Electric load forecasting using support vector machines for robust regression

Conference Paper · April 2013

5 authors, including:


Renato De Leone
University of Camerino


Erik Kropat


Silja Meyer-Nieberg
Universität der Bundeswehr München


Stefan Wolfgang Pickl
Universität der Bundeswehr München

Electric Load Forecasting using Support Vector Machines for Robust Regression

Sonia De Cosmis[1], Renato De Leone[1], Erik Kropat[2], Silja Meyer-Nieberg[2], Stefan Pickl[2]
[1] School of Science and Technologies, University of Camerino,
via Madonna delle Carceri, Camerino, Italy
[2]Fakultät für Informatik
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
85577 Neubiberg, Germany
sonia.decosmis@unicam.it, renato.deleone@unicam.it, erik.kropat@unibw.de silja.meyer-nieberg@unibw.de, stefan.pickl@unibw.de

Keywords: electricity, load forecasting, robustness, regression, support vector machines

Abstract

Load forecasting is at the core of nearly all decisions made in energy markets. The electricity load demand is influenced by numerous factors - ranging from weather conditions over seasonal effects to socio-economic influences. In this paper, we present first computational results using a linear approach supported by support vector machines for robust regression.

1. INTRODUCTION

Electric load forecasting is an important tool in the energy sector considering the deregulation of energy markets that has further increased the need for accurate forecasts [1, 5]. Load forecasting is at the core of nearly all decisions made in energy markets, and due to the high importance of accurate load forecasting different approaches have been introduced. The approaches can be classified into two main categories: models and methods which follow a more classical approach, i.e., which apply concepts stemming from time series and regression analysis and methods belonging to artificial and computational intelligence [3].

This paper is structured as follows: First, a brief overview of SVMs for robust regression is given before a new approach for robust regression is presented. After a detailed discussion of SVMs for robust regression, the application domain -load forecasting- is presented in more detail. The next section describes the application of the model to electricity load forecasting. The computational results are presented and discussed. The paper closes with conclusions and outlooks.

1.1. Support Vector Machines for Regression

Support vector machines (SVMs) [15] and support vector regression (SVR) are machine learning methods for data classification and regression recently introduced to the field of load forecasting [8, 9, 12].

Given training data

$$\left(x^1, y_1\right), \left(x^2, y_2\right), \ldots, \left(x^l, y_l\right) \qquad x^i \in \mathbb{R}^n, y_i \in \mathbb{R}$$

where l is the number of samples, we are looking for a vector $w \in \mathbb{R}^n$ and a scalar b such that the quantities $f(x^i) = w^T x^i + b$ for each $i = 1, \ldots, l$ are as close as possible to the target y_i with an allowed tolerance of ε . In addition, it is also required that the function $f(x)$ be as flat at as possible, in other words the norm of w has to be minimized. More specifically, the following standard quadratic optimization problem needs to be solved:

$$\begin{array}{ll} \min\limits_{w,b,\xi^+,\xi^-} & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(\xi_i^+ + \xi_i^-\right) \\ \text{subject to} & \left(w^T x^i + b\right) - y_i \leq \varepsilon + \xi_i^+, i = 1, \ldots, l \\ & y_i - \left(w^T x^i + b\right) \leq \varepsilon + \xi_i^-, i = 1, \ldots, l \\ & \xi^+ \geq 0, \ \xi^- \geq 0 \end{array} \tag{1}$$

The corresponding dual problem is:

$$\begin{array}{ll} \min\limits_{\alpha^+,\alpha^-} & \frac{1}{2}\left(\alpha^+ - \alpha^-\right)^T Q \left(\alpha^+ - \alpha^-\right) + \varepsilon \sum\limits_{i=1}^{l}\left(\alpha_i^+ + \alpha_i^-\right) + \\ & \sum\limits_{i=1}^{l} y_i \left(\alpha_i^+ - \alpha_i^-\right) \\ \text{subject to} & \sum\limits_{i=1}^{l}\left(\alpha_i^+ - \alpha_i^-\right) = 0 \\ & \alpha_i^+ \in [0, C], \qquad i = 1, \ldots, l \\ & \alpha_i^- \in [0, C], \qquad i = 1, \ldots, l \end{array} \tag{2}$$

where $Q_{ij} := {x^i}^T x^j$.
Applying Karush-Kuhn-Tucker [10] conditions, we can rewrite the vector w and the function f totally described by a linear combination of training patterns x^i:

$$w = \left[\sum_{i=1}^{l}\left(\alpha_i^+ - \alpha_i^-\right) x^i\right] \tag{3}$$

and the function f is:

$$f(x) = \left[\sum_{i=1}^{l}\left(\alpha_i^+ - \alpha_i^-\right) x^i\right]^T x^h + b. \tag{4}$$

