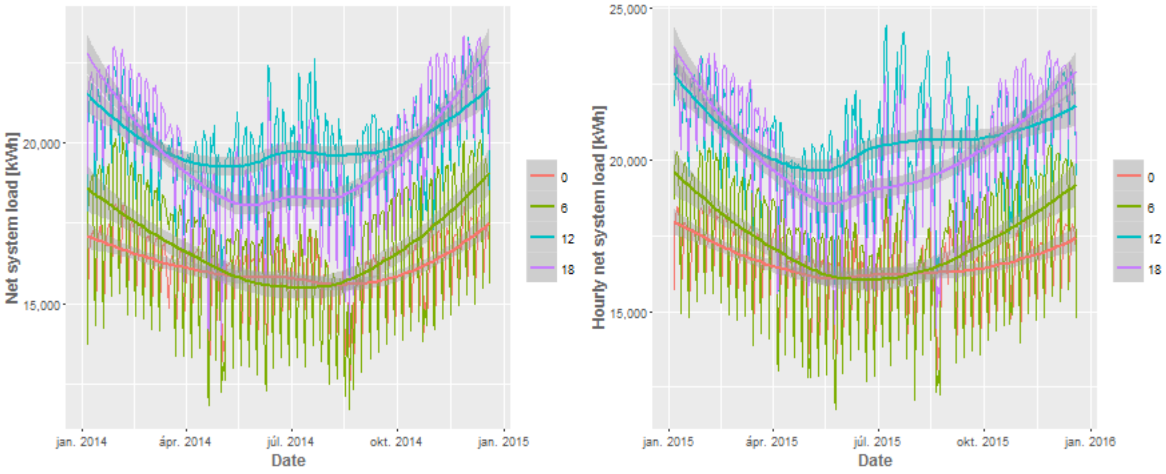
2. *A4*– **The daily and monthly MAPE of the one-day-ahead forecast of the first specification of my own model**



