This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines

Ervin Ceperic, Vladimir Ceperic, *Student Member, IEEE*, and Adrijan Baric, *Member, IEEE*

***Abstract*—This paper presents a generic strategy for short-term load forecasting (STLF) based on the support vector regression machines (SVR). Two important improvements to the SVR based load forecasting method are introduced, i.e., procedure for generation of model inputs and subsequent model input selection using feature selection algorithms. One of the objectives of the proposed strategy is to reduce the operator interaction in the model-building procedure. The proposed use of feature selection algorithms for automatic model input selection and the use of the particle swarm global optimization based technique for the optimization of SVR hyper-parameters reduces the operator interaction. To confirm the effectiveness of the proposed modeling strategy, the model has been trained and tested on two publicly available and well-known load forecasting data sets and compared to the state-of-the-art STLF algorithms yielding improved accuracy.**

***Index Terms*—Short-term load forecasting, support vector machines.**

I. Introduction

LOAD forecasting has become one of the most critical issues in planning and operation of electric power industry. An accurate estimation of load is essential for electricity price forecast. Load forecasts can have significant implications on energy transactions, market shares and profits in competitive electricity markets.

In this paper, we focus on short-term load forecasting (STLF). Short-term load forecasting time horizon if usually from one hour up to one week. STLF has specific uses in addition to the load forecasting applications, e.g., model power generation systems rely on the day-ahead STLF to help reduce the spinning reserve capacities and to efficiently schedule device maintenance plans. Furthermore, STLF is essential to the reliability of power systems because the load forecasting results are a basis for off-line network analysis that is performed in order to determine if the system is vulnerable. Naturally, the accuracy of such predictions can significantly affect the economic operation and reliability of the power system and, if not adequate, it can lead to allocation of insufficient reserve capacity and use of expensive peaking units, or to unnecessarily large reserve capacity, which are both related to increased operating cost. In an open-market environment, STLF gains even more attention because precise forecasting is the basis of electrical energy trading and spot price establishment, aiming to gain the minimum electricity purchasing cost. In deregulated electricity markets, STLF is important for reliable power system operation [1]. Unfortunately, the load demand is a non-stationary process influenced by numerous factors ranging from weather conditions over seasonal effects to socioeconomic factors and random effects [2] which makes load very difficult to predict.

Due to significant social, economic and environmental effects of STLF, many techniques have been already investigated [3], [4]. Some of them are: artificial neural networks [5]–[7], Fuzzy Logic Approach [6], [8], Bayesian Network Approach, various hybrid approaches [9], and many others, including classical statistical approaches like multiple liner regression and automatic regressive moving average (ARMA) [6]. The overviews of the load forecasting methods can be found in [2], [3], and [10].

During the last ten to fifteen years, neural network approaches combined with other methods (such as evolutionary or fuzzy methods) are most frequently used [3]. Unfortunately, severe issues appear in designing an STLF system based on neural networks for "real world" applications, mostly because of the two effects related to neural networks called "overfitting" (i.e., when a model describes random error or noise instead of the underlying relationship) and "curse of dimensionality" (problem caused by the exponential increase in complexity associated with adding extra dimensions to the ANN input). In such conditions, the forecasting system can yield poor results. Support vector regression (SVR) machines are proposed in [11] and [12] and used for STLF [13], [14] to tackle these problems.

The application of SVR for the load forecasting depends on the available data and desired forecasting horizon. This paper uses SVR machines as a base and proposes an adaptive model and modeling strategy with improved accuracy, when compared with the state-of-the-art methods. The emphasis of the paper is to propose a generic strategy for adaptive model building procedure with limited user interaction requirement. Two well-known and publicly available data sets, New England data set [1] and North-American electric utility data set [15]–[17], are used to compare, test and value the proposed strategy and models. The goal is to predict high level load but the proposed strategy can be also used at lower levels, e.g., at the distribution level.

The paper is organized as follows. In Section II, the proposed STLF strategy and modeling method based on SVR machines is presented. Section III presents testing results. Section IV discusses modeling results while Section V gives conclusions.

Manuscript received November 19, 2012; revised April 25, 2013; accepted June 13, 2013. Paper no. TPWRS-01289-2012.

E. Ceperic is with HEP d.d., Rijeka,51000, Croatia (e-mail: ervin.ceperic@hep.hr).

V. Ceperic is with the University of Zagreb, Faculty of Electrical Engineering and Computing (FER), 10000 Zagreb, Croatia, and also with ESAT-MICAS, KU Leuven, B-3001 Heverlee, Belgium (e-mail: vladimir.ceperic@ieee.org).

A. Baric is with the University of Zagreb, Faculty of Electrical Engineering and Computing (FER), 10000 Zagreb, Croatia (e-mail: adrijan.baric@fer.hr).

Digital Object Identifier 10.1109/TPWRS.2013.2269803

0885-8950/$31.00 © 2013 IEEE

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

II. Proposed Model and Modeling Procedure

In this paper we propose a strategic, seasonality-adjusted, support vector regression machines based model (SSA-SVR) for electric load forecasting. The proposed strategy for short-term load forecasting consists of:

- SVR as the base of the model—because of the favorable properties of SVR when compared to other functional approximation or regression methods [12], [18], [19], as discussed in Section II-A.
- Optimization of SVR hyper-parameters by non-derivative based optimization technique, discussed in Section II-A1, to automate the model building procedure and improve accuracy.
- Using the parallel model architecture instead of recursion to simplify the model building and improve accuracy and robustness, discussed in Section II-B.
- Generation of model inputs to include seasonal effects and most important factors that influence the production and consumption of electrical energy and to prepare for feature selection stage, discussed in Section II-C.
- Selection of model inputs and its lagged (delayed) versions using feature selection algorithms to further automate the model building procedure and improve accuracy, discussed in Section II-D.

