

Article

A New Two-Stage Approach to Short Term Electrical Load Forecasting

Miloš Božić ^{1,*}, Miloš Stojanović ², Zoran Stajić ¹ and Dragan Tasić ¹

¹ Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, Niš 18000, Serbia; E-Mails: zoran.stajic@alfatec.rs (Z.S.); Dragan.Tasic@elfak.ni.ac.rs (D.T.)

² School of Higher Technical Professional Education, Aleksandra Medvedeva 20, Niš 18000, Serbia; E-Mail: milosstojanovic10380@gmail.com

* Author to whom correspondence should be addressed; E-Mail: milos1bozic@gmail.com; Tel.: +381-69-222-3343; Fax: +381-18-293-921.

Received: 28 December 2012; in revised form: 11 March 2013 / Accepted: 1 April 2013 /

Published: 18 April 2013

Abstract: In the deregulated energy market, the accuracy of load forecasting has a significant effect on the planning and operational decision making of utility companies. Electric load is a random non-stationary process influenced by a number of factors which make it difficult to model. To achieve better forecasting accuracy, a wide variety of models have been proposed. These models are based on different mathematical methods and offer different features. This paper presents a new two-stage approach for short-term electrical load forecasting based on least-squares support vector machines. With the aim of improving forecasting accuracy, one more feature was added to the model feature set, the next day average load demand. As this feature is unknown for one day ahead, in the first stage, forecasting of the next day average load demand is done and then used in the model in the second stage for next day hourly load forecasting. The effectiveness of the presented model is shown on the real data of the ISO New England electricity market. The obtained results confirm the validity advantage of the proposed approach.

Keywords: short-term load forecasting; least-squares support vector machines; average daily load; two-stage approach

1. Introduction

With the deregulation of the energy market and the promotion of the smart grid concept, load forecasting has gained even more significance. Generation capacity scheduling, coordination of hydro-thermal systems, system security analysis, energy transaction planning, load flow analysis and so on are all tasks which rely on accurate short-term load forecasting (STLF) [1]. On the other hand, electric load is a random non-stationary process which is influenced by a number of factors, including: economic factors, time, day, season, weather and random effects, all of which leads to load forecasting being a challenging subject of inquiry.

During the past few decades, a wide variety of models have been proposed for the improvement of STLF accuracy. Conventional methods include: linear regression methods [2], exponential smoothing [3] and Box–Jenkins ARIMA approaches [4] which are linear models which cannot properly represent the complex nonlinear relationships between loads and their various influential factors. Artificial intelligence-based techniques are employed because of the good approximation capability for non-linear functions. These methods include: Kalman filters [5], fuzzy logic [6,7], knowledge-based expert system models [8], artificial neural network (ANN) models [9,10] and support vector machines (SVMs) [11,12]. No single model has performed well in STLF and hybrid approaches are being proposed to take advantage of the unique strength of each method. An adaptive two-stage hybrid network with a self-organized map and support vector machines is presented in [13]. A hybrid method composed of a wavelet transform, neural network and evolutionary algorithm is proposed in [14]. A combined model based on the seasonal ARIMA forecasting model, the seasonal exponential smoothing model and weighted support vector machines is presented in [15] with the aim of effectively accounting for the seasonality and nonlinearity shown in the electric load. Another seasonal model which combines the seasonal recurrent support vector regression with a chaotic artificial bee colony algorithm is proposed in [16] to determine the appropriate values of three parameters of SVRs.

In spite of all the performed research in the area of STLF, more accurate and robust load forecast methods are still required. One can also highlight some interesting works in this area, especially in recent years. A combined aggregative STLF method for smart grids which obtain a global forecasting by summing up the forecasts on the compounding individual loads is introduced in [17], with three new approaches proposed: bottom-up, top-down and regressive aggregation. A new singular value decomposition based exponential smoothing method is presented in [18], where it is applied to the intraweek cycle, which leads to a simpler and potentially more efficient model formulation. The new method is similar to the Holt-Winters exponential smoothing method, but both were outperformed by the unrestricted form of intraday cycle exponential smoothing. A combined forecast model constructed as the simple average of the weather-based method, the Holt-Winters exponential smoothing and proposed method, obtained the best results at all horizons. Also, these univariate methods outperformed a weather-based method up to about five hours ahead. In [19] an integrated approach which combines a self-organizing fuzzy neural network method with a bilevel optimization method is proposed for STLF. The proposed approach uses self-organizing fuzzy neural network advantage to automatically determine both the model structure and parameters, and bilevel optimization method advantage to automatically select the best pre-training parameters to ensure that the best fuzzy neural networks are identified. In [20], the comparison between the frequently used radial basis function network in STLF

and the modified radial basis function network with a genetic algorithm for weight estimation and a nonsymmetrical penalty function with different penalties for over-forecasting and under-forecasting is presented. The obtained results show the efficiency of the proposed method with the new forecasting metric which is the extension of the conventional sum of the squared error metric. Two methodologies for bus load forecasting, *i.e.*, multimodal load forecasting are proposed in [21], where one individually forecasts the local loads while the second forecasts the global load and then individually forecasts the load participation factors to estimate the local loads. In both methodologies a modified general regression neural network with automatic feature selection to reduce the number of inputs of the artificial neural networks is used.

In order to improve forecasting accuracy, in this paper emphasis is placed on model features in the context of machine learning models. It is well known that the balance between the size of the feature set and the quality of the chosen features is important, regardless of which method is used for modeling. A small feature set cannot provide enough information about the load and, on the other hand, too many features do not necessarily provide more information, but may bring noise to the model. The selection of appropriate model features which carry the right information about load behavior is one of the most important tasks. An analysis of what kind of information should be included in the model for mid-term load forecasting was done in [11] and a winning model feature set consists of calendar weekday features and time-series past load demand features. The approach in [22], in addition to the weekday calendar features, proposed using the hour of the day feature in STLF problems, and also suggested the use of temperature as the most important weather variable because of the strong correlation between temperature and load. Other weather variables (wind velocity and cloud cover) are also analyzed but in the end are neglected. The final feature set consists of an hour indicator, day indicator and estimated temperature at the hours k , $k - 1$ and $k - 2$, without using time-series past load. As load time series indicated a clear daily and weekly seasonality, in [23] the effects of the days of the week and special days, such as holidays, are included in the model. To model these effects, several features are introduced besides weekday features such as holidays, working days after or before a holiday, work only during the mornings or only during the afternoons, the Saturday after a holiday, special holidays and so on. Also, in order to choose the appropriate feature subset which best describes the load, in some papers the choice of features is not done manually, and it is common to use some of the algorithms for feature selection. In [24], ant colony optimization is applied to yield optimal feature subsets. The initial feature set is composed of 38 features which are selected to describe hourly and weekly load behavior and the correlation with weather variables. Some included features are the maximum, minimum and average temperatures during the last seven days, six temperature points on the forecasted day, forecasted day rainfall, wind speed, humidity, cloud cover, month, season, week, whether the day is a holiday or not, whether the day is a weekend or not and so on. At the end of the feature selection, 21 features were dropped from the initial set. The features have been selected by using a cross-correlation analysis in [25]. The feature set is composed of the previous hour load, the load of the previous day, load of the previous week and the load from two weeks ago.

It may be noted that the list of used features is wide and varies from work to work but they all have the same goal, to improve the model and achieve the best forecast accuracy. With the same aim, in this paper a new approach to STLF is proposed. An additional feature, next day average load demand, is appended to the STLF model feature set. As this feature is unknown for the next day, in the first stage,

the forecasting of the average daily load is carried out. Then, in the second stage, the forecasted average daily load is incorporated into the STLF model and the forecasting of the hourly load for the next day is carried out. It is important to emphasize here that the proposed approach is distinguished from others by the use of the average load in the model, such as for example the Box-Jenkins approach, in terms of using it in the context of the machine learning model, more concretely the LS-SVM. In this way this feature has direct influence in the training phase of the model formation. The results obtained from experiments on the real electricity market data indicate the validity and advantage of this approach.