Nonlinear classifiers can be obtained by applying the kernel trick [13]. After defining a map $\psi : x \in \mathbb{R}^n \mapsto \psi(x) \in F$ where F is an Hilbert space of finite or infinite dimension equipped with a scalar product $< .,. >$ in the new feature space, the training data are now $\{(\psi(x^i), y_i\}$, $i = 1, \dots, l$.

In recent years, an important line of research has focused on linear programming to solve the regression problem [11]. Considering the previous standard quadratic formulation for a regression problem (1), the main idea is to replace the 2-norm with the 1-norm [4] to have a linear function: a first possibility is using again the function $f(x) = w^T x + b$ and solve the following optimization problem:

$$\begin{array}{ll} \min\limits_{w,b,\xi^+,\xi^-} & \sum\limits_{k=1}^{n} |w_k| + C \sum\limits_{i=1}^{l} \left(\xi_i^+ + \xi_i^-\right) \\ \text{subject to} & \sum\limits_{k=1}^{n} w_k x_k^i + b - y_i \le \varepsilon + \xi_i^+, i = 1, \dots, l \\ & \sum\limits_{k=1}^{n} w_k x_k^i + b - y_i \ge -\varepsilon - \xi_i^-, i = 1, \dots, l \\ & \xi^+ \ge 0, \xi^- \ge 0 \end{array} \tag{5}$$

and an equivalent linear programming problem can be obtained through standard transformations. A second possibility is to use the support vector (SV) expansion (3) of the vector w [14] and the corresponding linear programming problem becomes:

$$\begin{array}{ll} \min\limits_{\alpha^+,\alpha^-,b,\xi^+,\xi^-} & \sum\limits_{i=1}^{l} \left(\alpha_i^+ + \alpha_i^-\right) + C \sum\limits_{i=1}^{l} \left(\xi_i^+ + \xi_i^-\right) \\ \text{subject to} & \left(\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) x^h\right)^T x^i + b - y_i \le \varepsilon + \xi_i^+, \\ & i = 1, \dots, l \\ & \left(\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) x^h\right)^T x^i + b - y_i \ge -\varepsilon - \xi_i^-, \\ & i = 1, \dots, l \\ & \alpha^+ \ge 0, \alpha^- \ge 0, \\ & \xi^+ \ge 0, \xi^- \ge 0 \end{array} \tag{6}$$

We note that only scalar products ${x^h}^T x^i$ are involved and, therefore, it is possible to move to kernel functions.

1.2. Robust Regression

The goal of a realistic classification and regression predictive model is to reduce effects of potential noise on the input data. So our aim is to obtain a regression which has good generalization properties and at the same time is not influenced by bounded perturbation. Using the perturbation on the nominal input vectors x^i such that $x^i = \bar{x}^i + g^i$ with $\left\|g^i\right\|_p \le \rho_i$,and $x^h = \bar{x}^h + g^h$ with $\left\|g^h\right\|_p \le \rho_h$ we propose now a new approach using the expansion of the vector w, (3) in terms of α_h^+ and α_h^-.

More specifically using the $2-$norm, adding a perturbation to x^i and x^h to the i^{th} constraint,

$$\left(\sum_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) x^h\right)^T x^i + b - y_i \le \varepsilon + \xi_i^+ \tag{7}$$

we obtain the following chain of equality and inequality through scalar product expansion and reduction:

$$\begin{array}{rl} & \left(\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) x^h\right)^T x^i + b - y_i \\ = & \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) (\bar{x}^h + g^h)\right]^T (\bar{x}^i + g^i) + b - y_i \\ = & \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) \bar{x}^h\right]^T \bar{x}^i + \\ & \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) g^h\right]^T g^i + \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) \bar{x}^h\right]^T g^i + \\ & \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) g^h\right]^T \bar{x}^i + b - y_i \\ \le & \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) \bar{x}^h\right]^T \bar{x}^i + \sum\limits_{h=1}^{l} \left(\alpha_h^+ + \alpha_h^-\right) \left[\left\|\bar{x}^h\right\|_2 \rho_i + \right. \\ & \left. \left\|\bar{x}^i\right\|_2 \rho_h + \rho_h \cdot \rho_i\right] + b - y_i \\ \le & \varepsilon + \xi_i^+. \end{array} \tag{8}$$