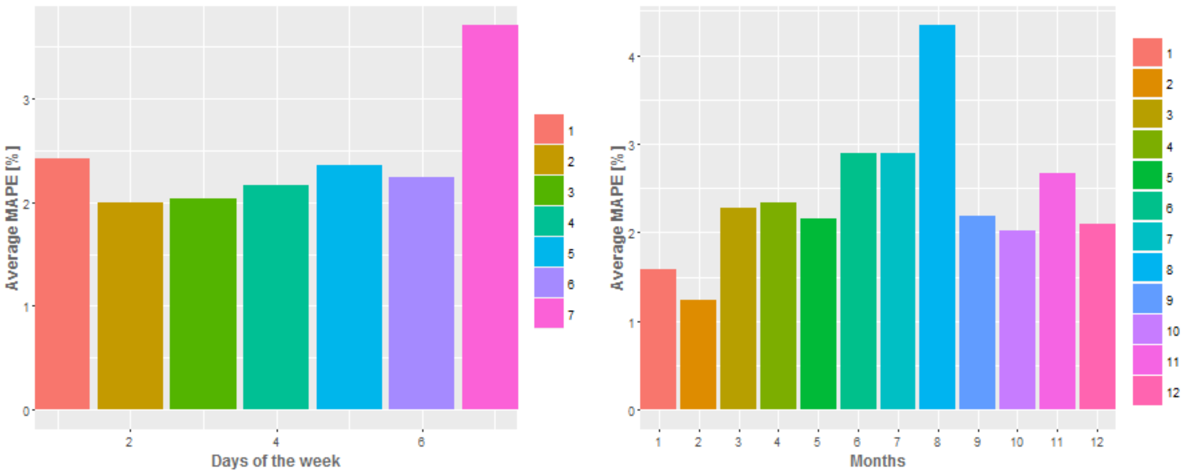
3. *A4*– **The result of the PE-test**

Hours	Linear difference in logarithmic equation	p-values	Log difference in linear equation	p-values
1	0,00	0,00	14 106,60	0,03
2	0,00	0,00	15 441,11	0,01
3	0,00	0,00	6 283,15	0,29
4	0,00	0,00	11 017,44	0,07
5	0,00	0,62	-5 082,38	0,21
6	0,00	0,80	-7 541,57	0,06
7	0,00	0,02	-25 259,73	0,00
8	0,00	0,05	-29 692,05	0,00
9	0,00	0,43	-31 753,47	0,00
10	0,00	0,92	-26 585,17	0,04
11	0,00	0,45	-16 900,98	0,22
12	0,00	0,15	-8 654,14	0,48
13	0,00	0,24	-19 055,75	0,05
14	0,00	0,52	-23 966,96	0,01
15	0,00	0,80	-18 172,16	0,03
16	0,00	0,67	-17 089,43	0,02
17	0,00	0,07	-8 112,05	0,23
18	0,00	0,01	890,86	0,87
19	0,00	0,00	2 529,75	0,63
20	0,00	0,02	-3 266,86	0,60
21	0,00	0,10	-21 484,57	0,00
22	0,00	0,27	-30 005,62	0,00
23	0,00	0,60	-22 269,42	0,00
24	0,00	0,04	-11 941,15	0,03

4. *A4*– **The evolution of the net load of the given hours in 2014 and 2015**



5. *A4*– **The daily and monthly MAPE of the one-day-ahead forecast of the logarithmic specification of my own model**