The rest of the paper is organized as follows: Section 2 presents the basics of least-squares support vector machines (LS-SVM) used in the regression. Next, Section 3 shows electrical load data features and presents the proposed STLF approach. Section 4 includes a variety of experiments to verify the proposed approach. Finally Section 5 outlines the conclusions.

2. Least Squares Support Vector Machines Model

The brief basic concepts of LS-SVMs are introduced. SVMs were proposed by Vapnik in [26], and are widely used for load forecasting, in addition to ANNs which also show a good approximation capability for non-linear functions. However, SVMs are based on the structural risk minimization principle in order to minimize the upper limit of the estimation error, rather than the empirical risk minimization which minimizes the training error used by ANNs. Consequently, by solving the quadric programming (QP) optimization problem, SVMs always manage to achieve the global optimum solution, instead of possibly stocking the local optimum like ANNs models. This approach, by using nonlinear kernels, leads to a very good generalization performance and sparse solutions. LS-SVMs, defined in [27], as reformulations of standard SVMs instead of solving the QP problem, which is complex to compute, obtain a solution from a set of linear equations. Therefore, LS-SVMs have a significantly shorter computing time and they are easier to optimize.

Let us consider a given training set $\{x_k, y_k\}$, $k=1, \dots, n$ with inputs $x_k \in R^p$ and outputs $y_k \in R$. The following regression model can be built by using a non-linear mapping function $\phi(\cdot): R^p \rightarrow R^{P_h}$ which maps the input space into a high-dimensional feature space and constructs a linear regression in it. The regression model in primal weight space is expressed as follows:

$$y(x) = \omega^T \phi(x) + b \quad (1)$$

where ω represents the weight vector and b is a bias term.

LS-SVM formulates the optimization problem in primal space presented as follows:

$$\min J_p(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^n e_k^2 \quad (2)$$

subject to equality constrains expressed as follows:

$$y_k = w^T f(x_k) + b + e_k, k = 1, \dots, n \quad (3)$$

where e_k represents error variables; γ is a regularization parameter which gives the relative weight to errors and should be optimized by the user.

In order to solve the optimization problem defined with Equations (2) and (3), it is necessary to construct a dual problem using the Lagrange function. Once the mathematical calculations were carried out, described in detail in [27], the following linear system was obtained:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + I\gamma^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (4)$$

In Equation (4), $y = [y_1, \dots, y_n]^T$, $1_v = [1, \dots, 1]^T$, $\alpha = [\alpha_1, \dots, \alpha_n]^T$ there are Lagrange multipliers, I is an identity matrix and $\Omega_{kl} = \phi(x_k)^T \phi(x_l) = K(x_k, x_l)$, $k, l = 1, \dots, n$ denotes the kernel matrix.

Once the system defined in Equation (4) is solved, the solutions for α and b are obtained. It is shown in [27] that usually all Lagrange multipliers are non-zero, which means that all training data participate in the solution, *i.e.*, every data point represents a support vector. Compared with SVM, the LS-SVM solution is not sparse.

The resulting LS-SVM model for function estimation in dual form is defined as follows:

$$y(x) = \sum_{k=1}^n \alpha_k K(x, x_k) + b \quad (5)$$

The dot product $K(x, x_k) = \phi(x)^T \phi(x_k)$ is known as a kernel function. Kernel functions that satisfy Mercer's condition enable computation of the dot product in a high-dimensional feature space by using data inputs from the original space, without explicitly computing $\phi(x)$.

A commonly used kernel function in non-linear regression problems, one that is employed in this study, is a radial basis function represented as follows:

$$k(x, x_k) = e^{-\frac{\|x-x_k\|^2}{\sigma^2}} \quad (6)$$

where the kernel parameter σ^2 denotes the squared variance of the Gaussian function.

When choosing the RBF kernel function with the LS-SVM, the optimal parameter combination (γ , σ) should be established, where γ denotes the regularization parameter and σ is a kernel parameter. It can be noticed that only two additional parameters (γ , σ) need to be optimized, instead of three (γ , σ , ε) as in SVM. Parameter selection is the most significant part during the formation of the LS-SVM regression model, because it has a significant effect on the performance, both in terms of accuracy and computing time. Accordingly, for this purpose, a grid search algorithm in combination with k -fold cross validation was used in this study.

3. Model Formation

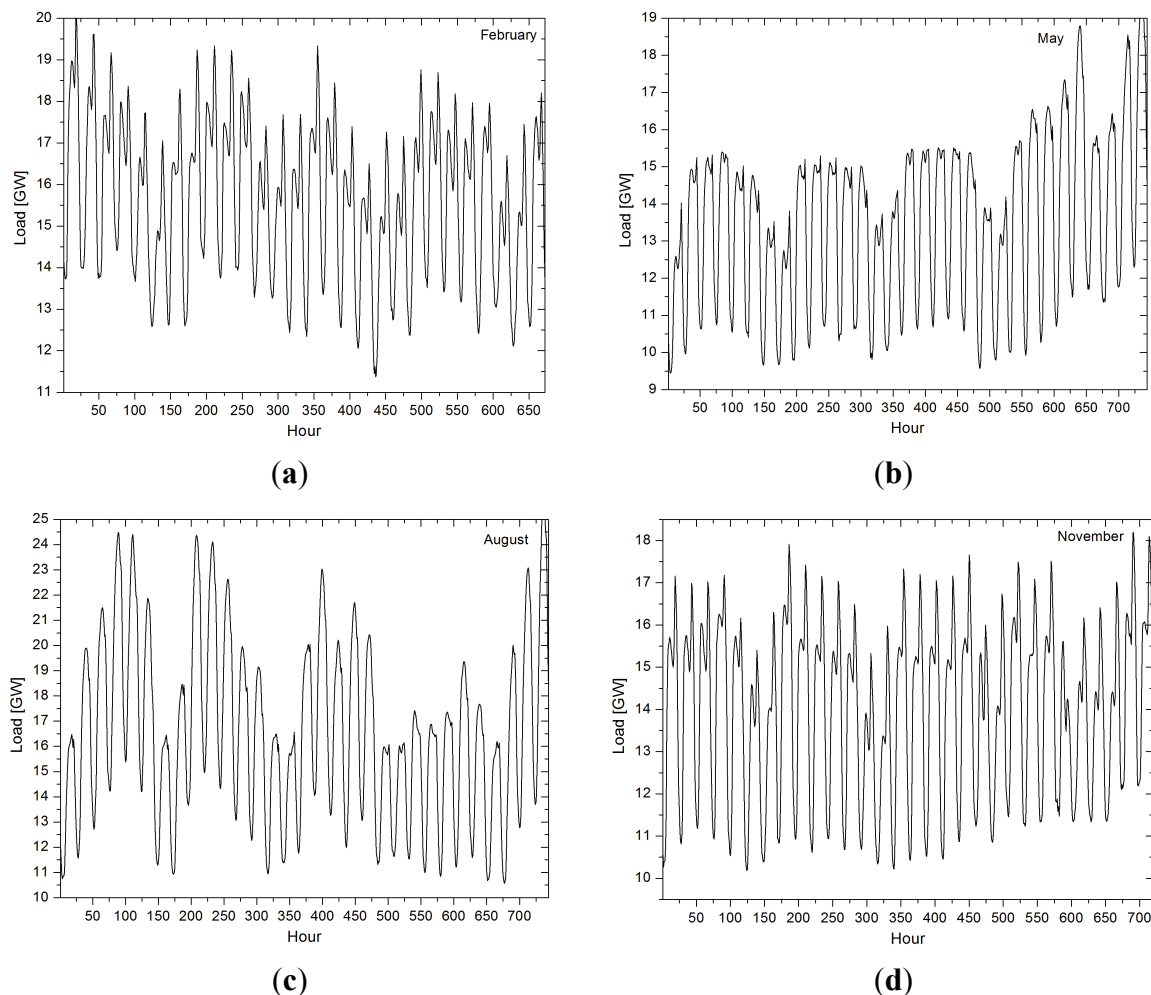
3.1. Features of Electric Load

The electric load is a random non-stationary process influenced by a number of factors which makes it difficult to model. Choosing appropriate input features to build the model is an important task in load forecasting. There is no general approach to conduct this problem, but load curve analyses and statistical analyses can be helpful for choosing key features to build a good load forecasting model.