The steps required to build the model using the proposed strategy, illustrated in Fig. 1, are the following:

1) Obtain the required data, i.e., past and forecasted weather information and past electric load information.
2) Split the data to the training and validation data set.
3) Generate the model input variables as described in Section II-C. Select maximum number of lagged variables used by the features selection algorithm. More lagged values can be used if the final accuracy of the model is not satisfactory.
4) Use the parallel model architecture to separate the problem into 24 one-hour models.
5) Optimize the SVR hyper-parameters for each hour of the day, i.e., build 24 models, to fit the validation data set. These 24 SVR models are built using only the training data set.
6) Test the model. If the accuracy of the model is not satisfactory, increase the maximum number of lagged values before the feature selection step and repeat the procedure. It should be noted that for all the test cases described in this paper there was no need to increase the number of lagged values in the Step 6) and repeat the procedure from Step 3).
7) Deployment of the model in case satisfying accuracy has been obtained.

In this paper, we have retrained the model for each month during the testing phase for the ISO New England test case while in the North-American test case we did not retrain the model in two-year testing period. In the production environment, more frequent retraining may be a good practice. The overall training time of the proposed model is at maximum 40 min for the test cases presented in Section III, using Linux based workstation with two Intel Xeon X5675 CPUs. The maximum training time was 40 min for the ISO New England test case with 1311 training samples and 60 validation samples on which the model was optimized. The training time is, therefore, smaller than the forecasting period of 1 hour, i.e., the model could be retrained every hour which could improve the model ability to adapt to latest changes in the environment or the power supply system. The evaluation of the trained model is nearly instant, e.g., less than 3 seconds for the two-year period in the North-American test case.

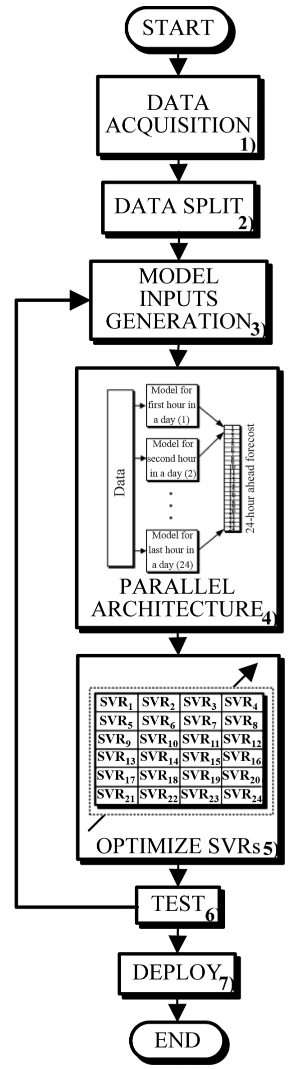

Fig. 1. Overview of the strategy for SVR based STLF.

A. Support Vector Regression

This subsection briefly presents the support vector regression (SVR). Cortes and Vapnik [20] presented support vector machines (SVM), a set of related supervised learning methods used for classification and regression. This approach was originated from Vapnik's statistical learning theory [21]. Since then, support vector machines have generated a lot of interest in the machine learning community due to their excellent performance in a variety of problems, e.g., text categorization [22], image analysis [23], solving problems in bioinformatics [24], bankruptcy prediction [25]. Some of the additional reasons for the wide use

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

of SVMs are: theoretical guarantees about their performance, lower susceptibility to local minima and higher immunity to increased model complexity that is usually associated with adding extra dimensions to the model input space.

SVMs were initially applied to classification problems [20], but soon after the introduction, the formulation was extended to regression problems [11], [12]. Support vector machines seem to offer excellent generalization properties on real-life regression and classification problems. The common formulation of SVM regression is Vapnik's ε-tube support vector regression (ε-SVR) [11], [12] which is the formulation also used in this paper. ε-SVR regression performs linear regression in the high-dimensional feature space created by a kernel function using ε-insensitive loss and, at the same time, tries to reduce model complexity by minimizing model coefficients, i.e., weight vector $\mathbf{w}$ method. In the ε-SVR method, the goal is to find a function f that has the maximum deviation ε from all training data and at the same time is as flat as possible:

$$f(\mathbf{x}) = \sum_{i=1}^{l} \mathbf{w}_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{1}$$

where $\mathbf{x}_i$ is the p-dimensional input vector associated to the output and K is the kernel function. Let l in (1) denotes the number of the training data samples.

The flatness of the function f means that one seeks a small weight vector $\mathbf{w}$. An overview of flatness definitions of such functions is given in [26]. A simple way to ensure required flatness is to minimize the norm $\|\mathbf{w}\|^2 = \langle \mathbf{w}, \mathbf{w} \rangle$.

The vectors $\mathbf{x}_i$ corresponding to non-zero $\mathbf{w}_i$ are called the support vectors. The evaluation complexity of a support vector machine is dependent on the number of support vectors.

The weight $\mathbf{w}$ is usually calculated by transferring the SVR optimization problem to the dual optimization problem that equals the constrained quadratic problem and applying quadratic programming [11], [12].

The kernel function on two vectors $\mathbf{v}$ and $\mathbf{z}$ is the function $K : X \to \mathbb{R}$ that satisfies

$$K(\mathbf{v}, \mathbf{z}) = \langle \Phi(\mathbf{v}), \Phi(\mathbf{z}) \rangle . \tag{2}$$

Kernel function enables the transformation of the input space into high-dimensional feature space where it is possible to apply the linear SV regression algorithm [18]. In this paper, the radial basis function (RBF) is used as a kernel function:

$$\begin{aligned} K(\mathbf{v}, \mathbf{z}) &= \exp\left(-\gamma \|\mathbf{v} - \mathbf{z}\|^2\right) \\ &= \exp\left(-\gamma\left(\langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{z}, \mathbf{z}\rangle - 2\langle \mathbf{v}, \mathbf{z}\rangle\right)\right), \text{for } \gamma > 0. \end{aligned} \tag{3}$$

The RBF kernel function performs well over a wide range of machine learning problems [18], [27]–[30] and is therefore also used in this paper, although other kernel functions can be used with the proposed strategy.