6. *A4*– The logarithmic specification of my own model

Heteroscedastic models	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	17,372	17,651	16,478	16,555	19,362	20,329	24,985	16,003	15,656	18,575	24,658	25,437	34,551	34,802	37,185	38,580	40,023	38,551	35,832	23,165	21,725	25,623	24,073	25,989
_7DAYS_MIDTEMP_F	-0,001	-0,001	-0,001	-0,001	-0,001	-0,002	-0,004	-0,002	-0,002	-0,002	-0,001	-0,001	-0,001	-0,001	-0,002	-0,002	-	-	-0,004	-0,004	-0,005	-0,003	-0,004	-0,004
AFTER_HOLIDAY	-0,043	-0,040	-0,033	-0,029	-0,025	-0,040	-0,032	-0,040	-0,035	-0,040	-0,035	-0,030	-	-	-	-	-	-	-	-	-	-	-	-
APRIL	-	-	-	-	-	0,906	1,531	-	-	-	-	-	-	-	1,164	1,335	-	-	-	-	-	-	-	-
APRIL*HUMIDITY	-	-	-	-	-	-	-0,001	-	-	-	-	-	-	-	-	-	-	-	-0,001	0,000	-	-	-	-
APRIL*TEMP_F	-	-	0,000	0,000	0,000	-0,003	-0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,004	-0,005	-	-	-	-	-	-	-	-
AUGUST	-	-	-	-	-	-	-	-0,080	-1,011	-1,032	-0,988	-1,006	-0,043	-0,044	-0,050	-0,049	-	-	-	-	-	-	-	-
AUGUST*TEMP_F	-	-	-	-	-	0,000	0,000	-	0,003	0,003	0,003	0,003	-	-	-	-	-	-	-	-	-	-	-	-
BEFORE_HOLIDAY	-	-	-	-	-	-	-	-	-	-	-0,019	-0,020	-0,023	-0,022	-0,024	-0,028	-0,022	-0,028	-0,024	-0,024	-	-	-	-
DECEMBER	-	-	-	-	-	-	-	1,409	1,630	1,497	-	0,041	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*ILLUM	-	-	-	-	-	-	-	-	-	-0,001	-0,001	-0,001	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*TEMP_F	-	-	-	-	-	-	-	-0,005	-0,006	-0,005	0,000	-	-	-	-	-	-	-	-	-	-	-	-	-
FEBRUARY	-0,625	-0,619	-0,521	-0,476	-0,651	-0,536	-0,664	-	-	-	-	-	-0,798	-0,907	-1,029	-1,191	-2,056	-1,926	-1,614	-0,489	-	-	-	-
FEBRUARY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,001	0,001	-	-	-	-	-	-	-	-
FEBRUARY*HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	-	-	-	-	-	-	-	-
FEBRUARY*TEMP_F	0,002	0,002	0,002	0,002	0,002	0,002	0,002	-	-	-	-	-	0,003	0,003	0,004	0,004	0,007	0,007	0,006	0,002	-	0,000	-	0,000
FRIDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-0,013	-0,017	-0,020	-0,020	-	-	-	-
HOLIDAY	-0,082	-0,082	-0,083	-0,086	-0,101	-0,154	-0,194	-0,273	-0,184	-0,173	-0,159	-0,156	-0,127	-0,128	-0,130	-0,136	-0,104	-0,098	-0,085	-0,076	-0,146	-0,133	-0,125	-0,136
ILLUM	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY	-0,007	-0,009	-0,010	-0,009	-0,335	-	-	-	-	-0,011	-0,756	-0,762	-1,345	-1,349	-1,291	-1,599	-2,683	-2,523	-1,552	-0,014	-	-	-	-
JANUARY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY*TEMP_F	-	-	-	-	0,001	0,000	0,000	-	-	-	0,003	0,003	0,005	0,005	0,005	0,006	0,010	0,009	0,006	-	-	-	-	-
JULY	-	-	-0,672	-0,749	-1,194	-1,234	-1,829	-1,542	-1,710	-1,547	-1,268	-2,038	-1,082	-1,289	-1,516	-1,439	-	-	-	-	-	-	-0,723	-
JULY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	0,001	0,001	0,001	0,001	0,001	-	-	-	-	-	-	-	-
JULY*TEMP_F	-	-	0,002	0,003	0,004	0,004	0,006	0,005	0,006	0,005	0,004	0,007	0,003	0,004	0,005	0,005	-	-	-	-	-	-	0,003	-
JUNE	-	-	-	-	-	-	-	-1,780	-2,377	-2,245	-1,901	-1,917	-	-	-	-	-	-	-	-	-	-	-	-
JUNE*HUMIDITY	-	-	-	-	-	0,000	-0,001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JUNE*TEMP_F	-	-	-	-	-	-	0,000	0,006	0,008	0,008	0,006	0,007	-	-	-	-	-	-	-	-	-	-	-	-
MA(1)	-0,815	-0,805	-0,794	-0,754	-0,724	-	-	-	-	-	-	-	-	-	-	-	0,279	0,351	0,393	0,451	0,610	0,530	0,566	0,598
MA(2)	0,788	0,780	0,744	0,703	0,467	-	-0,214	-	-	-	-	-	-	-	-	-	0,283	0,218	0,345	0,497	0,590	0,351	0,368	0,464
MA(3)	-0,442	-0,473	-0,463	-0,456	-0,413	-	-0,174	-	-	-	-	-	-0,227	-0,232	-0,264	-0,276	0,205	0,232	0,324	0,453	0,615	0,327	0,345	0,481
MA(4)	-	-	-	-	-	-	-	-	-	-	-	-	-0,272	-0,280	-0,381	-0,373	-	-	-	0,222	0,415	-	-	0,212
MA(5)	-	-	-	-	-0,312	-	-0,265	-	-	-	-	-	-0,332	-0,331	-0,326	-0,325	-	-	-	-	0,247	-0,054	-0,101	-
MA(6)	-	-	-	-	-	-	-0,333	-	-	-	-	-	-0,138	-0,127	-	-	0,178	0,218	-	-	0,172	-	-	-
MA(7)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