The real-life STLF test case is considered in this paper to evaluate the performance of the proposed forecast approach. This STLF test case is related to the ISO New England power system, which is an

electricity market in the U.S. The employed data for the load in this test case are publicly available data obtained from a website [28]. Figure 1 shows the power load curves for four months, which are typical representatives of each quarter of the year.

Figure 1. Power load curves. (a) February 2011; (b) May 2011; (c) August 2010; (d) November 2010.



In Figure 2, the hourly load during the week is presented for four weeks in February, May, August and November. It is obvious that the daily load on work days is greater than the load on weekends. The reasons for this are people's behavior during the week, and this pattern is periodically repeated each week. All this imposes using the day of the week for the features in the model.

Figure 3 shows hourly load during the day for each day in one week in February, May, August and November. This curve is influenced and shaped by people's daily habits. The load changes from hour to hour during the day, indirectly following consumer behavior. This brings one more important variable to the feature set, and that is the hour of the day. Also, it can be noticed that the curves have a similar shape but different magnitude from day to day in the week. This also confirms the validity of using the day of the week for the model feature with the aim of mapping this property. However, from Figures 2 and 3 it can be observed that the daily load curve is different for the four given months. This difference is reflected not only in load magnitude but also in the shape of the load curves.

Figure 2. Hourly load during the week in (a) February 21–27; (b) May 16–22; (c) August 16–22; (d) November 15–21.

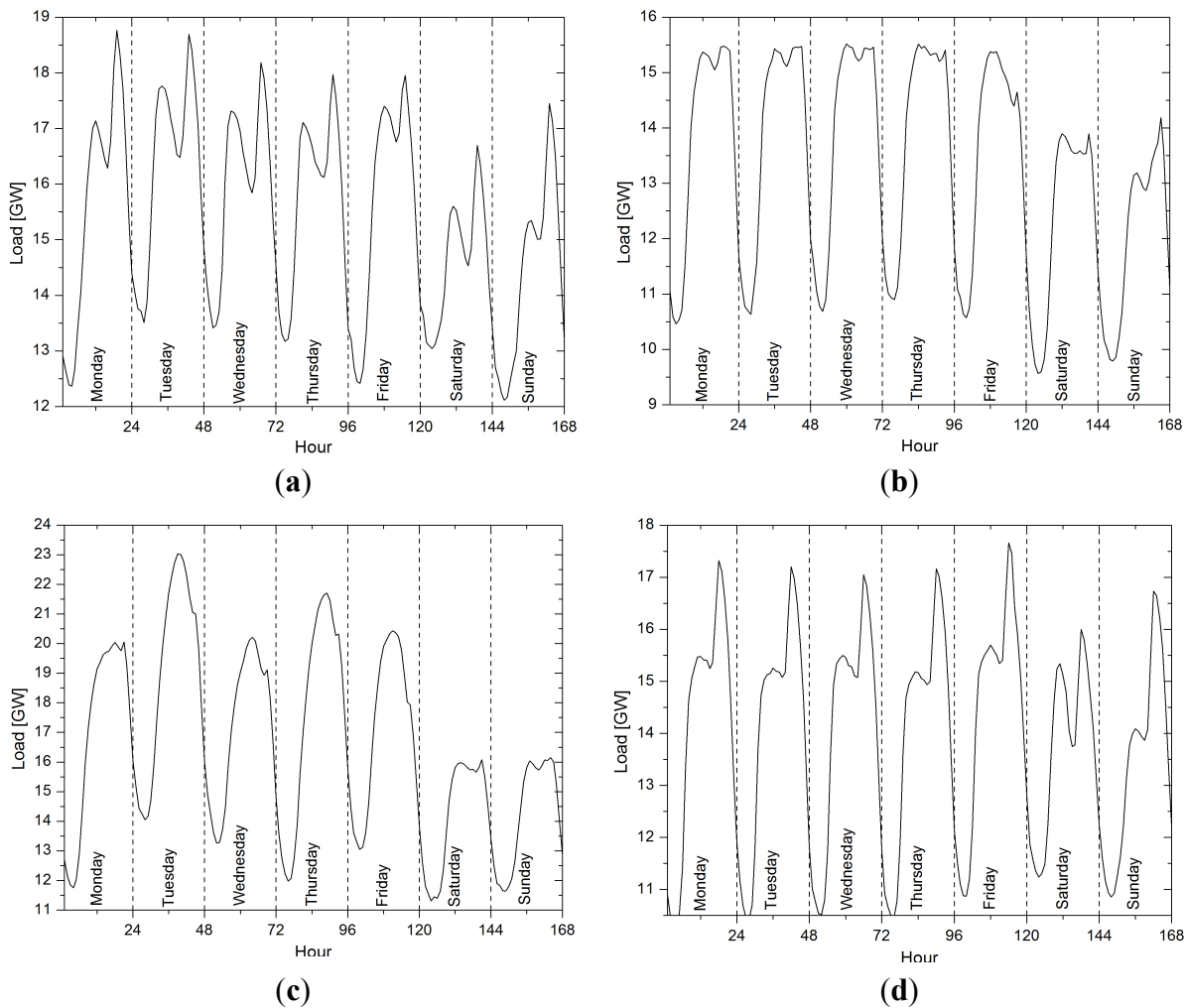


Figure 3. Hourly load during the day. (a) February 21–27; (b) May 16–22; (c) August 16–22; (d) November 15–21.

