Abstract—In this paper, we present a new approach for very short term electricity load demand forecasting. In particular, we apply support vector regression to predict the load demand every 5 minutes based on historical data from the Australian electricity operator NEMMCO for 2006-2008. The results show that support vector regression is a very promising approach, outperforming backpropagation neural networks, which is the most popular prediction model used by both industry forecasters and researchers. However, it is interesting to note that support vector regression gives similar results to the simpler linear regression and least means squares models. We also discuss the performance of four different feature sets with these prediction models and the application of a correlation-based sub-set feature selection method.

I. INTRODUCTION

ElectrIcItY load forecasting is an important component for any modern energy management system. Electricity market operators and participants use load forecasting for many reasons such as to make unit commitment decisions, reduce spinning reserve capacity and schedule device maintenance plan properly to ultimately safeguard system security and achieve the highest possible reliability of supply while optimising economic cost. For instance, an over-prediction of load to meet security requirements could involve the start up of too many units resulting in an unnecessary increase in reserve, and hence operating costs, as market operators must allocate such demand [1-2].

Load forecasting can serve different planning objectives and assist system operators to meet all critical conditions in the short, medium and long term. It can be classified into four types with respect to the future time window of the forecasting task:

1) Long-term - Typically 1 to 10 years; used to identify needs for major generation planning and investment, since large power plants may take a decade to become available due to the challenging project requirements and the needs to design, finance and build them;

2) Medium-term - Typically between several months to a year; used to ensure that security and capacity constraints are met in the medium term;

3) Short-term (STLF) - A day ahead; used to assist planning and market participants;

4) Very short-term (VSTLF) - hours and minutes ahead; used to assist trading and eventually dispatch.

In this paper we focus on VSTLF. In Australia, the VSTLF plays an important role in the operation of the national electricity market, as its market operator NEMMCO must issue every 5 minutes the production schedule of the generators [3-4]. The 5-minute regional demand forecasting relates the demand at the beginning of a dispatch interval to the target at the end [4]. Based on this load forecasting, generators and network operators are required to notify NEMMCO of their maximum supply capacity and availability, and this information is matched against regional demand forecasts. This then enables the remainder of market participants to respond to potential supply shortfalls by increasing their generation or network capacity to meet expected market demand. All offers to supply (bids) are then collated so that potential shortfalls of supply against expected demand can be identified and published. Participants in the market use this information as the basis for any re-bids of the capacity they wish to bring to the market.

In this paper, we deal with the 5-minute VSTLF using data from 2006, 2007 and 2008 provided by NEMMCO. Our main objective is to study VSTLF in power system since there have been only a few published articles [5-12]. We then present an application of support vector regression (SVR) for 5-minute electricity demand forecasting. SVR has been successfully used for STLF but has not been reported in the open technical literature for VSTLF. The performance of SVR is compared with standard statistical methods such as linear regression (LR) and least means squares (LMS) and also with multilayer perceptron trained with the backpropagation neural network (BPNN) algorithm. The latter is a very popular model for electricity load forecasting, predominantly reported in the research literature and it is also the main method used by industry forecasters. We also compare several feature sets and study the effect of a correlation-based feature selector.

The paper is organised as follows. Section II discusses the characteristics of VSTLF and previous research on VSTLF. Section III presents SVR algorithm and the other prediction algorithms used for comparison. Section IV introduces the modelling process including feature sets, evaluation procedure and performance measures. The results are presented and discussed in section V before concluding in Section VI.
II. VERY SHORT TERM ELECTRICITY LOAD FORECAST

A. Characteristics of Electricity Load

Time, random effects and anomalous days are some of the
main factors influencing the VSTLF.

Time: Electricity demand during the day differs from the
demand during the night, and demand during weekdays
differs from the demand during weekends. However, all
these differences have cyclic nature, as the electricity
demand on the same weekday and time but on a different
date is likely to have the same value. Any shift to and from
daylight saving time and the start of school year also
changes the previous load profiles.

Random effects: A power system is continuously
subjected to random disturbances and transient phenomena.
In addition to a large number of very small disturbances,
there are large load variations caused by devices such as
steel mills, synchrotrons and wind tunnels, as their hours of
operation are usually unknown to utility dispatchers. There
are also certain events such as widespread strikers,
shutdown of industrial facilities and special television
programs whose occurrences are not known a priori but
affect the load.