MARCH	-	-	-	-	-	-	-	-	-	-	-	-	0,900	1,001	1,079	1,083	-	-	-	-	-	-	-	-
MARCH*HUMIDITY	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	0,000	0,000	0,000	0,001	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*ILLUM	-	-	-	-	-	-	-	-	-	-	-	0,000	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*TEMP_F	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	-0,003	-0,004	-0,004	-0,004	-	-	-	-	-	-	-	-
MAX_HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	-	-	-	-	-	-	-	-
MAX_HUMIDITY(-1)	-	-	-	-	-	-	0,001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MAX_ILLUM	-	-	-	-	0,000	0,000	0,000	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MAX_TEMP_F	-	-	-	-	-	-	-	-	-	-	-	-	-0,183	-0,185	-0,197	-0,205	-	-	-	-	-	-	-	-
MAX_TEMP_F_S	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MAY	-	-	-	-	-	-	-	-0,035	-0,482	-0,510	-0,458	-0,451	-	-	-	-	-	-	-	-	-	-	-	-
MAY*HUMIDITY	0,000	0,000	0,000	0,000	0,000	0,000	-0,001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000
MAY*TEMP_F	-	-	-	-	-	-	-	-	0,002	0,002	0,001	0,001	0,000	0,000	0,000	0,000	-	-	0,000	0,000	-	-	-	-
MEAN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	-	-	-	-	-	-	-	-
MEAN_ILLUM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-0,001	-0,001	-0,001	-0,001	-	-	-	-
MIN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MONDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NAT_HOLIDAY	-	-	-	-	-	-	-0,067	-	-0,073	-0,061	-0,068	-0,069	-0,088	-0,098	-0,094	-0,097	-0,129	-0,129	-0,121	-0,117	-	-	-	-
NET_LOAD(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-2)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-4)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-5)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-7)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-6)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NOVEMBER	-	-	-	-	-	-	-	-	-	-	-	-	0,029	0,029	0,027	0,029	-0,640	-0,943	-0,975	-	-	-	-	-
NOVEMBER*TEMP_F	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	0,002	0,004	0,004	0,000	0,000	0,000	0,000	0,000
OCTOBER	-	-	-	-	-	-	-0,982	-	-	-	-	-	0,903	0,835	0,724	0,772	-	-	-	-	-	-	-	-
OCTOBER*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OCTOBER*TEMP_F	-	0,000	0,000	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000	0,000	-0,003	-0,003	-0,002	-0,003	-	0,000	0,000	0,000	-	-	-	-
SATURDAY	-0,020	-0,022	-0,023	-0,028	-0,048	-0,079	-0,158	-0,167	-0,133	-0,106	-0,095	-0,090	-0,098	-0,114	-0,124	-0,135	-0,139	-0,141	-0,120	-0,112	-0,094	-0,090	-0,092	-0,102
SUMMER_HOL	-0,030	-0,030	-0,030	-0,031	-0,032	-0,025	-0,026	-	-	-	-	-	-	-	-	-	-0,030	-0,030	-0,046	-0,040	-	-0,034	-0,040	-0,048
SUNDAY	-0,097	-0,087	-0,080	-0,079	-0,089	-0,138	-0,236	-0,261	-0,208	-0,175	-0,148	-0,139	-0,143	-0,164	-0,167	-0,176	-0,172	-0,163	-0,137	-0,124	-0,109	-0,106	-0,102	-0,115
SUNDAY_AFTER_WORKSAT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,036	0,027	-	-	-	-
TEMP_F	-0,091	-0,093	-0,084	-0,085	-0,103	-0,097	-0,110	-0,052	-0,041	-0,064	-0,107	-0,113	-	-	-	-	-0,233	-0,228	-0,185	-0,114	-0,084	-0,116	-0,099	-0,104
TEMP_F_S	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
THURSDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,005	-	-	-	-	-	-	-	-
TUESDAY	-	-	-	-	-	-	-	-	-	0,018	0,016	0,015	-	-	-	-	-	-	-	-	-	-	-	-
WORKING_SAT	0,035	0,026	0,020	0,019	0,025	0,035	0,072	0,076	0,061	0,049	0,049	0,050	0,049	0,053	0,052	0,059	0,067	0,057	0,049	0,029	-	0,026	0,038	0,054