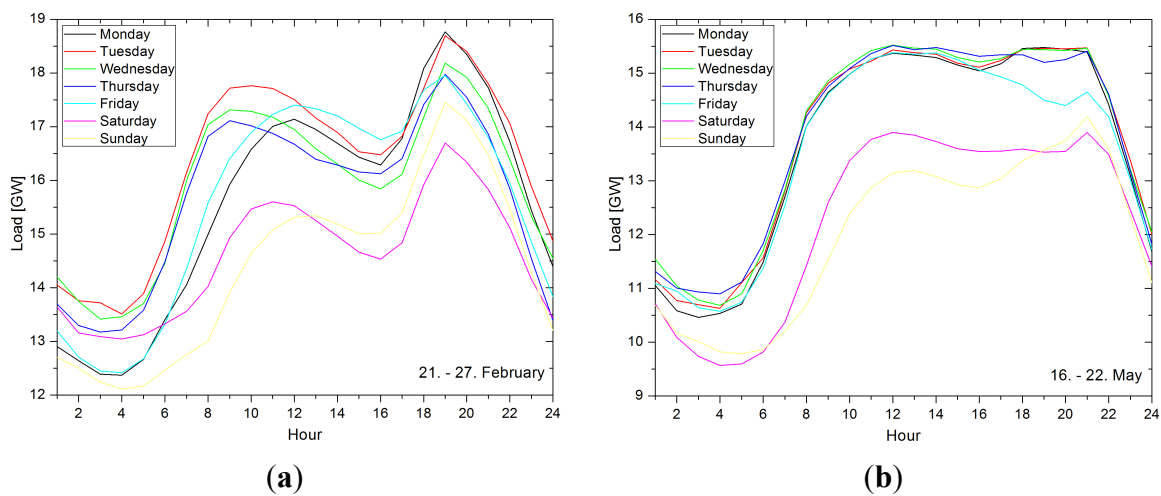
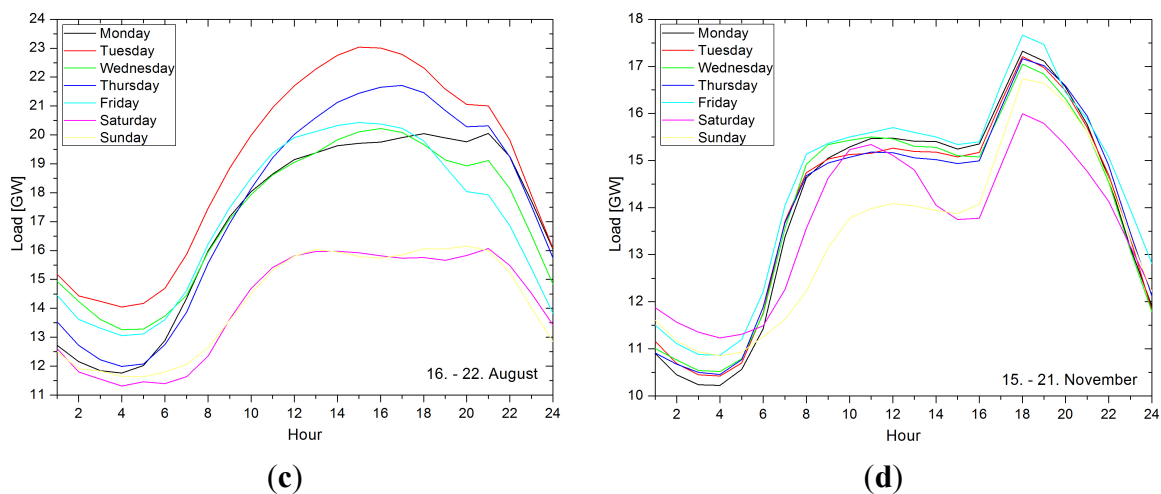
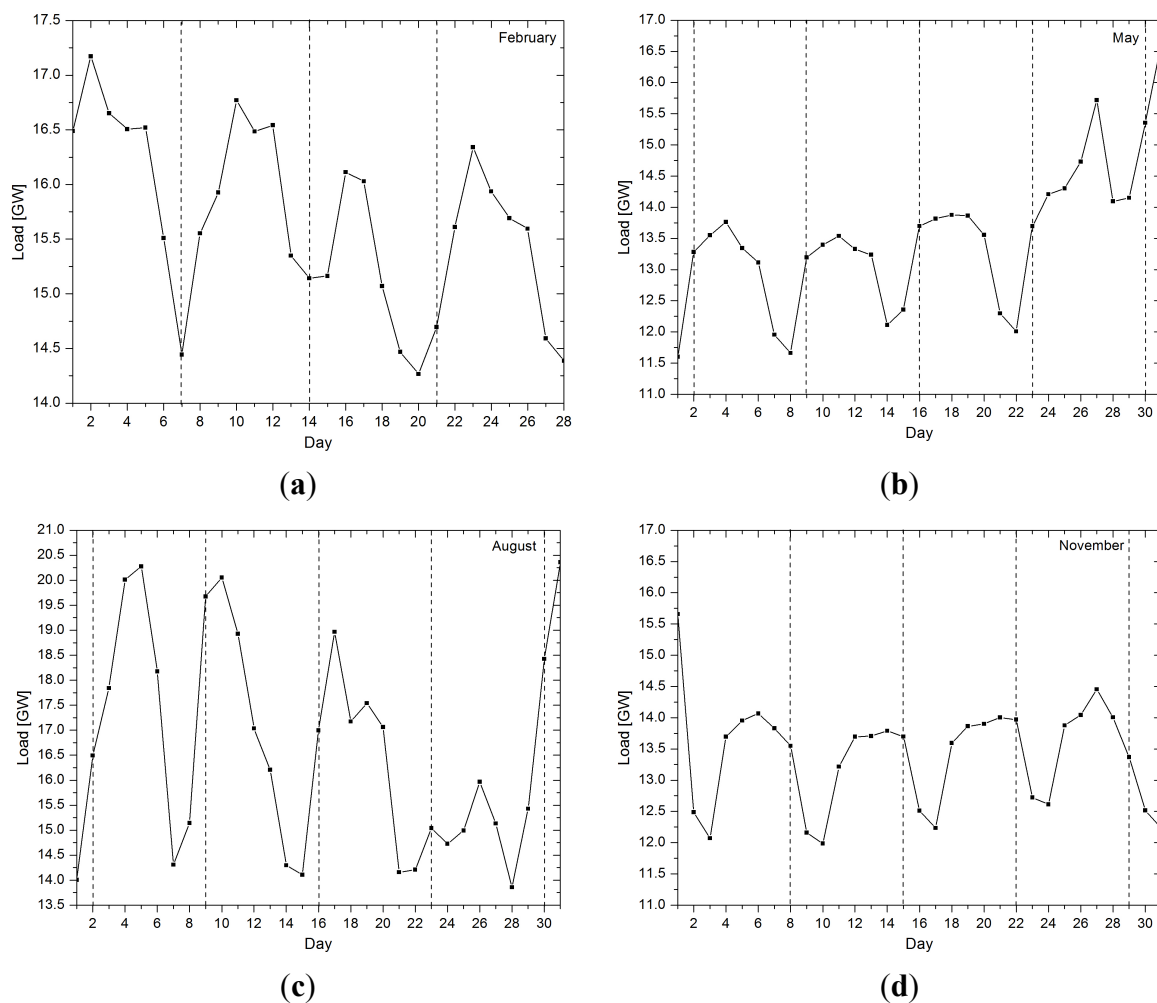


Figure 3. Cont.



In Figure 4, the average daily loads in February, May, August and November are presented. The start of the week (Monday) is marked with dashed vertical lines.

Figure 4. Average daily load in (a) February 2011; (b) May 2011; (c) August 2010; (d) November 2010.

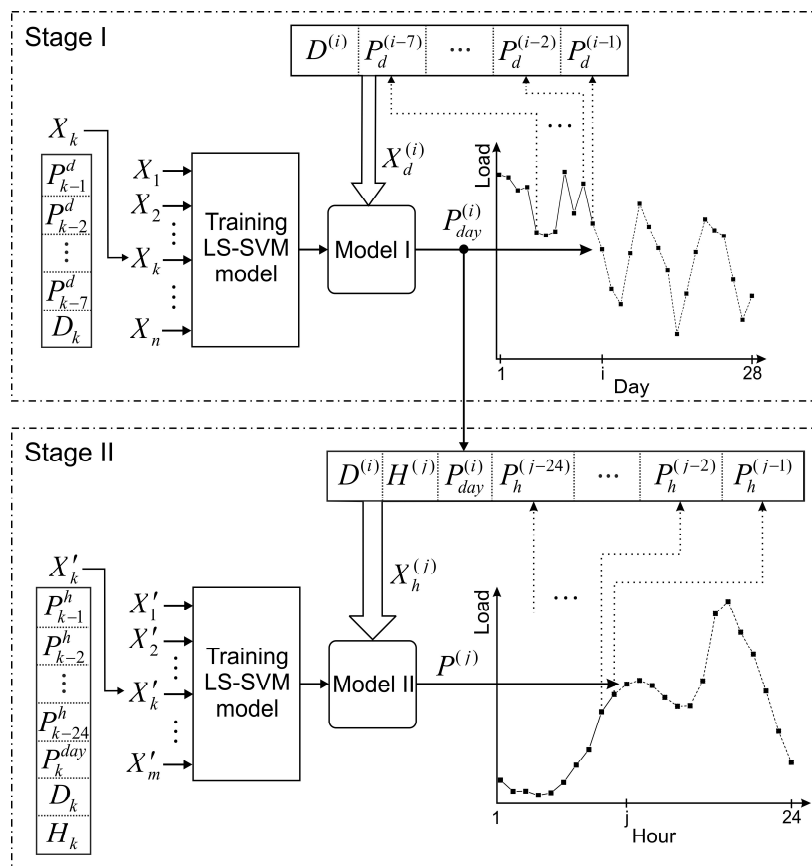


It is clear that the average daily load on the weekends is smaller than on week days. It also can be seen that power consumption on Tuesday and Wednesday is much greater than on the other days.

3.2. The Proposed Approach

As previously described, electric load is a nonlinear, time variant and multi-variable function. It is very difficult to capture the correct mapping function of such a signal in all the time spans. To solve this problem, a new two-stage STLF approach based on least squares support vector machines with the architecture shown in Figure 5 is proposed in this paper. Beside the Figure 5 which shows graphical representation of the proposed approach, a step-by-step procedure of two-stage LS-SVM model training and forecasting is given in algorithm 1.