The ε-SVR implementation used in this paper is based on the LIBSVM tool [31].

A detailed analysis why SVR usually performs better than other modeling methods in time series prediction can be found in [27], [32], and [33].

1) Optimization of the SVR Hyper-Parameters: For the ε-tube support vector regression machine, three parameters are the most critical:

- The regularization parameter C. The constant C determines the balance between the flatness of f and the amount of tolerated deviations larger than ε. In other words, the constant C can be considered as a tuning parameter between model complexity and the required training data set model accuracy.
- The width of the ε-tube, i.e., the parameter ε.
- Kernel function additional parameters, e.g., the parameter γ in (3).

The SVR parameters and the additional kernel parameters are together often referred to as the hyper-parameters.

The selection of SVR hyper-parameters is crucial and it imposes an obvious influence on load forecasting. For example, if the value of hyper-parameter C is too large, an overfitting phenomenon may appear.

In this paper we optimize the SVR parameters on the validation data set (and not on the training data set). The validation data set is chosen to be closest in time to the test data set or actual production deployment time. Some researchers, e.g., [34], use cross-validation technique to optimize the model parameters. If the data is not stationary, then using a validation data set closest to the actual deployment time (or testing data set) may improve the model accuracy because the model parameters are optimized on the latest available data.

The standard approach to find a near optimal C and ε is to perform a grid search (GS), as suggested in [35]. It should be noted that in [35] the grid search for optimal parameters is performed for classification problems. This grid search approach is adopted by many researchers and used also for SVM regression, e.g., [36]. Hsu *et al.* in [35] suggest exponentially growing sequences of SVR parameters as a practical method to identify optimal parameters.

The main problem with the grid search strategy is that it is very time consuming [36].

In the last few years, many researchers used the particle swarm optimization (PSO) method to optimize the hyper-parameters of the SVM [37]–[39]. Particle swarm optimization is an optimization algorithm based on the principles of swarm intelligence and it finds a solution to an optimization problem in the search space based on mimicking social behavior of swarms. The use of PSO enables faster and more accurate identification of near-optimal hyper-parameters [37]–[39] compared to conventional methods such as, e.g., the grid search method.

In this paper, we use the particle swarm pattern search method (PSwarm) for SVR hyper-parameter optimization, which is a PSO inspired technique. PSwarm, a hybrid algorithm for global minimization introduced in [40], [41], combines a heuristic approach for global optimization (particle swarm) with a rigorous method (pattern search) for local minimization. PSwarm enjoys the global convergence properties of pattern search and the efficiency of PSO. In [40] and [41], it is shown that PSwarm is the most robust among all global optimization solvers tested and that it is highly competitive in efficiency.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

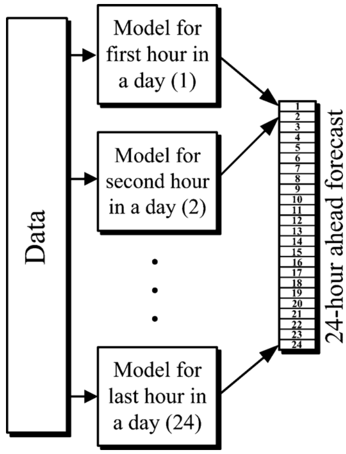

Fig. 2. Parallel model architecture for 24-hour ahead forecasting.

B. Parallel Model Architecture for 24-Hour Load Prediction

In this paper, a nonlinear autoregressive model with exogenous inputs (abbreviated, NARX) is used for the load forecasting.

NARX model can be described as

$$y_t = E\Big(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, \mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \mathbf{x}_{t-3}, \ldots\Big) + e_t \tag{4}$$

where E indicates an unknown nonlinear function, y_t is the prediction on the day t, e_t is the model error (sometimes called noise), and $\mathbf{x}$ is the regression vector of exogenous variables.

The next day hourly load is determined from the previous loads, as well as from the influence of exogenous variables (commonly temperature or some other variables in order to follow weekly or seasonal patterns).

In the 24-hour load prediction scenario we can use recursive or parallel model. In the parallel model, each hour is predicted separately.

In this paper we have used the parallel model architecture (as shown in Fig. 2) due to the following reasons:

- Smaller data set per each SVR with no loss of any informative training examples. In this way it is possible to handle even large data sets and reduce the overall training and execution time. Also, the model for each hour is simpler than the recursive model.
- Reduced influence of the forecast error in the first hours of the 24-hour forecasting period due to the non-existence of feedback loops that use the model outputs as inputs for the next hour prediction.
- No need to include inputs that indicate the current hour of the day as in [34], thus effectively reducing the number of model binary inputs by 24 (for each hour of the day) and therefore, further simplifying the model.
- The problem of finding a suitable model for one hour of the day is simpler than finding a unique model for all 24 hours of the day. The usage of a simpler model can improve generalization properties and increase accuracy due to reduced possibility of overfitting.

C. Input Variables

A wealth of information is available in the load forecasting domain, such as present and past electrical load (which is the variable that we are forecasting) and the factors that influence the production and consumption of electrical energy such as temperature, humidity, air pressure, seasonal period, holiday season and other data.

A very important factor in load forecasting is the seasonal variation, e.g., the energy consumption is higher during the winter due to increased heating requirements. Several options can be used to exploit this factor:

- Use separate model for each season.
- Divide data by a suitable clustering technique such as the self-organizing neural networks (SOM) [9].
- Usage of binary variables to indicate the season [34].