R-squared	0,88	0,9	0,91	0,92	0,94	0,96	0,98	0,97	0,966	0,960	0,950	0,950	0,96	0,96	0,96	0,966	0,96	0,96	0,97	0,97	0,94	0,94	0,94	0,948
AIC	-4,5	-4,67	-4,85	-4,9	-5,18	-5	-4,83	-4,62	-4,91	-5,030	-5,060	-5,090	-5,15	-5,1	-5,08	-5,080	-4,93	-4,8	-4,97	-5,06	-4,98	-5,36	-5,28	-5,3
DW	2,1	2,05	2	1,96	1,9	2,1	2	2,03	1,9	1,800	1,700	1,760	1,95	1,95	1,94	1,920	1,99	1,95	1,97	1,99	1,95	1,98	1,92	1,97
LM-test (p-value of F-stat)	0,01	0,2	0,73	0,76	0,05	0,3	0,79	0,7	0,4	0,100	0,010	0,040	0,7	0,66	0,63	0,500	0,98	0,18	0,66	0,9	0,03	0,64	0,05	0,41
RESET (p-value of F-stat)	0,00	0,00	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,04	0,49	0,34	0,00	0,00	0,00	0,00	0,06	0,13	0,14	0,04	0,00	0,02	0,00	0,00

P-values of the heteroscedastic models	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
C	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
_7DAYS_MIDTEMP_F	0,020	0,010	0,003	0,012	0,008	0,000	0,000	0,000	0,000	0,000	0,014	0,042	0,006	0,011	0,000	0,000	-	-	0,000	0,000	0,000	0,000	0,000	0,000
AFTER_HOLIDAY	0,003	0,002	0,009	0,013	0,038	0,000	0,002	0,000	0,005	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
APRIL	-	-	-	-	-	0,016	0,002	-	-	-	-	-	-	-	0,020	0,014	-	-	-	-	-	-	-	-
APRIL*HUMIDITY	-	-	-	-	-	-	0,011	-	-	-	-	-	-	-	-	-	-	-	0,000	0,031	-	-	-	-
APRIL*TEMP_F	-	-	0,020	0,002	0,000	0,014	0,003	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,018	0,012	-	-	-	-	-	-	-	-
AUGUST	-	-	-	-	-	-	-	0,000	0,003	0,002	0,004	0,003	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
AUGUST*TEMP_F	-	-	-	-	-	0,000	0,000	-	0,006	0,003	0,006	0,005	-	-	-	-	-	-	-	-	-	-	-	-
BEFORE_HOLIDAY	-	-	-	-	-	-	-	-	-	-	0,050	0,028	0,003	0,005	0,004	0,001	0,010	0,006	0,001	0,000	-	-	-	-
DECEMBER	-	-	-	-	-	-	-	0,003	0,004	0,023	-	0,001	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*ILLUM	-	-	-	-	-	-	-	-	-	0,012	0,016	0,000	-	-	-	-	-	-	-	-	-	-	-	-
DECEMBER*TEMP_F	-	-	-	-	-	-	-	0,003	0,004	0,027	0,033	-	-	-	-	-	-	-	-	-	-	-	-	-
FEBRUARY	0,000	0,001	0,001	0,002	0,000	0,005	0,004	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,040	-	-	-	-
FEBRUARY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,004	0,001	-	-	-	-	-	-	-	-
FEBRUARY*HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,035	0,005	-	-	-	-	-	-	-	-
FEBRUARY*TEMP_F	0,000	0,001	0,001	0,002	0,000	0,005	0,004	-	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,039	-	0,080	-	0,000
FRIDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,001	0,000	0,000	0,000	-	-	-	-
HOLIDAY	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ILLUM	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY	0,011	0,002	0,000	0,000	0,011	-	-	-	-	0,010	0,003	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,060	-	-	-	-
JANUARY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JANUARY*TEMP_F	-	-	-	-	0,014	0,000	0,000	-	-	-	0,004	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-
JULY	-	-	0,003	0,001	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,001	0,000	0,001	-	-	-	-	-	-	0,009	-
JULY*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	0,023	0,001	0,001	0,000	0,000	-	-	-	-	-	-	-	-
JULY*TEMP_F	-	-	0,003	0,001	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,009	0,001	0,000	0,001	-	-	-	-	-	-	0,007	-
JUNE	-	-	-	-	-	-	-	0,027	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
JUNE*HUMIDITY	-	-	-	-	-	0,000	0,001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JUNE*TEMP_F	-	-	-	-	-	-	0,009	0,030	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
MA(1)	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MA(2)	0,000	0,000	0,000	0,000	0,000	-	0,003	-	-	-	-	-	-	-	-	-	0,000	0,007	0,000	0,000	0,000	0,000	0,000	0,000
MA(3)	0,000	0,000	0,000	0,000	0,000	-	0,003	-	-	-	-	-	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000
MA(4)	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	0,001	0,000	-	-	0,001
MA(5)	-	-	-	-	0,000	-	0,001	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	0,002	0,370	0,097	-
MA(6)	-	-	-	-	-	-	0,000	-	-	-	-	-	0,006	0,011	-	-	0,001	0,001	-	-	0,010	-	-	-
MA(7)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