Irregular days: These days include public holidays,
consecutive holidays, and days preceding and following the
holidays, days with extreme weather or sudden weather
change and special event days [14].

B. Previous Work on VSTLF

Table 1 lists the forecasting methods previously used for
STLF and VSTLF. The majority of work has been on STLF.

### Table 1. Summary of Load Forecasting Techniques Used

<table>
<thead>
<tr>
<th>Method</th>
<th>STLF</th>
<th>VSTLF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autoregressive integrated moving average</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Autoregressive moving average</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>General exponential smoothing</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Kalman filter</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multiple linear regression</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>State space</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Stochastic time series</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Support vector regression</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Artificial intelligence - based methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artificial neural networks</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Expert system</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fuzzy inference</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fuzzy neural system</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

The most popular prediction models used for VSTLF are
based on neural networks. In [5] a set of BPNN was used
where each network forecasts the load for a particular time
lead and for a certain period of the day. Single BPNNs
showed good accuracy in [9] and [11]. A hierarchical neural
network algorithm was used in [12] for 15-minutes load
forecasting. A combination of several BPNNs was shown to
outperform LR in [7].

In [10] an approach combining fuzzy logic with chaos
theory was applied for 15 minutes ahead forecasting. It uses
data from the previous two weeks to predict the load in the
current week and was shown to compare favourably with a
neural network based approach in terms of accuracy. In [6]
three approaches were compared (fuzzy logic, BPNN and
autoregressive model) and the first two were found to be
more accurate than the third one. A self-organizing fuzzy
neural network was used in [8] and shown to outperform the
standard BPNN and fuzzy neural network.

C. Requirement and Difficulties in VSTLF

A good VSTLF system should fulfil the requirement of
high accuracy and speed. However, there are several
challenges. First, it is not clear how to select the best
prediction algorithm and the best feature set. Good feature
selection is the key to the success of a prediction algorithm.
It is needed to reduce the number of features by selecting the
most informative and discarding the irrelevant features.
Second, overfitting is a common problem in load prediction,
especially for the predominantly used neural network-based
prediction algorithms. It means that the error on the training
data (the historical data used to build the prediction model)
is low but the error on the new data is high. Thirdly, the
neural network algorithms have many parameters that
require manual tuning and greatly influence their
performance.

III. SVR AND OTHER PREDICTION ALGORITHMS

A. Support Vector Regression

SVR is an extension of the support vector machine
(SVM) [19] algorithm for numeric prediction. SVM is a
state-of-the-art machine learning algorithm that is applicable
to classification tasks. Using the training data, it finds the
maximum margin hyperplane between two classes by
applying an optimization method. The decision boundary is
defined by a subset of the training data, called support
vectors. By using a kernel function, nonlinear decision
boundaries can be formed while keeping the computational
complexity low.

SVR also produces a decision boundary that can be
expressed in terms of a few support vectors and can be used
with kernel functions to create complex nonlinear decision
boundaries. Similarly to linear regression, SVR tries to find
a function that best fits the training data. In contrast to LR,
SVR defines a tube around the regression line using a user-
specified parameter $\varepsilon$ where the errors are ignored and it
also tries to maximize the flatness of the line (in addition to
minimizing the error) [17].

SVM and SVR offer several advantages. Similarly to
network based algorithms they are capable of forming
complex decision boundaries. However, unlike neural
networks they do not overfit the training data as the decision
boundary depends only on a few training instances. Also,
they have a much smaller number of parameters to optimise.

The main parameters in SVR are $\varepsilon$ and $C$. As mentioned
above, $\varepsilon$ defines the error-insensitive tube around
the regression function, and thus controls how well the function
fits the training data [16, 17]. The parameter $C$ controls the
trade off between training error and model complexity; a
smaller $C$ increases the number of training errors; a larger $C$ increases the penalty for training errors and results in a behavior similar to that of a hard-margin SVM [21].

We used WEKA’s SVR implementation which is based on a sequential optimization algorithm [16] with the following parameters, which were set empirically: polynomial kernel, $\varepsilon=0.001$ and $C=1$.