So the new linear regression robust problem becomes:

$$\begin{array}{ll} \min\limits_{\alpha^+,\alpha^-,b,\xi^+,\xi^-} & \sum\limits_{i=1}^{l} \left(\alpha^+ + \alpha^-\right) + C \sum\limits_{i=1}^{l} \left(\xi_i^+ + \xi_i^-\right) \\ \text{subject to} & \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) \bar{x}^h\right]^T \bar{x}^i + \sum\limits_{h=1}^{l} \left(\alpha_h^+ + \alpha_h^-\right) \\ & \left[\left\|\bar{x}^h\right\|_2 \rho_i + \left\|\bar{x}^i\right\|_2 \rho_h + \rho_i \rho_h\right] + b - y_i \le \varepsilon + \xi_i^+, \\ & i = 1, \dots, l \\ & y_i - \left[\sum\limits_{h=1}^{l} \left(\alpha_h^+ - \alpha_h^-\right) \bar{x}^h\right]^T \bar{x}^i - \sum\limits_{h=1}^{l} \left(\alpha_h^+ + \alpha_h^-\right) \\ & \left[\left\|\bar{x}^h\right\|_2 \rho_i \left\|\bar{x}^i\right\|_2 \rho_h + \rho_i \rho_h\right] - b \le \varepsilon + \xi_i^-, \\ & i = 1, \dots, l \\ & \alpha^+ \ge 0, \quad \alpha^- \ge 0, \\ & \xi^+ \ge 0, \quad \xi^- \ge 0. \end{array} \tag{9}$$

For the optimal solution, we note that $(\alpha_h^+ + \alpha_h^-) = |(\alpha_h^+ - \alpha_h^-)|$ and this new linear robust regression problem uses as objective function the $1-$norm of the x^i components' multipliers of the vector w and the constraints rather are expressed using the $2-$norm.

Once again using 1−norm and ∞−norm and the same approach we have that for

$$x^i = \bar{x}^i + g^i \qquad \left\|g^i\right\|_\infty \leq \eta_i \quad \left\|g^i\right\|_1 \leq \sigma_i$$

the same constraint (1) becomes:

$$\begin{aligned}
&\left[\sum_{h=1}^{l}\left(\alpha_h^+ - \alpha_h^-\right)\bar{x}^h\right]^T \bar{x}^i + \\
&\sum_{h=1}^{l}\left(\alpha_h^+ + \alpha_h^-\right)\left[\min\left\{\|\bar{x}^i\|_\infty \sigma_h, \|\bar{x}^i\|_1 \eta_h\right\} + \right. \\
&\left.\min\left\{\|\bar{x}\|_\infty^h \sigma_i, \|\bar{x}^h\|_1 \eta_h\right\} + \min\left\{\eta_i \sigma_h, \eta_i \sigma_h\right\}\right] + \\
&b - y_i \leq \varepsilon + \xi_i^+
\end{aligned} \tag{10}$$

So the final problem is now:

$$\begin{aligned}
\min_{\alpha^+,\alpha^-,b,\xi^+,\xi^-} \quad & \sum_{i=1}^{l}\left(\alpha_i^+ + \alpha_i^-\right) + C\sum_{i=1}^{l}\left(\xi_i^+ + \xi_i^-\right) \\
\text{subject to} \quad & \left[\sum_{h=1}^{l}\left(\alpha_h^+ - \alpha_h^-\right)\bar{x}^h\right]^T \bar{x}^i + \\
& \sum_{h=1}^{l}\left(\alpha_h^+ + \alpha_h^-\right)\left[\min\{\|\bar{x}^i\|_\infty \sigma_h, \|\bar{x}^i\|_1 \eta_h\} + \right. \\
& \left.\min\{\|\bar{x}\|_\infty^h \sigma_i, \|\bar{x}^h\|_1 \eta_h\} + \min\{\eta_i \sigma_h, \eta_i \sigma_h\}\right] + \\
& b - y_i \leq \varepsilon + \xi_i^+ \\
& y_i - \left[\sum_{h=1}^{l}\left(\alpha_h^+ - \alpha_h^-\right)\bar{x}^h\right]^T \bar{x}^i + \\
& \sum_{h=1}^{l}\left(\alpha_h^+ + \alpha_h^-\right)\left[\min\{\|\bar{x}^i\|_\infty \sigma_h, \|\bar{x}^i\|_1 \eta_h\} - \right. \\
& \min\{\|\bar{x}\|_\infty^h \sigma_i, \|\bar{x}^h\|_1 \eta_h\} - \\
& \left.\min\{\eta_i \sigma_h, \eta_h \sigma_i\}\right] - b \leq \varepsilon + \xi_i^- \\
& \alpha^+ \geq 0, \quad \alpha^- \geq 0, \\
& \xi^+ \geq 0, \quad \xi^- \geq 0.
\end{aligned} \tag{11}$$

1.3. Load Forecasting

The electricity load forecast is becoming an increasingly important issue in the energy sector in the last years. In fact an underestimation of the energy demand by a supplier may lead to high operational costs. The additional demand has to be met by procuring energy in the market, on the other hand an overestimation wastes scarce resources [1]. It is important to note that the day-to-day operation of the power system requires the prediction of load for a day ahead. The decision whether to undertake major structural investments or not necessitates a longer prediction horizon.

Therefore, forecasts can be distinguished mainly by the time horizon or the lead time [7]. The most commonly used is called short-term load forecasts (STLF): predict the load one day or one week ahead. However, more recently, attention is focused with a wide time horizon or in other words to the medium-term load forecast (MTLF): is used for forecasting from one week to one year. Forecasting aiming at load prediction for more than one year ahead are usually called long-term load forecasts (LTLF). The differences in lead times have consequences for the models and methods applied and for the input data necessary to built reliable tools [5, 1].

The decision maker is not only faced with the task of selecting an appropriate model type, but also with determining important external factors ranging from weather conditions over seasonal effects to socio-economic factors. All weather-dependent factors have a significant influence on the prediction, but there is a common agreement in considering the air temperature as the most important element. Medium-term load forecasts usually incorporate several additional influences, such as demographic and economic factors. In the case of long-term load forecasts, even more indicators for the demographic and economic development have to be taken into account.

2. APPLYING ROBUST SVR TO LOAD FORECASTING

This section describes the application of the new robust SVR to load forecasting. First, the data and the task are described. This is followed by the experimental set-up and a discussion of the results.

2.1. Data and Task Description

We considered data sets proposed in a forecasting competition organized by EUNITE (European Network of Intelligent Technologies for Smart Adaptive System) network, where the goal were to predict load maximum daily of electrical demand for the month of January 1999 [2]. Given information were:

- the past two-year load demand data, recorded every half hour, from January 1, 1997 to December 31, 1998;
- the previous four-year daily temperature, average from 1995 to 1998;
- the local holiday events, from 1997 to 1999.

In order to compare models, it was chosen as a measure of the error metric the minimum value of mean absolute percentage error (MAPE), given by

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{L_i - \hat{L}_i}{L_i}\right|$$

where n is the number of working days in the month of January 1999, L_i and $\hat{L}_i$ are respectively the real and the predicted value of maximum daily electrical load on the i^{th} day of the year 1999.

Our load demand data set has some important properties: first of all, climatic conditions play a central role in the load prediction, they include for example temperature, humidity, and illumination. In fact the demand has a seasonal variation that consists of high demand for electricity in the winter and low demand in the summer. A load periodicity in every week was observed, more specifically a load demand on weekends is usually lower than that of weekdays (from Monday to Friday), while electricity demand on Saturday is a little higher than that on Sunday.

Finally, local events like national holidays and festivities also effect the load demand: they may require higher or lower demand with respect to working days. Therefore, all weekends and holidays from our data set were deleted.