MARCH	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MARCH*HUMIDITY	0,004	0,001	0,000	0,000	0,000	0,005	0,000	-	0,022	0,018	0,048	0,014	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*ILLUM	-	-	-	-	-	-	-	-	-	-	-	0,031	-	-	-	-	-	-	-	-	-	-	-	-
MARCH*TEMP_F	-	-	-	-	-	-	-	0,000	0,000	0,000	0,001	0,003	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MAX_HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,026	0,006	-	-	-	-	-	-	-	-
MAX_HUMIDITY(-1)	-	-	-	-	-	-	0,009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MAX_ILLUM	-	-	-	-	0,001	0,002	0,000	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MAX_TEMP_F	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MAX_TEMP_F_S	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
MAY	-	-	-	-	-	-	-	0,000	0,036	0,024	0,036	0,039	-	-	-	-	-	-	-	-	-	-	-	-
MAY*HUMIDITY	0,001	0,012	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,013	0,006
MAY*TEMP_F	-	-	-	-	-	-	-	-	0,049	0,030	0,045	0,049	0,000	0,000	0,000	0,000	-	-	0,000	0,007	-	-	-	-
MEAN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,001	-	-	-	-	-	-	-	-
MEAN_ILLUM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-
MIN_HUMIDITY(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MONDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NAT_HOLIDAY	-	-	-	-	-	-	0,016	-	0,021	0,022	0,006	0,006	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-
NET_LOAD(-1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-2)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-4)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-5)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-7)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NET_LOAD(-6)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NOVEMBER	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	0,040	0,026	0,010	-	-	-	-	-
NOVEMBER*TEMP_F	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	0,027	0,017	0,006	0,011	0,013	0,001	0,000	0,004
OCTOBER	-	-	-	-	-	-	0,000	-	-	-	-	-	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
OCTOBER*HUMIDITY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OCTOBER*TEMP_F	-	0,005	0,002	0,001	0,000	0,017	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	0,033	0,000	0,001	-	-	-	-
SATURDAY	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SUMMER_HOL	0,000	0,000	0,000	0,000	0,000	0,000	0,030	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	0,000	-	0,000	0,000	0,000
SUNDAY	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SUNDAY_AFTER_WORKSAT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,000	0,032	-	-	-	-
TEMP_F	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TEMP_F_S	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,000	0,000	-	-	-	-	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
THURSDAY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,042	-	-	-	-	-	-	-	-
TUESDAY	-	-	-	-	-	-	-	-	-	0,000	0,000	0,000	-	-	-	-	-	-	-	-	-	-	-	-
WORKING_SAT	0,014	0,026	0,034	0,034	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,010	-	0,026	0,000	0,000