In this paper the use of binary variables is adopted, as suggested in [34]. We have observed increased accuracy of the model by including the binary input variables such as the day of week, the holiday indicator and other indicators in the input selection pool and the proposed strategy always includes these binary variables because they can implicitly describe seasonal variations. The support vector machines are well suited to handle the increased dimensionality of the model inputs because the complexity of the SVR optimization problem is not inherently sensitive to the number of model inputs, as discussed in Section II-A.

A number of input variables is generated. The following real-valued variables are generated:

- Past electricity load values L—168 lagged (past) values (spanning one week) are generated and the feature-selection algorithm is used to choose the most important lagged values w.r.t. the accuracy these values introduce.
- Temperatures—three temperature-based sets of variables are created to capture the effect of cooling and heating requirements [42] on the load. The variable $T_{CR} = \max(T - 20°C, 0)$ captures the cooling requirement, while the heating and extra-heating variables are defined using $T_{HR} = \max(16.5°C - T, 0)$ and $T_{XHR} = \max(5.0°C - T, 0)$ where the temperature thresholds are based on the standard techniques used in the energy industry [34], [42]. From each variable, 96 inputs are generated, spanning three past days (72 inputs) and one day ahead (24 inputs). The total number of generated temperature based variables is 288. It should be noted that for the training, past forecasted temperatures are used instead of the measured temperatures. This step further enhances the accuracy as shown in Section III and Tables IV and VI.
- Humidity index I. Humidity index is used as described in [1] for the ISO New England test case. The data needed to build the humidity index was not available for the North-American test case.

The following binary variables as generated:

- 7 binary variables for each day in a week ($W_{1,\ldots,7}$)—this step enables the model to be sensitive to variations during the week, e.g., $W_{1,\ldots,7} = [1,0,0,0,0,0,0]$ marks Monday.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
MODEL INPUT VARIABLES

Input variable	Mark	Description
Past electricity load values	L	168 lagged values.
Cooling requirement	T_{CR}	Captures the effect of cooling requirements on the load.
Heating requirement	T_{HR}	Captures the effect of heating requirements on the load.
Extra heating requirement	T_{XHR}	Captures the effect of extra-heating requirements on the load.
Humidity index*	I	Only current value of humidity index is used for load forecasting.
Day of the week	$W_{1,\ldots,7}$	7 binary values for each day in a week.
Month	$M_{1,\ldots,12}$	12 binary values for each month in a year.
Holidays	$H_{1,\ldots,4}$	4 binary values for indicating if the previous, current and two days in future are holidays.

* For the ISO New test case.

TABLE II
COMPARISON TO THE STATE-OF-THE-ART—MAPE FOR NEW ENGLAND 2006 LOAD USING ACTUAL WEATHER

	ANN*	SIWNN*	SSA-SVR*
Jan	2.01%	1.6%	1.31%
Feb	1.5%	1.43%	1.12%
March	1.55%	1.47%	0.93%
April	1.51%	1.26%	1.34%
May	1.69%	1.61%	1.1%
June	2.3%	1.79%	1.55%
July	3.72%	2.7%	1.86%
Aug	3.33%	2.62%	1.5%
Sep	1.6%	1.48%	1.15%
Oct	1.52%	1.38%	1.2%
Nov	1.73%	1.39%	1.09%
Dec	1.91%	1.75%	1.61%
Average	2.03%	1.71%	1.31%

* ANN and SIWNN results are reported in [1].

TABLE III
COMPARISON TO THE STATE-OF-THE-ART—MAPE FOR NEW ENGLAND 2006 LOAD USING FORECASTED WEATHER

	ANN*	SIWNN*	SSA-SVR*
Jan	2.12%	1.65%	1.53%
Feb	1.66%	1.48%	1.32%
March	1.66%	1.52%	1.13%
April	1.59%	1.32%	1.37%
May	1.76%	1.66%	1.12%
June	2.42%	1.83%	1.56%
July	3.86%	3.24%	2.06%
Aug	3.48%	2.67%	1.86%
Sep	1.66%	1.52%	1.19%
Oct	1.58%	1.43%	1.32%
Nov	1.83%	1.44%	1.14%
Dec	2.03%	1.81%	1.68%
Average	2.14%	1.80%	1.44%

* ANN and SIWNN results are reported in [1].

- 12 binary variables for each month in a year ($M_{1,\ldots,12}$)—this step enables the model to be sensitive to seasonal variations and circumvents the need to build a separate model for each season, e.g., $W_{1,\ldots,12} = [1,0,0,0,0,0,0,0,0,0,0,0]$ marks January.

TABLE IV
INFLUENCE OF THE VARIABLES $H_{1,\ldots,4}$, T_{CR}, T_{HR}, T_{XHR} AND TRAINING USING ONLY MEASURED TEMPERATURE DATA—MAPE FOR NEW ENGLAND 2006 LOAD USING FORECASTED WEATHER

	SSA-SVR without $H_{1,\ldots,4}$	SSA-SVR without T_{CR}, T_{HR} and T_{XHR}	SSA-SVR trained using measured temperature data
Jan	1.82%	1.76%	1.58%
Feb	1.41%	1.65%	1.44%
March	1.13%	1.27%	1.32%
April	1.39%	1.45%	1.36%
May	1.21%	1.25%	1.29%
June	1.95%	2.20%	2.39%
July	2.26%	2.54%	3.33%
Aug	1.84%	2.27%	2.13%
Sep	1.28%	1.35%	1.32%
Oct	1.26%	1.54%	1.32%
Nov	1.29%	1.25%	1.27%
Dec	2.09%	2.16%	1.87%
Average	1.58%	1.72%	1.72%

- 4 binary variables for the holidays. The first variable marks if the current day is a holiday, the second marks if the previous day was a holiday. The third and forth variable mark if the next two days are holidays. In this way, the influence of the holidays on the electric load can be taken into account.

Some researchers, e.g., Espinoza *et al.* [34] use the additional variables to indicate the hour of the day. In our case, this is not needed because we use the parallel model architecture for 24-hour forecasting.