2.2. Experimental Setup

Let y_i be the load target value for prediction of the i^{th} day, and let x^i be the corresponding vector that includes previous target values $y_{i-1},...,y_{i-\Delta}$ as attributes. In the training phase all values y_i are known, while for future prediction $y_{i-1},...,y_{i-\Delta}$ can be the results of previous predictions. For instance if we assume $\Delta = 3$, after obtaining an approximate load of January 1, 1999, this value is used together with the loads of December 31, 1998 to forecast that of January 2, 1999. Then, we can proceed in the same way until finding an approximate load of January 31, 1999. In particular we used $\Delta = 1$ as number of back days used to forecast. Moreover, there are some parameters to choose in order to train an SVM model. They influence the performance of an SVM model, therefore, in order to get a good model they have to be determined experimentally. They are:

- cost of error C,
- the width of range of robustness ρ,
- the width of insensitive ε,
- how many previous days are included for one training data.

The aim of competition was to obtain the best possible results, therefore, we considered first of all the mean absolute percentage error (MAPE) of January 1999, to select the best parameters to minimize this error.

At the beginning of an Excel file we had the electrical loads of all the days between January 1, 1997 and January 31, 1999, recorded every half hour, i.e., 00:30 to 24:00. Considering that the load demand on holidays is lower than on working days we deleted all data of weekends from our input data and, afterwards, we deleted also spring and summer seasons. We focused our attention on the first value available of the day, particularly, we considered the load forecasting for 00:30 of each day. However, the predictions for subsequent days could be made in the same way. An important part of this work was to select the best value for the involved parameters. First of all we fixed $\varepsilon = 0$ because we used a robust regression, in other words we have a radius of uncertainty sphere around the points given by the bounded perturbation, so we do not need an insensitive tube. Secondly, we chose the parameter C between the interval $[0.1, 100]$, considering the minimum obtained value for MAPE with $\rho = 0.0001$. Similarly, we made the same considering ρ equal to 0.001 and 0.01. The minimum MAPE considered is calculated on December 1998 instead of January according with a more realistic approach for which the load of January is unknown. The model used is the new robust linear approach (11) implemented in for Matlab.

2.3. Computational Results

Two different approaches were applied: first of all we took into account all working days available from January 1997 to January 1999, secondly we used only "winter" seasons. In the first approach, three different training sets were considered: the first contained all the days of the years 1997 and 1998 excluding national holidays and weekends (in total 499 days), the second ultilized around 75% of the available data (395 days), and finally the third consisted of only 300 days.

The following Tables 1, 2, and 3 show some of best values of the parameter C obtained for each value of ρ at the time 00:30 on December 1998 (MAPE December), and after that the corresponding values of the MAPE on the total available days (MAPE total) and on January 1999 (MAPE January). Figures 1, 2, and 3 show a comparison between the estimates and real electricity loads. First, the graphs for all working days considered in 1997, 1998, and January 1999 are shown, followed by the differences between the estimate and real values for the 19 working days in January 1999 and corresponding percentage error. Three cases are presented:

- Case 1: number of training element 499. The results are reported in Table 1 and in Fig. 1,
- Case 2: number of training element 395. The results are reported in Table 2 and in Fig. 2,
- Case 3: number of training element 300. The results are reported in Table 3 and in Fig. 3.

In the second approach, we deleted national holidays and weekends, also spring and summer seasons, from March 21 to September 21 of each year, to be more coherent with the objective: to forecast the load of the month of January 1999

Table 1. Best parameters C with different values for ρ using 499 days

ρ	C	MAPE total	MAPE 98 December 98	MAPE January 99
0.0001	8.5	3.5767	3.1948	2.7032
0.0001	0.9	3.8106	3.3328	2.6715
0.001	1.5	3.8316	3.7946	3.1970
0.001	0.7	3.8408	3.6912	3.0493
0.01	0.1	4.6820	4.6546	3.3016

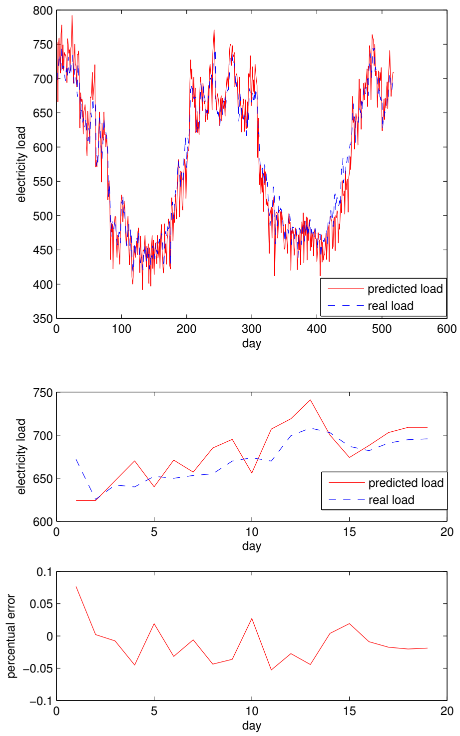

Figure 1. Load and forecasted load (total period and January) at 00:30 with $\rho = 0.0001, C = 0.9$