2015.01.05-2015.12.20	Naive benchmark with seasonal differencing	"SARMA"	Exponential smoothing	Exact-day matching with daily seasonal differencing	Exact-day matching with dummy	EGRV	Own model linear	Own model log-linear
Type of forecast	*one-step ahead*	*one-step ahead*	*one-step ahead*	*one-step ahead*	*one-step ahead*	*one-day ahead*	*one-day ahead*	*one-day ahead*
Monday	45%	466%	1,14%	41%	1,24%	3,18%	2,65%	2,42%
Tuesday	98%	157%	0,91%	123%	0,64%	3,78%	2,53%	2,00%
Wednesday	137%	203%	0,63%	201%	0,76%	2,37%	2,70%	2,04%
Thursday	248%	91%	0,63%	411%	0,59%	2,68%	2,98%	2,16%
Friday	156%	110%	0,75%	221%	0,70%	2,81%	2,99%	2,36%
Saturday	22%	132%	0,99%	23%	0,98%	3,16%	2,73%	2,24%
Sunday	40%	79%	1,08%	51%	1,12%	3,54%	3,59%	3,70%
January	91%	127%	1,07%	111%	0,76%	2,85%	3,15%	1,58%
February	99%	143%	0,76%	140%	0,75%	2,04%	1,99%	1,24%
March	111%	93%	0,79%	151%	0,85%	2,75%	3,93%	2,28%
April	128%	261%	1,21%	161%	0,99%	3,83%	2,72%	2,33%
May	66%	133%	1,03%	62%	0,95%	4,11%	2,82%	2,16%
June	73%	136%	0,68%	91%	0,82%	2,04%	1,96%	2,90%
July	58%	76%	0,62%	85%	0,82%	3,08%	3,28%	2,89%
August	57%	66%	0,79%	79%	0,86%	5,21%	5,47%	4,33%
Spetember	73%	616%	1,12%	86%	0,84%	2,30%	2,72%	2,19%
October	306%	213%	0,99%	484%	1,08%	3,56%	2,22%	2,02%
November	109%	131%	0,74%	226%	0,84%	2,13%	1,84%	2,66%
December	101%	77%	0,61%	144%	0,70%	2,92%	2,62%	2,09%
MEAN of MAPE	**107%**	**173%**	**0,87%**	**151,67%**	**0,85%**	**3,07%**	**2,89%**	**2,39%**
Chow forecast test - winter time	0,67	0,64	NA	1,00	0,00	NA	NA	NA
Chow forecast test - summer time	0,00	0,00	NA	0,00	0,00	NA	NA	NA
Heteroscedasticity	yes	yes	?	yes	yes	no	yes	yes

Reference list

Chatfield, C. and Yar, M. (1988). Holt-Winters Forecasting: Some Practical Issues. *The Statistician*, 37(2), p.129.

Jongbyung Jun A. Tolga Ergün, (2011),"A more accurate benchmark for daily electricity demand forecasts", Management Research Review, Vol. 34 Iss 7 pp. 810 - 820

Gary, B. and Grag, W (2001). An introduction to the Kalman Filter. Course 8. University of North Carolina at Chapel Hill, Department of Computer Science, Chapel Hill, NC 27599-3175

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

Hippert, H.S., Bunn, D.W., Souza, R.C., 2005. Large neural networks for electricity load forecasting: Are they overfitted. International Journal of Forecasting 21,425–434.

Lahouar, A. and Ben Hadj Slama, J. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, 103, pp.1040-1051. Maglogiannis, I. (2007). *Emerging artificial intelligence applications in computer engineering*. Amsterdam: IOS Press

Piras, A., and Buchenel, B. (1999). A tutorial on short term load forecasting. Engineering Intelligent Systems, 1, 41–47.

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

Taylor, J. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19(4), pp.715-725.

Taylor, J., de Menezes, L. and McSharry, P. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1), pp.1-16.

Wooldridge, J. (2013). *Introductory econometrics*. Mason, OH: South-Western Cengage Learning.

Zhang, Y. and Luo, G. (2015). Short term power load prediction with knowledge transfer. *Information Systems*, 53, pp.161-169.