Table I presents the overview of the model input variables.

All input variables are normalized in the range $[-1, 1]$ as suggested in [43].

D. Feature Selection Algorithms for Selection of Model Inputs

When building an STLF model one of the critical steps is the selection of input variables.

Although the selection of variables based on experience can yield very good results, the process requires (along with expert knowledge) numerous attempts to obtain satisfactory results. The use of feature selection (FS) algorithms can reduce the time needed to build a model, further automate the modeling process and even improve the accuracy of the model. A number of feature selection algorithms exists, e.g., feature selection based on genetic algorithms and information theory (GA&IT) [44], feature selection using mutual information and maximum-relevance minimum-redundancy criterion (MRMR) [45], feature selection using ReliefF algorithm [46], sequential feature selection with stepwise regression (SteepwiseFS) [47].

In this paper, the feature selection algorithms are used to select the near-optimal inputs of the model, with little or no user intervention. The analysis in Section III-C shows that the SteepwiseFS algorithm contributes most to the model accuracy in the test period.

III. TEST CASES

The proposed model is compared in Section III-A with the state-of-the-art model described in [1] on the daily elec-

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

trical loads in New England and in Section III-B with the state-of-the-art models described in [15]–[17] on the daily and hourly loads in North America. The comparison with simple methods, e.g., linear regression, autoregressive filter, is omitted because the state-of-the-art models that have been used for comparison claim to have better accuracy than the simple methods [1], [15]–[17].

It should be noted that all data used in this paper is publicly available which allows other researchers to perform comparison of their models to the models proposed in this paper and easy reproduction and validation of the results presented in this paper.

A. ISO New England

For the ISO New England test case, the variable being predicted is the actual system load in MW as determined by metering. The electric loads are collected from ISO New England.[1] The past predicted temperatures are from ECMWF,[2] obtained using Monitoring Atmospheric Composition and Climate (MACC) model with 1.25 degrees resolution. The forecasts are interpolated using least square linear regression in order to get more realistic forecasts for the New England area (targeted for STLF).

The data in the training data set is from March 2003 until January 2006. In order to test the proposed model, we have predicted the hourly loads for the period from January 1, 2006 to December 31, 2006. The data prior to the current forecasting point is used either for training or validation. The validation data set is always set to be 60 days prior to the test period. The model parameters, i.e., SVR hyper-parameters and the selection of the model inputs are tuned on the validation set.

The proposed model and modeling strategy is compared to the artificial neural networks (ANN) based model and the state-of-the-art model called similar day-based wavelet neural network (SIWNN), as presented in [1].

Table II compares the performance of the proposed SSA-SVR model to the ANN and SIWNN models on the New England data set using actual weather values, while Table III compares the models using the forecasted weather data. The error measure used is the mean absolute percentage error (MAPE). The comparison test data set includes 8760 samples (during 12-month testing period). On the given test data set, SSA-SVR model is 20% better when using forecasted weather data and 23.4% better when using actual weather data.

As shown in Fig. 3, the most difficult months for STLF are July and August. Nevertheless, MAPE of the SSA-SVR model does not exceed 2.1%.

Table IV shows the influence of the following factors on the accuracy of the model:

- 4 binary variables for the holidays $(H_{1,\ldots,4})$ (marked as "*SSA-SVR without* $H_{1,\ldots,4}$" in the Table IV). The MAPE is 9.7% worse if the variables are not included as model inputs.
- T_{CR}, T_{HR} and T_{XHR} (marked as "*SSA-SVR without* T_{CR}, T_{HR} *and* T_{XHR}" in the Table IV). The MAPE is 19.7% worse if non-processed temperature (T) is used instead of T_{CR}, T_{HR} and T_{XHR}.
- SSA-SVR training using only measured temperature data. Neglecting the influence of the error in weather forecasts in the SSA-SVR training yields 19.7% worse MAPE.

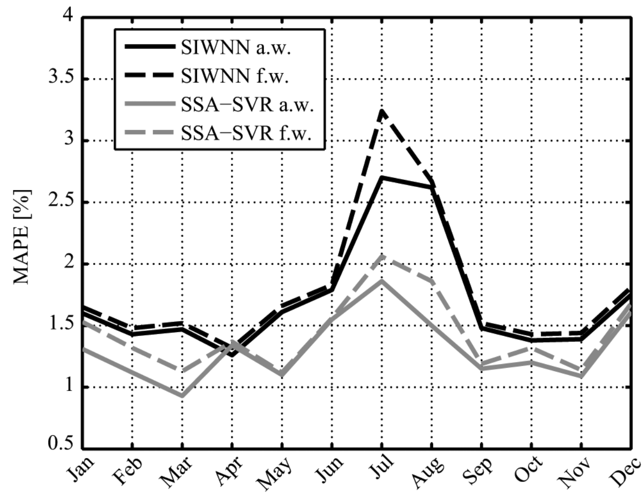

Fig. 3. Mean absolute percentage error (MAPE) for New England 2006 load data set—SIWNN versus SSA-SVR. Note: "f.w." is an abbreviation for the case when models use forecasted weather while "a.w." is an abbreviation for the case when models use actual weather data.

B. North-American Utility

In this section the proposed strategy is tested on the North-American utility data set[3][15]–[17]. The hourly load and temperature data set used in this test case belong to an electric utility in North America. The North American bulk power system consists of five AC power grids (Western, Eastern, Texas, Quebec and Alaska power grids) all connected together [17]. Due to legal reasons, the name of that electric utility is omitted as in [15]–[17].

The data in this test case is from January 1988 until October 12, 1992. In order to test the proposed model, we have predicted the hourly loads for the two-year period prior to October 12, 1992. The data prior to the current forecasting point is used either for training or validation. The validation data set is set to be 365 days prior to the test period. The model parameters, i.e., SVR hyper-parameters and the selection of the model inputs are tuned on the validation set.