SVR is yet to be tested for VSTLF. It has been found very efficient and accurate for STLF and shown to outperform autoregressive and BPNN algorithms [14]. The winners of the 2001 EUNITE (European Network on Intelligent Technologies) competition on electricity load forecasting also used SVR for STLF [15].

B. Other Prediction Algorithms Used for Comparison

The performance of SVR was compared with 3 prediction algorithms: LR, LMS and BPNN. LR is a standard statistical method; LMS is a modification of LMS; BPNN is the most popular algorithm for VSTLF prediction as discussed above. Table 2 summarizes the algorithms and parameters used. We used their WEKA's implementations [16].

<table>
<thead>
<tr>
<th>Prediction algorithm</th>
<th>Description and parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression (LR)</td>
<td>Approximates the relationship between the outcome and independent variables with a straight line. The least squares method is used to find the line which best fits the data. We used the M5's method for variable selection - starting with all variables at each step the weakest variable is removed based on its standardized coefficient until no improvement is observed in the error estimate.</td>
</tr>
<tr>
<td>Least mean squares (LMS)</td>
<td>A modification of LR using least squared regression functions generated from random subsamples of the data. The final prediction model is the least squared regression with the lowest median squared error. We used sample size=4.</td>
</tr>
<tr>
<td>Back-propagation Neural Network (BPNN)</td>
<td>A classical neural network trained with the back-propagation algorithm. We used 1 hidden layer with the standard heuristic for setting the number of neurons in it (average of the number of input and output neurons), sigmoid transfer functions, learning rate=0.3, momentum=0.2. The training stopped when a maximum number of epochs was reached (500).</td>
</tr>
</tbody>
</table>

IV. MODELLING

As stated already, we aim to predict the electricity load every 5 minutes for August 2008 based on the historical data for the period of 2006 and 2007 in the NSW region of Australia. In this section we describe the data, the extracted features for use with the prediction algorithms, the evaluation procedure and performance measures.

A. Historical Data

We have obtained data for the electricity load in NSW in 2006, 2007 and 2008. To predict the load for August 2008, we used the data from August 2006 and 2007 as training data, as discussed in the following sections.

B. Variable Selection

Variable selection plays a critical role in building a good forecasting model. It is important to first analyse the data to ensure all essential variables are included.

Fig. 1 shows a typical daily load graph. As it can be seen, there is a daily pattern of decreases and increases with two peaks around 8:00 and 17:00 and 2 lows around 1:00 and 16:00. More specifically, the load is low from 0:00 to 5:00; it increases from 5:00 to 8:00 and decreases from 8:00 to 17:00 and then gradually decreases till 1:00. The load during the weekend is lower than the load during the workdays.

Fig. 2 plots the load differences at 5-minute intervals for 2 consecutive days (Monday and Tuesday). It confirms that there are daily patterns in the electricity demand. These difference values can be used instead of the actual values or in addition to them.
Fig. 3 shows that there are seasonal differences in the electricity load in NSW. As expected, the loads for spring, summer and autumn are similar, while the load for winter is different and significantly higher. However, it has been shown that the weather (e.g. temperature) does not influence the accuracy of VSTLF [13] as it changes slowly within a 5-minute forecasting time window.

C. Feature Selection

Based on the results from the previous section, four feature sets were selected. They are summarised in Table 3.

To predict the load X_{t+1} at time t+1, feature set 1 uses the previous 5 loads from the same day X_{t},...,X_{t+4}, the load at the same time and weekday the previous week XX_{t},... XX_{t+4} and also the previous 5 loads XX_{t},... XX_{t+4}. Thus, feature set 1 takes into account the load similarity in the previous 25 minutes and the weekly load pattern.

Feature sets 2, 3 and 4 are supersets of feature set 1. Feature set 2 also takes into account the daily load pattern by encoding 4 daily intervals based on the pattern from Fig.1. Feature set 3 tries to exploit more the weekly load pattern by adding a feature that encodes the weekday (Monday, Tuesday, etc) for the forecast day. Feature set 4 tries to take advantage of the daily pattern by adding features based on more fine grained information about the time for which the forecast is made. It also adds the 5-minute load differences for the features from set 1.