Figure 5. The proposed two-stage model architecture.



In the first prediction stage, Stage I, forecasting of the next day average load is done. This is performed by Model I, whose inputs consist in total of $t + s$ features, where t is the number of past average daily load time-series features and s is the number of non-time series features. The past average daily load time horizon is set to $t = 7$, i.e., the model uses the last seven average daily loads from the prediction moment ($P_{k-i}^d, i = 1, \dots, 7$). To map the weekly load behavior, the day of the week feature ($D_k, D_k \in \{1, 2, \dots, 7\}$ where 1 corresponds to Monday, 2 to Tuesday and so on) is included in the feature set and this feature is the only non-time series feature, i.e., $s = 1$.

Algorithm 1. The two-stage LS-SVM model training and forecasting procedure

1. Stage I

- 1.1. Model I training set formation using daily average load data for the past three years. This training set contains 1095 vectors in total and each vector is composed of features from seven past average daily loads and the current day of the week indicator. Normalize all of the features in the [0–1] range by using min-max normalization,
- 1.2. Based on this training set and grid-search algorithm with a k -fold cross validation procedure ($k = 10$), obtain the optimal parameters γ and σ for the LS-SVM Model I,
- 1.3. Using Equations (5) and (6) and the previously optimized parameters γ and σ train the LS-SVM forecasting Model I,
- 1.4. In order to predict the average load for one step ahead, *i.e.*, for the next day, seven past average daily loads and the next day of the week indicator form the input test vector for model I,
- 1.5. At the end of stage I, the average load for the next day is obtained and passed on to Stage II.

2. Stage II

- 2.1. Model II training set formation using hourly load data for the corresponding months from three previous years. This training set contains 2016 vectors in total and each vector is composed of features from 24 past hourly loads, the current day of the *week* indicator, the current hour of the day indicator and the current average daily load. Normalize all of the features in the [0–1] range by using min-max normalization,
- 2.2. Based on this training set and grid—search algorithm with a k -fold cross validations procedure ($k = 10$) obtain the optimal parameters γ and σ for the LS-SVM model II,
- 2.3. Using expressions (5) and (6) and the previously optimized parameters γ and σ train the LS-SVM forecasting Model II,
- 2.4. Now, the input test vector is formed from the 24 past hourly loads, the next day of the week indicator, the next hour of the day indicator and the average load for next day, obtained from Model I in Stage I,
- 2.5. Employ model II with the test vector for the prediction of the hourly load for one step ahead, *i.e.*, for the next hour,
- 2.6. Update the test vector, first shift the 24 past hourly loads one place to the left and then add the prediction for the past hour in last place, then, update the hour of the day indicator (the day of the week indicator and the daily average load remains the same),
- 2.7. Go to Step 2.5. until the prediction of the hourly loads for the 24 steps ahead are obtained,
- 2.8. At the end of stage II, the hourly load for next day is obtained.

When the structure of the inputs is defined, the training set which contains an n number of inputs is formed. For Model I training, the total number of inputs is set to $n = 1095$, *i.e.*, the training set contains inputs for the previous three years before the prediction moment. As the experiment results will show, this value is sufficient to catch the evolving nature of the average load pattern. After establishing the training set, the training of the LS-SVM forecasting of Model I is performed. In order to have an optimal training of the model, the data set has to be normalized before training. This

prevents the dominance of any features in the output value and provides faster convergence and better accuracy of the learning process. Accordingly, all of the features are normalized within the range [0–1]. After that, the optimal (γ, σ) pair is determined on a training set using a grid search with k -fold cross validations, as mentioned in Section 2.

The training set is randomly subdivided into k disjoint subsets of approximately equal size and the LS-SVM model is built k times with the current pair (γ, σ) . Each time, one of the k subsets is used as the test set and the other $k-1$ subsets are put together to form a training set. After k iterations, the average model error is calculated for the current pair (γ, σ) . The entire process is repeated with an update of the parameters (γ, σ) until the given stopping criterion (e.g., Mean Squared Error) is reached. The parameters (γ, σ) are updated exponentially in the given range using predefined equidistant steps, according to the grid-search procedure. After obtaining the optimal (γ, σ) combination, values for a and b are obtained from Equation (4), and the LS-SVM Model I is formed according to Equations (5) and (6). The test vector is constructed in regard to the previously defined feature set structure and Model I is then employed for the prediction of the average load for one step ahead, *i.e.*, for the next day. When the next day average load is obtained, it is passed to Stage II where the forecasting of the next day hourly loads is done. The Model II feature set is also composed of both time-series and non time-series types of features. The past hourly load time horizon used for this model is $t = 24$, *i.e.*, the model uses the last 24 hour loads from the prediction moment $(P_{k-i}^h, i = 1, \dots, 24)$. In addition to these time-series parts, the model feature set contains three non time-series features ($s = 3$): the hour of the day H_k , $H_k \in \{1, 2, \dots, 24\}$, the day of the week D_k , $D_k \in \{1, 2, \dots, 7\}$ and average daily load P_k^{day} . In the training phase of Model II, the last mention feature is obtained as an average from the history of hourly loads, and therefore has an exact value, while in the prediction phase this value is obtained in Model I, and therefore represents a predicted value. After defining the structure of training inputs, the Model II training set is formed from approximately $m = 2016$ inputs, *i.e.*, the training set contains hourly inputs from three months in the past three years, e.g. if the hourly loads for each day in February 2012 need to be predicted, the training set consist of the inputs from February 2009, 2010 and 2011. This is not necessary but it is shown in [11] that the training set calendar congruence with the predicted period produces better forecasting accuracy and reduces the time needed for model formation. After establishing the training set, training of the LS-SVM forecasting Model II is performed in the same manner as the training of Model I. After Model II is trained, it is then committed with the test vector which is formed in regard with the previously defined feature set structure, and the prediction of load for one step ahead, *i.e.*, for the next hour is done. After that it is necessary to update the test vector for the next prediction step, *i.e.*, for the next hour. The update is needed because the exact values of the load for the past 24 hours are available only for the first prediction step. After that, for the next predictions, the predicted values from the previous steps are used instead of the exact ones, which are unknown at that moment. Accordingly, the test vector is first shifted left for one place, the hour feature is updated (the day and average load features remain for the current day) and prediction from the previous step is placed in the final position. The whole process is repeated 24 times and in the end, hourly predictions for the next day will be obtained.

4. Experimental Results

For the evaluation of the proposed STLF approach, the forecasting of hourly loads for four typical month representative of each quarter of the year was done for each day. The results are obtained for August 2011, November 2011, February 2012 and May 2012. This implies that the results from the Stage I forecasting model for the prediction of the next day average load, must first be obtained. Also, the evaluation of these results is important, because they directly influence final STLF accuracy and provide insight into the extent of this dependence, and that is a useful indicator of new feature contributions to STLF accuracy.

The prediction quality is evaluated using the Mean Absolute Percentage Error (MAPE), Maximum Error (ME) and Absolute Percent Error (APE) as follows, respectively:

$$MAPE[\%] = 100 \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - \hat{P}_i}{P_i} \right|, \quad (7)$$

$$ME = \max_i |P_i - \hat{P}_i|, \quad (8)$$

$$APE[\%] = \left| \frac{P_i - \hat{P}_i}{P_i} \right| \cdot 100, \quad (9)$$

where P_i and \hat{P}_i are the real and the predicted value of the load demand in the i^{th} hour and n is the number of hours.

Real and predicted average daily loads are shown in Figure 6 for August, November, February and May respectively. In the same Figure, daily APEs are given to illustrate the deviation in the prediction of next day average load.

Figure 6. Real and predicted average daily load with APE. (a) August 2011; (b) November 2011; (c) February 2012; (d) May 2012.