Table V compares the performance of the proposed SSA-SVR model to the models described in [15]–[17] on the North-American utility data set. The error measure used is the mean absolute percentage error (MAPE). On the given test data set, SSA-SVR model has from 2.5% up to 34.2% better MAPE accuracy when compared to the best available model.

Fig. 4 shows the accuracy per hour of the day of the SSA-SVR model compared to the accuracy of the model presented in [17] for North-American utility data set using noisy temperature data set.

Table VI shows the influence of the variables $H_{1,\ldots,4}$, T_{CR}, T_{HR} and T_{XHR} and the SSA-SVR training using only measured temperature data on the accuracy of the model. The use of

[1]Available at http://www.iso-ne.com/markets/hstdata/znl_info/hourly/index.html

[2]Available at http://data-portal.ecmwf.int/data/d/macc_reanalysis/

[3]Available at http://fkeynia.googlepages.com/loaddata

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
COMPARISON TO THE STATE-OF-THE-ART—MAPE FOR NORTH-AMERICAN UTILITY DATA SET

Model	Predictions using measured temperature		Prediction using noisy temperature*	
	1-hour ahead**	24-hour ahead**	1-hour ahead**	24-hour ahead**
[15]	1.1%	2.64%	1.11%	2.84%
[16]	0.99%	2.04%	n.a.	n.a.
[17]	1.14%	2.37%	1.21%	2.533%
SSA-SVR	0.72%	1.99%	0.73%	2.03%
Improvement of SSA-SVR	+27.2%	+2.5%	+34.2%	+19.9%

* The Gaussian noise with 1 °F standard deviation is added to the measured temperature as done in [15, 17].
** 24-hour ahead and 1-hour ahead mark the perdition of the electric load in 24- and 1-hour time intervals.

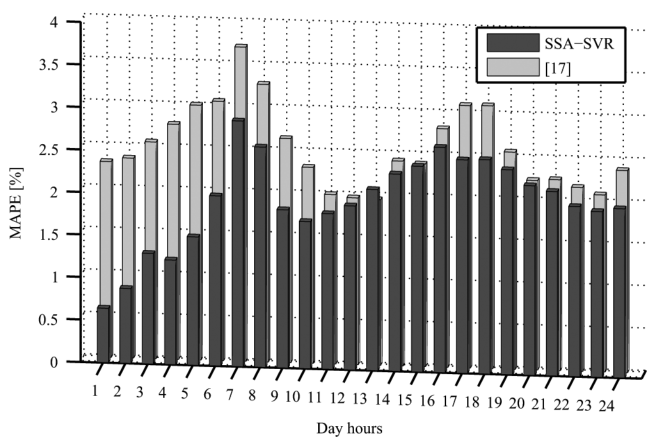

Fig. 4. MAPE for North-American utility data set with noisy temperature data set—SSA-SVR versus model presented in [17].

TABLE VI
INFLUENCE OF THE VARIABLES $H_{1,\dots,4}$, T_{CR}, T_{HR}, T_{XHR} AND TRAINING USING ONLY MEASURED TEMPERATURE DATA—MAPE FOR NORTH-AMERICAN UTILITY DATA SET

	SSA-SVR without $H_{1,\dots,4}$	SSA-SVR without T_{CR}, T_{HR} and T_{XHR}	SSA-SVR trained using measured temperature data
1-hour ahead	0.72% (0%)*	0.73% (+1.3%)*	0.76% (+5.5%)*
24-hour ahead	2.05% (+3%)*	2.12% (+6.5%)*	2.06% (+3.5%)*

* The values in the parentheses mark the percentage of change w.r.t. accuracy of the model built using proposed strategy with all variables, e.g. the value +3% means that the accuracy has decreased 3%.

all variables has a positive influence on the accuracy, although the influence is somewhat lower than in the case of ISO New England (described in Section III-A).

C. Feature Selection Algorithms for STLF

In this paper, various feature selection algorithms have been tested on the validation data set as shown in Table VII. The validation period is set to span 60 days before the testing period for the ISO New England test case and 365 days for the North-

TABLE VII
ACCURACY OF THE FEATURE SELECTION ALGORITHMS—MAPE

Algorithm	ISO New England* (24-hour ahead**)		North-American* (24-hour ahead)**		North-American* (1-hour ahead)**	
	Validation data set	Test data set	Validation data set	Test data set	Validation data set	Test data set
GA&IT	1.36%	1.58%	2.15%	2.06%	0.76%	0.75%
MRMR	1.34%	1.56%	2.12%	2.02%	0.79%	0.77%
ReliefF	1.44%	1.67%	2.54%	2.58%	0.76%	0.75%
SteepwiseFS	1.15%	1.31%	2.11%	1.99%	0.74%	0.72%
Manual***	1.32%	1.51%	2.24%	2.31%	0.75%	0.79%

* The data sets and the model setup is described in Section III.
** 24-hour ahead and 1-hour ahead mark the perdition of the electric load in 24- and 1-hour time intervals.
*** Best effort of the authors at selecting the features manually.

American utility test case. The feature selection algorithms are set to choose 20 loads and 20 temperature derived variables out of 456 available. The best effort of the authors at selecting the features manually is marked as $Manual$ in Table VII. The best performance on the validation data set is obtained by SteepwiseFS algorithm. The feature selection algorithm chosen on the validation data set is then used for building the model that will be tested on the test data set or deployed in production environment.

The analysis on two test cases, described in Sections III-A and III-B, shows that the SteepwiseFS contributes most to the model accuracy in the test period too, as shown in Table VII. It should be noted that the SSA-SVR model for the ISO New England data set built by using any of the tested feature selection algorithms performs better than the best reported model in [1], while only the ReliefF algorithm for the 24-hour North-American utility case performs worse than best reported models in [16] and [17].