Example: Predicting the load for 10:35 Monday based on the 11 features from feature set 1 and also using 4 variables representing the time interval for 10:35 (i.e. their values will be 0, 1, 1, 0).

Example: Predicting the load at 10:30 Monday based on the 11 features from feature set 1 and also using 4 variables representing the weekday for Monday (i.e. their values will be 1, 0, 0, 0).

Example: Predicting the load at 10:30 Monday based on the 11 features from feature set 1 and also using 4 variables representing the time interval for 10:30 (i.e. their values will be 0, 1, 0, 0).

D. Evaluation Procedure

The quality of each prediction algorithm was evaluated in two ways: using cross validation and by testing the prediction algorithm on new data.

We firstly used 10-fold cross validation on the data set from August 2006 and August 2007 (17,856 instances). 10-fold cross validation is the standard procedure used in machine learning [17]. The main idea is to use independent testing set for performance evaluation rather than the training data. This allows obtaining a more reliable estimate of the prediction performance. In 10-fold cross validation, the dataset is divided into 10 subsets of approximately equal size. Ten experiments are conducted: each time k-1 subsets are put together to form the training set which is used to build the model. The remaining test set is used to test the model calculating the performance measure (e.g. a mean performance measure) in the testing set.
absolute error). In each of the 10 runs, a different test set is used. The overall performance measure for the model is calculated as the average across the 10 runs.

We also evaluated the quality of the models on new data, i.e. the data from August 2006 and 2007 was used to build the models and then they were tested on the data from August 2008.

E. Performance Measures

We used the following performance measures:

1) Mean absolute error (MAE):

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| L_{\text{actual}} - L_{\text{forecast}} \right|
\]

where \(L_{\text{actual}}\) and \(L_{\text{forecast}}\) are the actual and forecasted electricity loads, respectively, at the 5-minute interval \(i\) and \(n\) is the total number loads that are being predicted.

2) Relative absolute error (RAE) – It measures the performance of the prediction model used, relative to a very simple predictor, the ZeroR algorithm. ZeroR predicts the mean value in the training data and is used as a baseline.

3) Mean absolute performance error (MAPE), an extension of MAE defined as

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{L_{\text{actual}} - L_{\text{forecast}}}{L_{\text{actual}}} \right|
\]

4) KPI - a performance improvement indicator which can be used to compare the MAPE performance of the predictor relative to no prediction (MAPE_{naive}):

\[
\text{KPI} = \frac{\text{MAPE}_{naive} - \text{MAPE}}{\text{MAPE}_{naive}}
\]

\[
\text{MAPE}_{naive} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{L_{\text{actual}} - L_{\text{actual}}}{L_{\text{actual}}} \right|
\]

An acceptable target for KPI would be 20%.

All results in Section V are presented in % using these performance measures.

V. RESULTS AND DISCUSSION

A. Performance Evaluation Using Cross Validation

Table 4 summarises the performance of the four prediction algorithms using the four feature sets and 10-fold cross validation as an evaluation method. Three of the algorithms (LR, LMS and SVR) performed similarly obtaining the best MAE (MAE=0.41%, i.e. 41 MW). BPNN was slightly worse with MAE=0.51-0.53%. LR was the fastest to build the model, followed by LMS, BPNN and SVR. In particular, it took approximately two hours to build SVR using a computer with a 2 GHz processor and 1.49 GB of RAM. In [20] SVR was found to be faster than BPNN but this may be due to the different stopping criterion used by BPNN and the smaller amount of training data.

A comparison across the feature sets shows that there was no difference in accuracy. Thus, the additional features such as the time of the day and week, when used with the classifiers we studied, did not improve the prediction accuracy. Feature set 1 is the smallest (11 features) while feature set 4 is the biggest (22 features). Naturally, the bigger the feature set, the longer the training time (this is not reflected in Table 4 for LR and LMS as these algorithms perform additional sub-set feature selection as mentioned in Table 2). All prediction models improved over the baseline (ZeroR) as indicated by RAE.