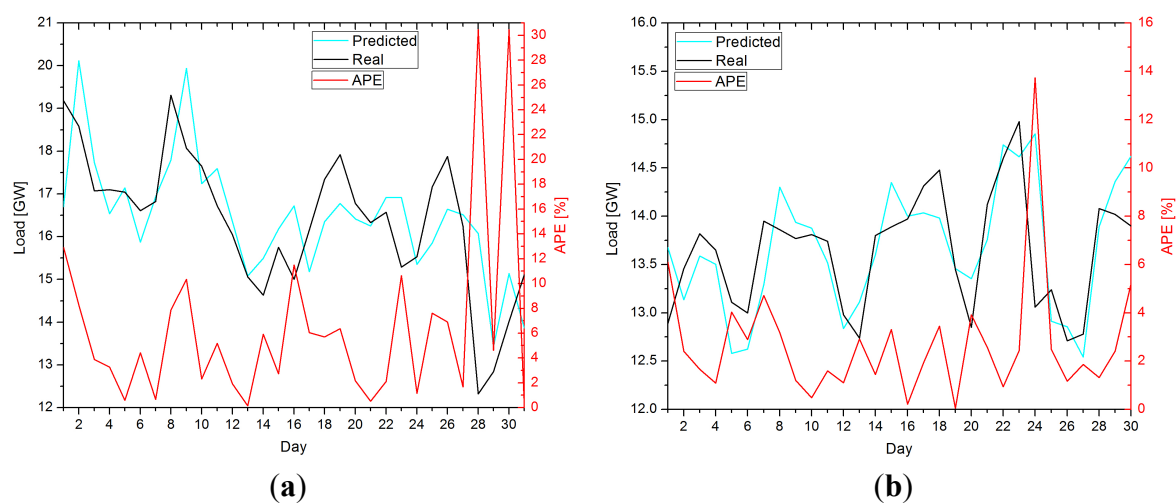
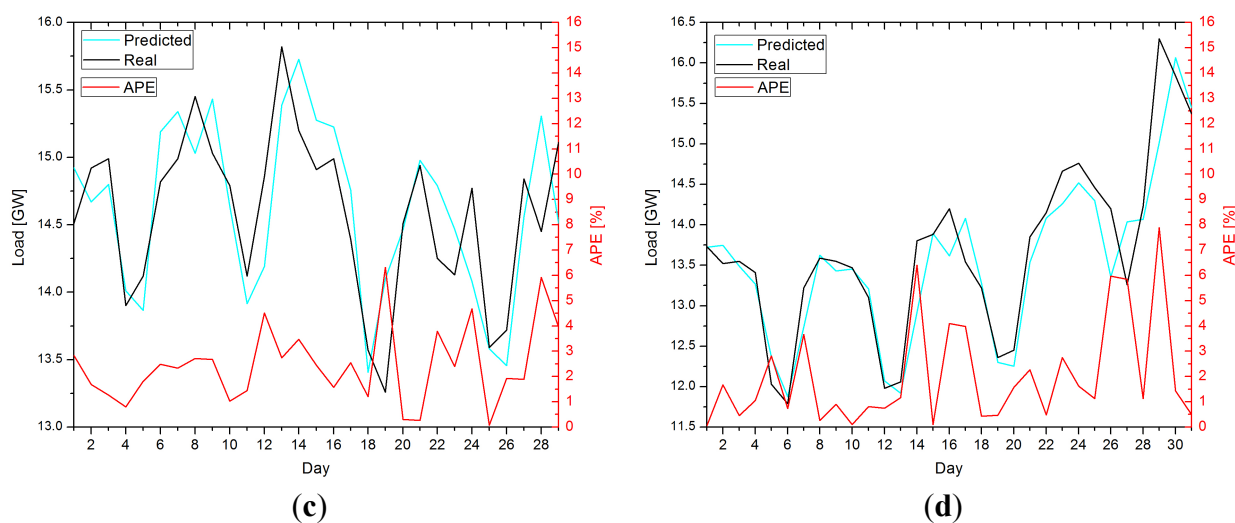


Figure 6. Cont.



In Table 1, minimum, average and maximum APE values of entire test sets are shown to also give an indication of the range of APE values in addition to the graphic representations. The first column indicates the test month set, while the second to forth indicate minimum, average and maximum monthly APE values. These APE values fall within scope of interest not because the development and evaluation of the next day average load forecasting model was carried out here, but because we are interested in how the proposed STLF model will behave using the predicted next day average load values in that range.

Table 1. Daily APE for next day average load prediction.

Set	APE		
	Minimum	Average	Maximum
August	0.14	6.12	30.47
November	0.06	2.72	13.73
February	0.08	2.52	6.32
May	0.05	2.02	7.88

Figure 6 and Table 1 give as a sense of the range of the forecasted average load APE for each day in test sets. Thus the days that do not have a satisfactory average load forecasting accuracy can be identified with the aim of monitoring the results of hourly load forecasting on these days. It is of interest because the forecasted average load at stage I is used as input at stage II, where hourly load forecasting is done, as stated above.

To examine the STLF model behavior when it uses the next day average load feature with different APE values, three sets for two test month of next day average loads were artificially generated using the reverse process of calculating APEs with respect to APE values of 2.5, 5 and 7.5%. This resembles a prediction of next day average loads, where the obtained values are in the range of 2.5, 5 and 7.5 of the APE for each day in the test set. When these artificially generated values are collected, they are used as a feature in the input vector for Model II and the forecasting of the next day hourly loads are carried out. In Table 2, the STLF results obtained using artificially generated values for next day average loads are shown. The first column indicates the test month set and the second, the artificially

generated value in the input vector, where I2.5 means an artificial next day average load with 2.5 APE, I5 with 5 and I7.5 with 7.5 APE. The remaining columns contain values for minimum, average and maximum monthly values of MAPE and ME. From this table it can be observed that the MAPE and ME values, regardless of whether they are minimum, average or maximum values, increase with the rise in the APE of the next day average load artificially generated values used in the input vector. Thus, it can be noted that the accuracy of the proposed STL model will increase with an increase in the next day average load forecasting model accuracy, *i.e.*, if the next day average load predicted value is closer to the real value, then the STL model will also give accurate predictions.

Table 2. Average, max and min daily MAPEs and MEs, obtained with artificial inputs during Stage II.

Set	Input	MAPE			ME		
		Minimum	Average	Maximum	Minimum	Average	Maximum
February	I2.5	2.43	2.92	4.38	0.58	0.96	2.05
	I5	4.48	5.33	7.04	0.92	1.55	2.14
	I7.5	6.57	7.43	8.58	1.49	2.1	2.91
May	I2.5	2.26	3.16	5.71	0.63	0.98	1.6
	I5	4.14	5.12	6.14	0.88	1.44	2.26
	I7.5	5.81	7.32	9.77	1.18	1.99	2.81

To give a graphic representation of the STL accuracy of the proposed approach, from its obtained results for test sets, daily MAPEs are calculated and shown in Figure 7. In this figure, five curves for each test month can be seen, each corresponding to the LSSVM-I, LSSVM-TSTL, LSSVM-TS, DS-ARIMA and DS-EST model respectively. The LSSVM-I (least square support vector machines initial) model curves represent daily MAPEs for initial model forecasting, *i.e.*, a model whose feature set consists of 26 features: days of the week, hours of the day and 24 past load time-series features. In addition to the features in the LSSVM-I model, models LSSVM-TSTL (least square support vector machines two-stage true average load) and LSSVM-TS (least square support vector machines two-stage) have one more feature, the next day average daily load. Although the LSSVM-TSTL and LSSVM-TS models share the same model structure, they have different inputs in the prediction step. The LSSVM-TSTL model in the input vector for next day average load feature uses exact values, which cannot be used in the real scenario because this value is not known for the step forward, while the LSSVM-TS model uses previously predicted values from Stage I. In addition, due to the verification of performance of a proposed method, the double seasonal ARIMA model (DS-ARIMA) proposed by Taylor *et al.* [29] and the double seasonal exponential smoothing model (DS-EST) proposed by Taylor [30], are also involved in the comparison.