IV. DISCUSSION

The effectiveness of the strategy proposed in this paper has been confirmed on two publicly available and well-known load forecasting data sets. The use of publicly available and free data sets allows other researchers to easily compare their models to the models proposed in this paper.

The proposed strategy yields increased accuracy in comparison to the state-of-the-art STLF algorithms. There are several factors that contribute to the increased accuracy, e.g., use of specific input variables, feature selection algorithms, hyper-parameter optimization procedure on the validation data set closest to the actual deployment time and other.

The strategy proposes that the inputs of the model are first generated in a way to enable the capturing of the seasonal effects. The addition of binary variables, describing holidays periods (days) prior, during and after holidays, as inputs further enhances model accuracy. The importance of inclusion of temperature derived variables T_{CR}, T_{HR} and T_{XHR} instead of plain temperature is confirmed as well as the beneficial influence the use of weather forecasts instead of measured data in the phase of model training.

One of the objectives of the proposed strategy is to reduce the operator interaction in the model-building procedure. The proposed use of feature selection algorithms for automatic model

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

input selection and the use of highly accurate and efficient global optimization technique PSwarm for the optimization of SVR hyper-parameters further reduces the operator interaction and in addition, improves accuracy when compared to manual (expert) based input selection. The data set closest to the actual deployment data is used as a validation data set on which model parameters are optimized using the PSwarm algorithm. In this way model parameters are optimized on the latest available data which can improve results if the data is not stationary.

The training time of the model is less than the model forecasting horizon of one hour, which enables frequent retraining of the model in the production environment and inclusion of the recent changes in the environment and power supply system. The effects of such retraining in the production environment are to be examined in the future.

V. Conclusion

A generic short-term load forecasting strategy based on the support vector regression machines is proposed in this paper. To confirm the effectiveness of the proposed modeling strategy, the models have been trained and tested on two publicly available well-known load forecasting data sets and compared to the state-of-the-art algorithms. The improvements proposed by the strategy are made by taking into account that the effectiveness of the SVR based load forecasting depend significantly on the structure of the available data. The modeling results are encouraging and yield an accuracy improvement from 20% to 23.4% when compared to the state-of-the-art models on the ISO New England load forecasting test cases and from 2.5% up to 34.2% on the North-American test cases.

Acknowledgment

The authors would like to thank the ECMWF and MACC project for supplying the weather forecasts, ISO New England for providing the load forecasting data, Prof. V. H. Ferreira and Prof. A. P. A. da Silva for sharing the North-American utility data set, and Prof. P. Luh for help with finding the ISO New England data set.

References

[1] Y. Chen, P. Luh, C. Guan, Y. Zhao, L. Michel, M. Coolbeth, P. Friedland, and S. Rourke, "Short-term load forecasting: Similar day-based wavelet neural networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 322–330, Feb. 2010.

[2] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: Tools for decision making," *Eur. J. Oper. Res.*, vol. 199, no. 3, pp. 902–907, 2009.

[3] E. A. Feinberg and D. Genethliou, *Load Forecasting*. New York, NY, USA: Springer Science, 2005, ch. 12.

[4] M. D. Felice and X. Yao, "Short-term load forecasting with neural network ensembles: A comparative study (application notes) ," *IEEE Computat. Intell. Mag.*, vol. 6, no. 3, pp. 47–56, Aug. 2011.

[5] D. Park, M. El-Sharkawi, I. R. J. Marks, L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, May 1991.

[6] K. Liu, S. Subbarayan, R. Shoults, M. Manry, C. Kwan, F. Lewis, and J. Naccarino, "Comparison of very short-term load forecasting techniques," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 877–882, May 1996.

[7] R. Abdel-Aal, "Short-term hourly load forecasting using abductive networks," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 164–173, Feb. 2004.

[8] M. Rejc and M. Pantos, "Short-term transmission-loss forecast for the Slovenian transmission power system based on a fuzzy-logic decision approach," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1511–1521, Aug. 2011.

[9] S. Fan and L. Chen, "Short-term load forecasting based on an adaptive hybrid method," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 392–401, Feb. 2006.

[10] S. K. Aggarwal, L. M. Saini, and A. Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation," *Int. J. Elect. Power Energy Syst.*, vol. 31, no. 1, pp. 13–22, 2009.

[11] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems 9*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA, USA: MIT Press, 1996, pp. 155–161.

[12] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Advances in Neural Information Processing Systems 9*. Cambridge, MA, USA: MIT Press, 1996, pp. 281–287.

[13] M. Mohandes, "Support vector machines for short-term electrical load forecasting," *Int. J. Energy Res.*, vol. 26, no. 4, pp. 335–345, 2002.

[14] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on eunite competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, Nov. 2004.

[15] A. Reis and A. A. d. Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 189–198, Feb. 2000.

[16] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," *Energy*, vol. 34, no. 1, pp. 46–57, 2009.

[17] A. Deihimi and H. Showkati, "Application of echo state networks in short-term electric load forecasting," *Energy*, vol. 39, no. 1, pp. 327–340, 2012.

[18] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.

[19] V. Kecman, *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. Cambridge, MA, USA: MIT Press, 2001.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[21] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 2095.

[22] T.-F. Kuo and Y. Yajima, "Ranking and selecting terms for text categorization via SVM discriminate boundary," *Int. J. Intell. Syst.*, vol. 25, no. 2, pp. 137–154, 2010.

[23] B. Guo, S. Gunn, R. Damper, and J. Nelson, "Customizing kernel functions for SVM-based hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 622–629, Apr. 2008.

[24] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch, "Support vector machines and kernels for computational biology," *PLoS Computat. Biol.*, vol. 4, no. 10, Oct. 2008.