Table 4. 10-Fold Cross Validation Performance Using Data from August 2006 and 2007

<table>
<thead>
<tr>
<th>Feature set 1</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time [seconds]</td>
<td>1.13</td>
<td>123.88</td>
<td>132.31</td>
<td>7712.42</td>
</tr>
<tr>
<td>MAE [%]</td>
<td>0.41</td>
<td>0.41</td>
<td>0.51</td>
<td>0.41</td>
</tr>
<tr>
<td>RAE [%]</td>
<td>4.32</td>
<td>4.33</td>
<td>5.33</td>
<td>4.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature set 2</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time [seconds]</td>
<td>1.05</td>
<td>240.39</td>
<td>105.99</td>
<td>7717.06</td>
</tr>
<tr>
<td>MAE [%]</td>
<td>0.41</td>
<td>0.41</td>
<td>0.51</td>
<td>0.41</td>
</tr>
<tr>
<td>RAE [%]</td>
<td>4.31</td>
<td>4.32</td>
<td>5.33</td>
<td>4.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature set 3</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (seconds)</td>
<td>0.45</td>
<td>109.72</td>
<td>410.14</td>
<td>7724.2</td>
</tr>
<tr>
<td>MAE [%]</td>
<td>0.41</td>
<td>0.41</td>
<td>0.53</td>
<td>0.41</td>
</tr>
<tr>
<td>RAE [%]</td>
<td>4.32</td>
<td>4.33</td>
<td>5.54</td>
<td>4.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature set 4</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (seconds)</td>
<td>1.09</td>
<td>266.86</td>
<td>455.61</td>
<td>7854.66</td>
</tr>
<tr>
<td>MAE [%]</td>
<td>0.41</td>
<td>0.41</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>RAE [%]</td>
<td>4.32</td>
<td>4.33</td>
<td>4.53</td>
<td>4.32</td>
</tr>
</tbody>
</table>

B. Performance Evaluation on New Data

Table 5 shows the accuracy when the models were used to predict new data (i.e. they were trained on data for August 2006-2007 and used to predict the load for August 2008).

Table 5. Performance in Predicting the Load for August 2008 Using Data from August 2006 and 2007

<table>
<thead>
<tr>
<th>Feature set 1</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE [%]</td>
<td>0.31</td>
<td>0.32</td>
<td>0.41</td>
<td>0.31</td>
</tr>
<tr>
<td>MAPE_{naive} [%]</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>KPI [%]</td>
<td>46.11</td>
<td>44.52</td>
<td>30.06</td>
<td>46.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature set 2</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE [%]</td>
<td>0.45</td>
<td>0.39</td>
<td>0.77</td>
<td>0.43</td>
</tr>
<tr>
<td>MAPE_{naive} [%]</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>KPI [%]</td>
<td>22.58</td>
<td>32.40</td>
<td>-31.99</td>
<td>26.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature set 3</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE [%]</td>
<td>0.31</td>
<td>0.32</td>
<td>0.73</td>
<td>0.31</td>
</tr>
<tr>
<td>MAPE [%]</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>KPI [%]</td>
<td>46.01</td>
<td>44.95</td>
<td>-25.72</td>
<td>45.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature set 4</th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE [%]</td>
<td>0.31</td>
<td>0.31</td>
<td>0.40</td>
<td>0.31</td>
</tr>
<tr>
<td>MAPE_{naive} [%]</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>KPI [%]</td>
<td>46.03</td>
<td>46.36</td>
<td>31.20</td>
<td>46.14</td>
</tr>
</tbody>
</table>
The results are consistent with the cross-validation results in Table 4. SVR, LR and LMS perform best obtaining the lowest MAPE. In particular, the best performing algorithms are: on feature set 1 - LR and SVM, MAPE=0.31%; on feature set 2: LMS, MAPE=0.39%; on feature set 3 – LR and SVR, MAPE=0.31%; on feature set 4: LR, LMS and SVR, MAPE=0.31%. Thus, the best result is MAPE=0.31% achieved by LR, LMS and SVR using one or more feature sets. BPNN is slightly worse; its MAPE is between 0.4% using feature set 4 and 0.77% with feature set 2.

A MAPE of 0.3-0.4% is a very good result. Fig. 4 shows the actual and predicted values for MAPE=0.33%. The prediction model is SVR with feature set 1 for the 7th of August 2008. As it can be seen, there is almost no difference between the two graphs.