at 00:30 trying to minimize the MAPE. In the following table, for each best found parameters C and ρ the corresponding value of MAPE for the total period (252 days) is given, as well as the MAPE for December 1998 and the MAPE for January 1999. Figures 4, 5, and 6 show plots of predicted values as well the real load demand, either on the entire period or only for January 1999 with the relative percentage error for three different values of robustness ρ.

3. CONCLUSIONS AND OUTLOOK

Load forecasting is a very important tool in decision processes not only in the electricity sector, but also for energy transaction and energy operative decisions. In this paper an

Table 2. Best parameters C with different values for ρ using 395 days

ρ	C	MAPE total	MAPE December 98	MAPE January 99
0.0001	10	3.5754	3.1839	2.6351
0.0001	6.5	3.6368	3.1735	2.5403
0.0001	0.9	3.8639	3.2925	2.5269
0.001	1.5	3.8260	3.6479	2.9538
0.001	0.8	3.8548	3.5949	2.8694
0.01	0.1	4.7192	4.4371	3.0895

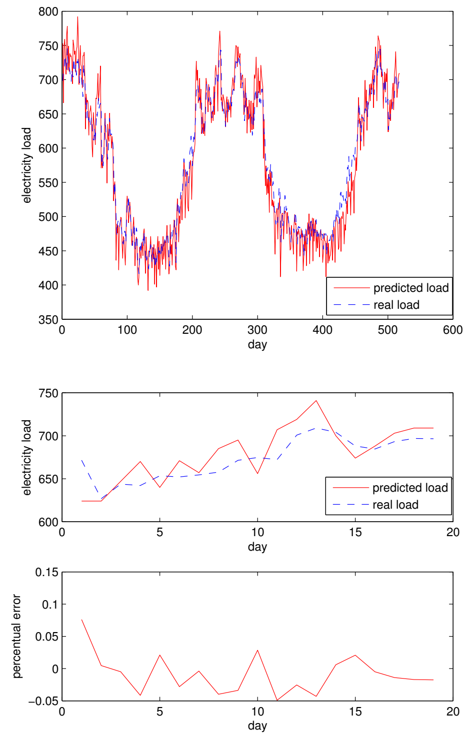

Figure 2. Load and forecasted load (total period and January) at 00:30 with $\rho = 0.0001, C = 0.9$

overview over support vector machines for robust regression that are a relatively new method for load forecasting was given. We considered data sets proposed in a forecasting competition organized by EUNITE (European Network of Intelligent Technologies for Smart Adaptive System) network, where the goal was to predict the daily maximum load of electricity for January 1999 [2].

We applied a robust model using support vector machines to forecast half hourly loads of January 1999, using data from 1997 and 1998. The paper presented exemplary the results for the first available half hour, i.e., 00:30. The error measure

Table 3. Best parameters C with different values for ρ using 300 days

ρ	C	MAPE total	MAPE December 98	MAPE January 99
0.0001	16	3.5687	3.0602	2.4709
0.0001	0.9	3.9590	3.2878	2.4256
0.0001	0.4	4.2198	3.3378	2.3712
0.0001	8.5	3.6718	3.1015	2.4292
0.001	10	3.8495	3.9254	3.3414
0.001	2.5	3.7881	3.4596	2.7383
0.01	0.1	4.9055	3.7966	2.5431

Table 4. Best parameters C with different values for ρ without spring and summer

ρ	C	MAPE total	MAPE December 98	MAPE January 99
0.0001	20	3.4996	3.0151	2.8359
0.0001	16	3.5214	2.9851	2.7818
0.0001	8.9	3.5889	3.0145	2.6979
0.001	10	4.4131	4.3114	3.9445
0.001	0.4	4.1037	3.1721	2.3794
0.01	0.2	4.4721	3.3330	2.1544