[25] W. K. Härdle, R. Moro, and L. Hoffmann, "Learning machines supporting bankruptcy prediction," in *Sonderforschungsbereich 649, Humboldt Universität zu Berlin SFB 649 Discussion Paper 2010–032*, Germany, 2010.

[26] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: MIT Press, 2001.

[27] S. Rüping, *SVM Kernels for Time Series Analysis*, ser. Technical report: Sonderforschungsbereich Komplexitätsreduktion in Multivariaten Datenstrukturen. Univ., SFB 475, 2001.

[28] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[29] A. Singhal and M. Singh, "Performance evaluation of kernels in multiclass support vector machines," *Int. J. Soft Comput. Eng.*, vol. 1, pp. 146–56, 2011.

[30] M. Hussain, S. Wajid, A. Elzaart, and M. Berbar, "A comparison of SVM kernel functions for breast cancer detection," in *Proc. 8th Int. Conf. Computer Graphics, Imaging and Visualization (CGIV)*, 2011, pp. 145–150.

[31] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.

[32] N. Sapankevych and R. Sankar, "Time series prediction using support vector machines: A survey," *IEEE Computat. Intell. Mag.*, vol. 4, no. 2, pp. 24–38, 2009.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[33] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *Proc. 7th Int. Conf. Artificial Neural Networks, ser. ICANN '97 2000*, London, U.K.: Springer-Verlag, 1997, pp. 999–1004.

[34] M. Espinoza, J. Suykens, R. Belmans, and B. D. Moor, "Electric load forecasting," *IEEE Control Syst.*, vol. 27, no. 5, pp. 43–57, Oct. 2007.

[35] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, A Practical Guide to Support Vector Classification, Dept. Comput. Sci. Inf. Eng. Tech. Rep., National Taiwan Univ., 2003.

[36] Y. Bao, Y. Lu, and J. Zhang, "Forecasting stock price by SVMs regression," in *Artificial Intelligence: Methodology, Systems, and Applications, ser. Lecture Notes in Computer Science*. Berlin/Heidelberg, Germany: Springer, 2004, pp. 295–303.

[37] M.-H. Jiang and X.-C. Yuan, "Construction and application of PSO-SVM model for personal credit scoring," in *Proc. 7th Int. Conf. Computational Science (ICCS '07) Part IV*, New York, NY, USA: Springer-Verlag, 2007, pp. 158–161.

[38] C. Cao and J. Xu, Q. Peng, K. C. P. Wang, Y. Qiu, Y. Pu, X. Luo, and B. Shuai, Eds., "Short-term traffic flow predication based on PSO-SVM," in *Proc.1st Int. Conf. Transportation Eng. Amer. Soc. Civil Eng. (ASCE*, Chengdu, China, 2007, vol. 246, pp. 28–28.

[39] S. Fei, C. Liu, Q. Zeng, and Y. Miao, "Application of particle swarm optimization-based support vector machine in fault diagnosis of turbo-generator," in *Proc. 2008 Sec. Int. Symp. Intell. Inf. Technol. Applicat. IEEE Comput. Soc. (IITA '08)*, Washington, DC, USA, 2008, pp. 1040–1044.

[40] A. Vaz and L. Vicente, "A particle swarm pattern search method for bound constrained global optimization," *J. Global Optimiz.*, vol. 39, no. 2, pp. 197–219, 2007.

[41] A. I. F. Vaz and L. N. Vicente, "PSwarm: a hybrid solver for linearly constrained global derivative-free optimization," *Optimiz. Meth. Softw.*, vol. 24, no. 4–5, pp. 669–685, Oct. 2009.

[42] R. F. Engle, C. W. J. Granger, J. Rice, and A. Weiss, "Essays in econometrics," in *Semiparametric Estimates of the Relation Between Weather and Electricity Sales*, E. Ghysels, N. R. Swanson, and M. W. Watson, Eds. New York, NY, USA: Cambridge Univ. Press, 2001, pp. 247–269.

[43] C.-w. Hsu, C.-c. Chang, and C.-j. Lin, A Practical Guide to Support Vector Classification, Tech. Rep., Dept. Comput. Sci., National Taiwan Univ., 2010. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers.html.

[44] O. Ludwig and U. Nunes, "Novel maximum-margin training algorithms for supervised neural networks," *IEEE Trans. Neural Netw.*, vol. 21, no. 6, pp. 972–984, 2010.

[45] G. Brown, "A new perspective for information theoretic feature selection," *J. Mach. Learn. Res.—Proc. Track*, vol. 5, pp. 49–56, 2009.

[46] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RRELIEFF," *Mach. Learn.*, vol. 53, no. 1-2, pp. 23–69, Oct. 2003.

[47] R. R. Hocking, "The analysis and selection of variables in linear regression," *Biometrics*, vol. 32, no. 1, pp. 1–49, 1976.

Ervin Ceperic received the M.Sc. degree in electrical engineering from the University of Trieste—Faculty of Electrical Engineering, Italy, in 2002.

He is currently a system engineer at HEP d.d. (Croatian electric power company). His current research interests include application of machine learning and signal processing to electric load forecasting, and administration and maintenance of IT systems.

Vladimir Ceperic (S'12) received the M.Sc. degree in electrical engineering from the University of Zagreb, Zagreb, Croatia, in 2007. Currently, he is pursuing the double Ph.D. degree from KU Leuven, Belgium, and the University of Zagreb.

His current research interests include machine learning, artificial intelligence methods and their applications for behavioral modeling of electronic circuits.

Adrijan Baric (M'98) received the Dipl.-Ing. and M.Sc. degrees in electrical engineering from the University of Zagreb, Zagreb, Croatia, in 1982 and 1985, respectively, and the Ph.D. degree in electronics from the Dublin City University, Dublin, Ireland, in 1995.

Since 1984, he has been with the University of Zagreb, where he is currently a Professor. His research interests include semiconductor device modeling, integrated circuit design, interconnect modeling and electromagnetic compatibility.