Bearing in mind the obtained results for average daily load in Figure 6, the days characterized by higher MAPEs can be recognized. This refers to the days when the MAPEs are at least twice the values of the average daily MAPEs for a given month. As can be seen in Figure 7, on these days daily MAPEs for the proposed model LSSVM-TS are higher compared to the model LSSVM-TSTL which uses a true next day average load, *i.e.*, prediction accuracy is reduced as a result of inaccurate next day average load forecasting at stage I. This behavior is especially pronounced in several days in each test month, so for example on days 1, 9, 16, 23, 28 in August, 1, 7, 24, 30 in November, 1, 6, 12, 19, 22,

24, 28 in February and 5, 7, 14, 16, 17, 26, 27, 29 in May. On these days the difference in MAPEs is significantly expressed compared to the LSSVM-TSTL model, but on the other hand on days when the predicted average daily load is nearly equal to the real average daily load, there was a significant improvement in the forecasting accuracy at stage II. This does not mean that the on previously mentioned days with a slightly larger MAPE at stage I there was no improvement compared to the initial LSSVM-I model, which does not use next day average load in the feature set. Also, it should be noted that there are days for the proposed LSSVM-TS model with obtained MAPEs greater than those of the initial LSSVM-I model. These are for example the following days: 1, 7, 17 in August, 15 in November, 5, 6, 19, 24 in February and 5, 6, 7, 14 in May. The reason for this is that on these days the inaccurate next day average load was used in stage II, *i.e.*, as can be seen in Figure 7 on these days in the LSSVM-TSTL model with real next day average load gain, better MAPEs were determined compared to the proposed LSSVM-TS model, but also compared to the initial LSSVM-I model. This is not entirely true for days 7 in August, 6 in February and 6 in May where the initial LSSVM-I model obtained better MAPEs than the LSSVM-TSTL model. That can be expected in some situation when the hourly load curve is not strongly correlated with the daily average load, which then gives faulty information to the model.

Figure 7. Daily MAPEs for all of STLF models. (a) August; (b) November; (c) February; (d) May.

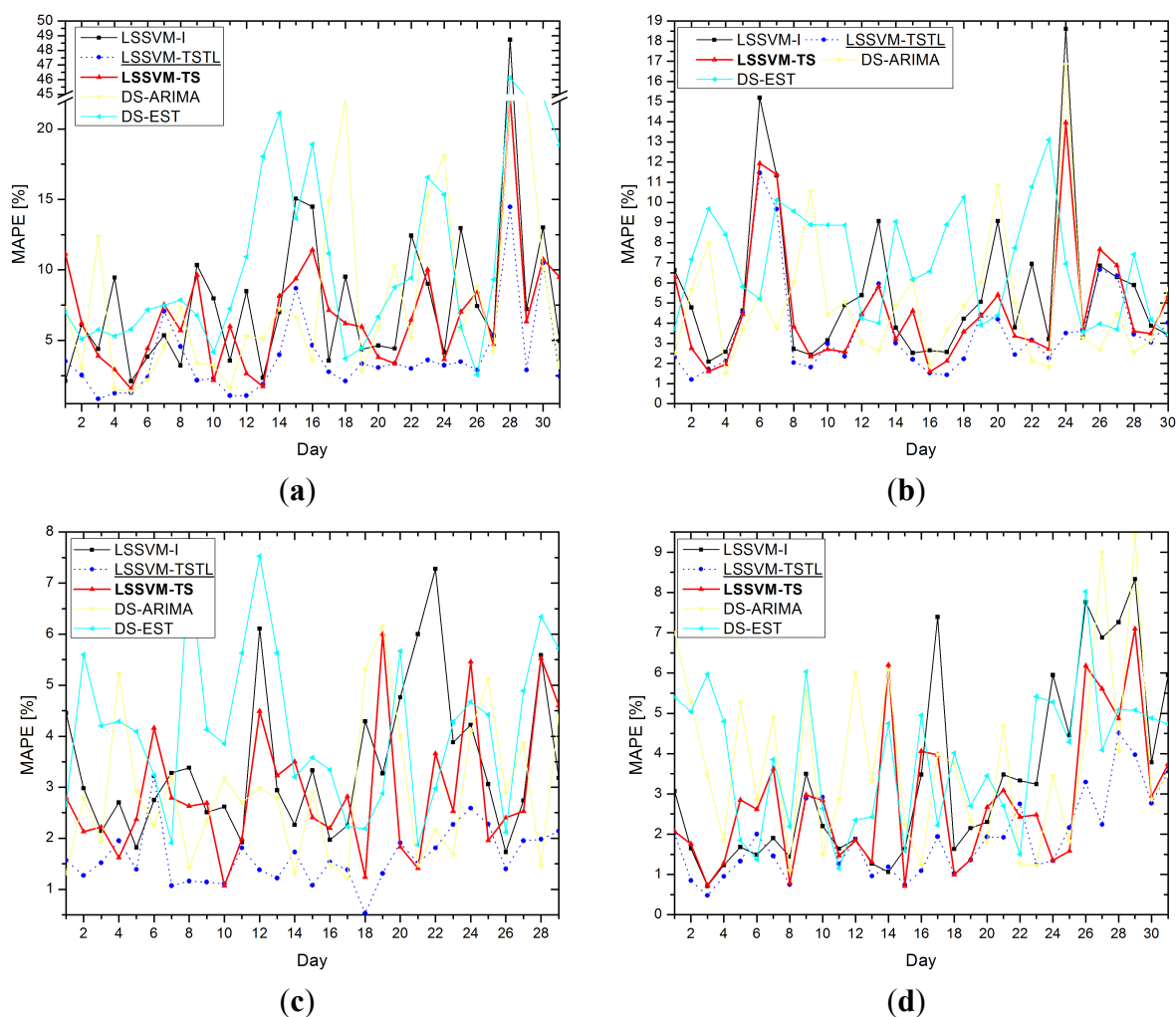


Table 3 shows the minimum, average and maximum values of MAPEs and MEs in the third to the fifth, *i.e.*, in the sixth to the eighth column, respectively, where the first column indicates the test set and the second column indicates the model. Table 3 provides a general overview of the behavior of the proposed LSSVM-TS model compared to not only the initial LSSVM-I model and LSSVM-TSTL model, but also compared to the DS-ARIMA and DS-EST models which take into account the time series trend and seasonality. The proposed LSSVM-TS model has smaller MAPE values than the LSSVM-I, DS-ARIMA and DS-EST models for all the test months. It should be noted that in Figure 7 there are days when the DS-ARIMA and DS-EST models gain better accuracy than the proposed LSSVM-TS model but on a monthly average the LSSVM-TS model is superior. The reasons why the proposed LSSVM-TS model has obtained smaller MAPEs can be found in several facts: the nonlinear mapping capabilities and structural risk minimization of LS-SVM model itself, the recurrent mechanism with superior capability to capture more data pattern information from the past load data and the indirect trend adjustment with an introduction of average daily load in the feature set. However, the proposed model prediction accuracy can be distorted because of these aforementioned facts, due to the using inaccurate prediction of the next day average load at Stage II.

Table 3. Average, max and min daily MAPEs and MEs.