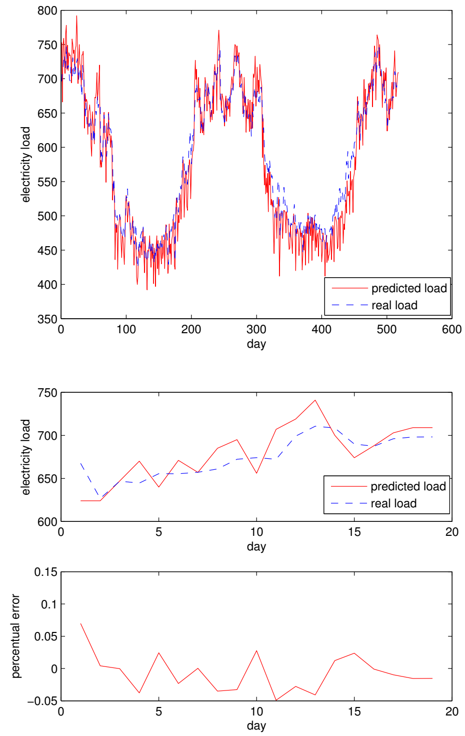

Figure 3. Load and forecasted load (total period and January) at 00:30 with $\rho = 0.0001, C = 0.9$

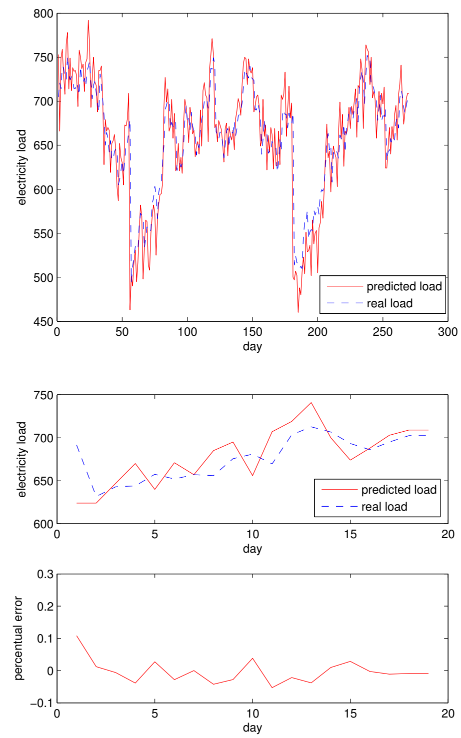

Figure 4. Load and forecasted load (total period and January) at 00:30 with $\rho = 0.0001, C = 8.9$

most used in literature to evaluate models is the mean absolute percentage error (MAPE). Therefore, it was also used in the present paper.

We observed that the load has some seasonal variation implying a great influence of climate conditions. For this reason, different approaches were applied: one on working days without weekends and holidays, and others for winter seasons: from January to March, and from September to December. Among different proposed solutions, we can conclude that selecting appropriate data sets increases the model performance. More specifically, considering only "winter" seasons, i.e. winter and autumn seasons, better forecasts that means a lower MAPE are obtained.

In most short-time forecasting (STLF) works [8, 9, 12], meteorological information which include temperature, humidity, illumination, etc., has been used to predict the load demand. However the prediction of a 30 days period during which the temperature does not change very much, trying to predict the temperature and incorporate it into the model is not more advantageous and useful [2]. Next steps will consider also daily temperature data set, and moreover will we

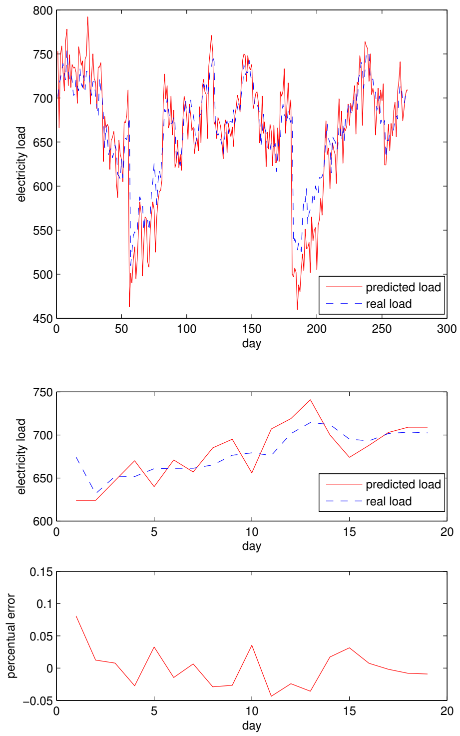

Figure 5. Load and forecasted load (total period and January) at 00:30 with $\rho = 0.001, C = 0.4$