A comparison across the feature sets shows that feature sets 1, 3 and 4 perform similarly while feature set 2 is slightly worse. Recall that feature set 1 is the smallest and the remaining feature sets are supersets of it. Thus, adding additional features do not seem to improve performance.

In terms of the KPI, the 20% target is met by all models for feature set 1 and 4, by LMS for feature set 2 and by all models except BPNN for feature set 3. Even though BPNN with feature set 1 and 4 meets the target, its performance is 14-16% lower than the other algorithms.

To summarize, feature set 1 in combination with SVR, LR and LMS produces the best results and significantly outperforms BPNN. The KPI accuracy results of the three models are 25-26% higher than the KPI target.

The good performance of LR and LSM indicates that there is a linear relationship between the independent variables (features used for prediction) and the dependent variable (the load that we are predicting). Fig. 5 shows scatter plots of $X_t$, $X_{t-1}$, $XX_{t+1}$ and $XX_t$ versus the dependent variable $X_{t+1}$ confirming the strong linear relationships.

We also investigated if the number of features in the best performing feature set (set 1) can be further reduced without deteriorating the prediction accuracy. We applied a simple, fast and efficient method for feature sub-set selection, called correlation-based feature selection (CFS) [18]. It searches for the "best" sub-set of features where "best" is defined by a heuristic which takes into consideration 2 criteria: 1) how
good the individual features are at predicting the outcome and 2) how much the individual features correlate with each other. Good feature subsets contain features that are highly correlated with the outcome and uncorrelated with each other.

To evaluate CFS, we used the data for 2006 and 2007. Table 6 shows the selected features using CFS. As it can be seen, 1 or 2 features were selected for each month, which is a drastic feature reduction from the original 11 features. The most important features for predicting the load $X_{t+1}$ are $X_t$, $XX_{t+1}$, $XX_t$, and $XX_{t-1}$, with $X_t$ (the load 5 minutes before) being always selected. There seem also to be a seasonal pattern that requires further investigation. Table 7 summarizes the performance of the prediction algorithms using the selected features and 10-fold cross validation as an evaluation procedure. A comparison with Table 4 shows that the accuracy results are worse - there is an increase in MAE of 0.14% for LR and LMS, 0.27% for BPNN and 0.12% for SVR. However, the time to build the prediction models is reduced, especially for BPNN and SVR.

**Table 6. Selected variables by CFS applied to feature set 1**

<table>
<thead>
<tr>
<th>Month</th>
<th>$X_t$</th>
<th>$XX_{t+1}$</th>
<th>$XX_t$</th>
<th>$XX_{t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>February</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>April</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>June</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>August</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>September</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>November</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>December</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 7. 10-fold cross validation performance using the features selected by CFS**

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>LMS</th>
<th>BPNN</th>
<th>SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time [s]</td>
<td>0.59</td>
<td>74.5</td>
<td>17.6</td>
<td>0.97</td>
</tr>
<tr>
<td>MAE [%]</td>
<td>0.55</td>
<td>0.55</td>
<td>0.78</td>
<td>0.52</td>
</tr>
<tr>
<td>RAE [%]</td>
<td>5.79</td>
<td>5.77</td>
<td>8.26</td>
<td>5.51</td>
</tr>
</tbody>
</table>

**VI. CONCLUSION**

We presented a new approach for the 5-minute electricity load demand forecasting using SVR. The results showed that SVR is a very promising prediction model, outperforming the BPNN prediction algorithms which is widely used by both industry forecasters and researchers. However, we found that simpler models such as LR and LMS produced similar accuracy results and were faster to train than SVR. We investigated the performance of four feature sets based on historical load data and information about the time and day of the week. The best performing feature set uses only historical load data, namely the load 25-30 minutes before the forecast time on the forecast day and the load on the same weekday of the previous week. Correlation-based feature sub-set selection applied to this feature set was able to further reduce the number of features, and thus decrease the time to build the model, but it also reduced the prediction accuracy.

**ACKNOWLEDGMENT**

We would like to thank NEMMCO for providing the data and all the information needed to conduct the research.

**REFERENCES**