Set	Model	MAPE (%)			ME (GW)		
		Min.	Avr.	Max.	Min.	Avr.	Max.
August	LSSVM-I	2.1	8.31	48.73	0.7	2.74	10.92
	LSSVM-TSTL	0.85	3.73	17.47	0.4	1.29	3.95
	LSSVM-TS	1.55	7.09	32.06	0.63	2.29	7.99
	DS-ARIMA	1.38	8.44	30.1	0.59	2.17	6.6
	DS-EST	2.55	12.22	46.14	1.06	3.22	10.23
November	LSSVM-I	2.09	5.56	18.62	0.62	1.64	5.41
	LSSVM-TSTL	1.2	3.67	11.46	0.45	1.17	3.42
	LSSVM-TS	1.59	4.69	13.96	0.44	1.5	4.25
	DS-ARIMA	1.5	4.94	16.83	0.51	1.34	4.52
	DS-EST	3.42	6.95	13.11	1.06	1.81	2.69
February	LSSVM-I	1.73	3.42	7.28	0.51	0.98	2.11
	LSSVM-TSTL	0.53	1.63	3.22	0.23	0.64	1.63
	LSSVM-TS	1.07	2.9	6	0.39	0.94	1.97
	DS-ARIMA	1.22	2.97	6.15	0.39	1	2.01
	DS-EST	1.87	4.16	7.53	0.59	1.31	2.32
May	LSSVM-I	0.71	3.35	8.33	0.26	1.01	2.93
	LSSVM-TSTL	0.48	1.89	4.51	0.24	0.67	1.83
	LSSVM-TS	0.71	2.82	7.1	0.21	0.85	2.24
	DS-ARIMA	1.22	3.71	9.44	0.12	0.96	1.74
	DS-EST	1.15	3.86	8.02	0.47	1.21	2.63

5. Conclusions

Electric load forecasting is a complex problem and electric load data present nonlinear data patterns caused by influencing factors. In order to overcome this, one approach for improving short-term load

forecasting is presented in this paper. The proposed approach is based on two LS-SVM prediction models, in two stages, where the first stage introduces a new feature, average daily load, into the second stage. The introduction of the average load into the feature set for the next day hourly load forecasting model is done with aim to examine its potential in the electric STLF. Moreover, this paper studied and revealed the influence of a new type of feature on STLF accuracy, besides the widely used calendar, climate and time-series features, and provided an efficient method for forecasting it.

Three other alternative models, LSSVM-I, DS-ARIMA and DS-EST models are used to compare the forecasting performance. The experiment results indicate that the proposed LSSVM-TS model has significant improvements among other alternatives in terms of forecasting accuracy. Furthermore, it has been shown that the quality of the proposed LSSVM-TS model directly depends on the quality of the next day average load predictions. As the experiment results have shown, by generating artificial average load samples, the accuracy of forecasting at stage II increases with an increase in the forecasting accuracy in stage I. Also, despite the usage of predicted or true value for next day average load, *i.e.*, LSSVM-TS or LSSVM-TSTL models, in both cases the generated STLF models generally performed better than the initial LSSVM-I model. Of course, usage of the exact next day average load in the STLF model input obtained the best forecasting results. However, this value is unknown and attempts should be made to obtain a value as close to the true value as possible, which would improve STLF accuracy.

Although the results are promising, further work could consider the development of a more advanced model for the prediction of average daily load for one day ahead in order to make it more accurate and thus improve STLF accuracy even more.

Acknowledgments

This work was supported by the Ministry of Science and Technological Development, Republic of Serbia (Project number: III 44006).

We would like to extend our thanks to all of the anonymous reviewers for all their helpful comments and suggestions, which have helped to improve the quality of this paper.

References

1. Soliman, S.A.; Alkandari, A.M. *Electrical Load Forecasting: Modeling and Model Construction*; Butterworth-Heinemann: Burlington, MA, USA, 2010.
2. Papalexopoulos, A.D.; Hesterberg, T.C. A regression-based approach to short-term system load forecasting. *IEEE Trans. Power Syst.* **1990**, *5*, 1535–1547.
3. Christiaanse, W.R. Short-term load forecasting using general exponential smoothing. *IEEE Trans. Power Appar. Syst.* **1971**, *PAS-90*, 900–911.
4. Vähäkylä, P.; Hakonen, E.; Léman, P. Short-term forecasting of grid load using Box-Jenkins techniques. *Int. J. Electr. Power Energy Syst.* **1980**, *2*, 29–34.
5. Irisarri, G.D.; Widergren, S.E.; Yehsakul, P.D. On-line load forecasting for energy control center application. *IEEE Power Eng. Rev.* **1982**, *PAS-101*, 71–78.
6. Mori, H.; Kobayashi, H. Optimal fuzzy inference for short-term load forecasting. *IEEE Trans. Power Syst.* **1996**, *11*, 390–396.

7. Ranaweera, D.K.; Hubele, N.F.; Karady, G.G. Fuzzy logic for short term load forecasting. *Int. J. Electr. Power Energy Syst.* **1996**, *18*, 215–222.
8. Rahman, S.; Bhatnagar, R. An expert system based algorithm for short term load forecast. *IEEE Trans. Power Syst.* **1988**, *3*, 392–399.
9. Dillon, T.S.; Sestito, S.; Leung, S. Short term load forecasting using an adaptive neural network. *Int. J. Electr. Power Energy Syst.* **1991**, *13*, 186–192.
10. Hippert, H.S.; Pedreira, C.E.; Souza, R.C. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Trans. Power Syst.* **2001**, *16*, 44–55.
11. Chen, B.-J.; Chang, M.-W.; Lin, C.-J. Load forecasting using support vector Machines: A study on EUNITE competition 2001. *IEEE Trans. Power Syst.* **2004**, *19*, 1821–1830.
12. Hong, W.-C. Electric load forecasting by support vector model. *Appl. Math. Model.* **2009**, *33*, 2444–2454.
13. Fan, S.; Chen, L. Short-term load forecasting based on an adaptive hybrid method. *IEEE Trans. Power Syst.* **2006**, *21*, 392–401.
14. Amjady, N.; Keynia, F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy* **2009**, *34*, 46–57.
15. Wang, J.; Zhu, S.; Zhang, W.; Lu, H. Combined modeling for electric load forecasting with adaptive particle swarm optimization. *Energy* **2010**, *35*, 1671–1678.
16. Hong, W.-C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy* **2011**, *36*, 5568–5578.
17. Borges, C.; Peña, Y.; Fernandez, I. Evaluating combined load forecasting in large power systems and smart grids. *IEEE Trans. Ind. Inf.* **2013**, doi:10.1109/TII.2012.2219063.
18. Taylor, J.W. Short-term load forecasting with exponentially weighted methods. *IEEE Trans. Power Syst.* **2012**, *27*, 458–464.
19. Mao, H.; Zeng, X.-J.; Leng, G.; Zhai, Y.-J.; Keane, J.A. Short-term and midterm load forecasting using a bilevel optimization model. *IEEE Trans. Power Syst.* **2009**, *24*, 1080–1090.
20. Kebriaei, H.; Araabi, B.N.; Rahimi-Kian, A. Short-term load forecasting with a new nonsymmetric penalty function. *IEEE Trans. Power Syst.* **2011**, *26*, 1817–1825.
21. Nose-Filho, K.; Lotufo, A.D.P.; Minussi, C.R. Short-term multinodal load forecasting using a modified general regression neural network. *IEEE Trans. Power Deliv.* **2011**, *26*, 2862–2869.
22. Kandil, N.; Wamkeue, R.; Saad, M.; Georges, S. An efficient approach for short term load forecasting using artificial neural networks. *Int. J. Electr. Power Energy Syst.* **2006**, *28*, 525–530.
23. Soares, L.J.; Medeiros, M.C. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *Int. J. Forecast.* **2008**, *24*, 630–644.
24. Niu, D.; Wang, Y.; Wu, D.D. Power load forecasting using support vector machine and ant colony optimization. *Expert Syst. Appl.* **2010**, *37*, 2531–2539.
25. Kelo, S.M.; Dudul, S.V. Short-term Maharashtra state electrical power load prediction with special emphasis on seasonal changes using a novel focused time lagged recurrent neural network based on time delay neural network model. *Expert Syst. Appl.* **2011**, *38*, 1554–1564.
26. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297.
27. Suykens, J.A.K.; Gestel, T.V.; Brabanter, J.D.; Moor, B.D.; Vandewalle, J. *Least Squares Support Vector Machines*; World Scientific Publishing Company: Singapore, 2002.

28. ISO New England Historical Data. Available online: http://www.iso-ne.com/markets/hst_rpts/hstRpts.do?category=Hourly (accessed on 12 April 2013).
29. Taylor, J.W.; de Menezes, L.M.; McSharry, P.E. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *Int. J. Forecast.* **2006**, *22*, 1–16.
30. Taylor, J.W. Short-term electricity demand forecasting using double seasonal exponential smoothing. *J. Oper. Res. Soc.* **2003**, *54*, 799–805.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).