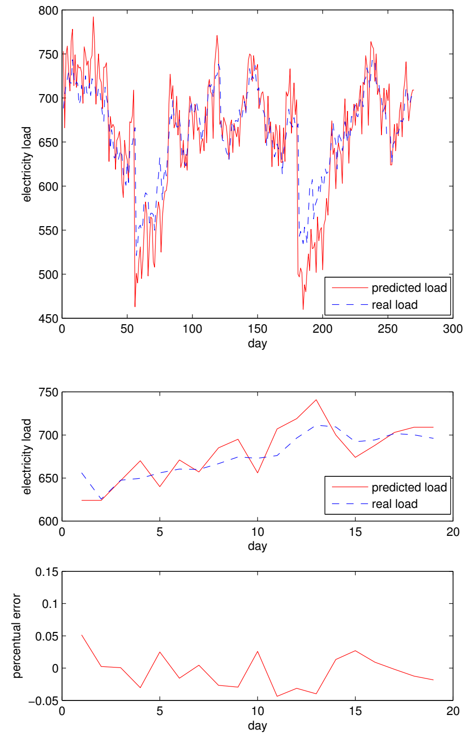

Figure 6. Load and forecasted load (total period and January) at 00:30 with $\rho = 0.01, C = 0.2$

work on average daily and maximum daily load to forecast January 1999.

REFERENCES

[1] Atzberger P. J., Kramer P. R., and Peskin C. S.: A stochastic immersed boundary method for fluid-structure dynamics at microscopic length scales. Journal of Computational Physics, **224**, pp. 1255-1292, June 10, 2007.

[2] 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., **19**, NO. 4, pp. 1821-1830, Nov. 2004.

[3] De Cosmis S., De Leone R., Kropat E., Meyer-Nieberg S., Pickl S., Rosati J.: Electric Load Forecasting: An Experimental Comparison. In CTW2012 Proceedings, May **29-31**, 2012, pp. 100-103.

[4] De Leone R.: Support Vector Regression for Time Series Analysis. In Operations Research Proceeding 2010, B. Hu et al., editors, pp.33-38, Springer 2011.

[5] Farnell L., Gibson W. G.: Monte Carlo simulation of diffusion in a spatially nonhomogeneous medium: a biased random walk on an asymmetrical lattice. Journal of Computational Physics, **208**, pp. 253-265, September 1, 2005.

[6] Gilbert R.C., Trafalis T. B.: Robust classification and regression using support vector machines. European Journal of Operational Research, **173**, 2006

[7] Hahn H., Meyer-Nieberg S., Pickl S.: Electrical load forecasting methods: tools for decision making. European Journal of Operational Research, **199**, pp. 902-907, 2009.

[8] Hippert H. S., Pedreira C. E., Souza R. C.: Neural networks for short-term load forecasting: A review and evaluation. IEEE Trans. Power Syst., **16**, pp.44-55, Feb.2001.

[9] Jabbour K., Riveros J. F. V., Landsbergen D., Meyer W. : Alfa: Automated load forecasting assistant.In IEEE Trans. Power Syst., **3**, pp. 908-914, Aug. 1988.

[10] Mangasarian O.L.: Nonlinear Programming. McGraw-Hill, New York, 1969.

[11] Montgomery D. C., Peck E. A.: Introduction to Linear Regression Analysis. Wiley, 1982, New York.

[12] Murto P.: Neural network models for short-term load forecasting. Master thesis, Dept. Eng. Phys. Math., Helsinki Univ. Technology, Helsinki, Finland, 1998.

[13] Schölkopf B., Smola A. J.: Learning with kernels. Ph.D. thesis, Technical University Berlin, 1998.

[14] Stitson M., Gammerman A., Vapnik V., Vovk V., Watkins C., Weston J.: Advances in Kernel Methods Support Vector Learning. Cambridge, MA, 1999. MIT Press.

[15] Vapnik V.: The Nature of Statistical Learning Theory. Springer, New York, 1995.