Short-Term Load Forecasting Using Wavelet Transform and Support Vector Machines

J. Pahasa, and N. Theera-Umpon, *Senior Member, IEEE*

***Abstract*--This paper presents a new technique in short-term load forecasting (STLF.) The proposed method consists of the discrete wavelet transform (DWT) and support vector machines (SVMs.) The DWT splits up load time series into low and high frequency components to be the features for the SVMs. The SVMs then forecast each component separately. At the end we sum up all forecasted components to produce a final forecasted load. The data from Bangkok-Noi area in Bangkok, Thailand, is used to verify on the one-day ahead load forecasting. The performance of the algorithm is compared with that of the SVM without DWT, and neural networks with and without DWT. The experimental results show that the proposed algorithm yields more accuracy in the STLF than the others.**

***Index Terms*--Discrete wavelet transform, Electric power systems, Short-term load forecasting, Support vector machine, Support vector regression.**

I. Introduction

SHORT-term load forecasting (STLF) is an essential issue for many schemes in power systems especially in economic dispatch, energy transactions, and unit commitment. The characteristics of electrical load depend on many factors such as weather condition, calendar properties, etc. Moreover, the characteristics of the load in a specific area are unique because of local cultures, social condition, and habit of using electrical devices. These factors make load forecasting very complex and difficult to reach an optimal accuracy. The properties of the load are nonlinear and non-stationary which should be solved by the different manners. Many techniques were proposed to improve the accuracy of the STLF. Artificial intelligence (AI)-based techniques such as neural network (NN) [1]-[4], fuzzy linear regression [5], [6], and support vector machine (SVM) [7] are among the popular techniques. These models consider the relationship between the load data and other features that affect the characteristics of the load. Therefore, feature selection in the AI-based forecasting models is an important issue and has become an active research area in the STLF.

In order to solve the nonlinear function approximation problem, the SVM is a powerful technique. It uses the kernel function to induce the data sets from the input space up to a high dimensional feature space in which we can implement the problem in a linear form [8]-[11]. The SVM has been applied to many prediction problems [12], [13]. The structural risk minimization (SRM) in the learning stage of the SVM is more powerful than the empirical risk minimization (ERM) in neural networks.

The discrete wavelet transform (DWT) is regarded as a powerful method for handling non-stationary discrete signals [14]. The wavelet technique has an advantage over the Fourier transform in that it allows each frequency component to be considered with an appropriate temporal resolution. In previous works, researchers employed the DWT in many applications in power systems such as fault identification [15], power quality classification [16], and load forecasting [4].

In this paper, the DWT and the SVMs are applied for the one-day ahead electric load forecasting. The DWT is used for the feature extraction by splitting up load time series into high and low frequency components. We apply a set of SVMs, each of them forecasts each frequency component with the different features and SVM parameters from the others. The forecasted load is achieved by summing up all forecast components. The accuracy of the proposed model is compared with other three techniques, i.e., SVM without DWT, and NN with and without DWT.

The rest of this paper is organized as follows. Brief descriptions of DWT and SVM are given in section II. In section III, the proposed models are described. Section IV presents the data sets and validation description. The experimental results are shown in section V. Finally, section VI concludes the paper.

II. Methodology

The wavelet transform and support vector machines are well-known and widely available in the literature. We, therefore, will provide only brief descriptions here.

A. Discrete Wavelet Transform

Wavelet transform is a method of analyzing non-stationary signals in both time and frequency domains [14]. It uses mother wavelet as a prototype function. The discrete wavelet transform (DWT) is the wavelet for analyze discrete data. In practice, using multi-resolution analysis and two channel filter bank analysis, the DWT can be performed as shown in Fig. 1 which consists of decomposition and reconstruction stage. In the decomposition stage, signal (S) is recovered from its high-pass and low-pass filters. The output from the first high-pass filter is a detail coefficient (cD_1) and the output from the first low-pass filter is an approximation coefficient (cA_1). The

J. Pahasa is with the Department of Electrical Engineering, Naresuan University Phayao, Phayao 56000 Thailand (e-mail: jonglakp@pyo.nu.ac.th).

N. Theera-Umpon is with the Department of Electrical Engineering, Chiang Mai University, Chiang Mai 50200 Thailand (corresponding author* e-mail: nipon@ieee.org).

978-981-05-9423-7 © 2007 RPS

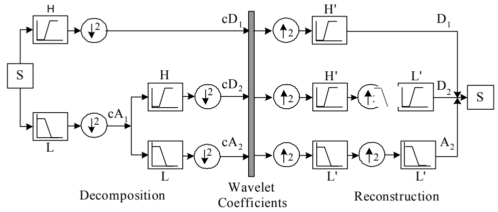

Fig. 1. Discrete wavelet decomposition and reconstruction.

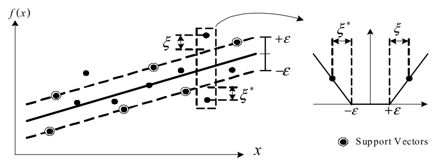

Fig. 2. Soft margin ε -insensitive loss setting in a linear SVR.

second high-pass and low-pass filters split up cA_1 into cD_2 and cA_2, respectively. In the reconstruction stage, the wavelet coefficients are used to reconstruct data in each frequency component. The components are summed up to produce the reconstructed data, in this example, $S=D_1+D_2+A_2$.

B. Support Vector Machine

Consider a set of training data, $\{\mathbf{x}_i, y_i\}_{i=1}^{\ell}, \mathbf{x}_i \in \Re^n$, $y_i \in \Re$ with l observations. The basic idea of SVM is to map the training data from the input space into a higher dimensional feature space via kernel function and then construct a separating hyperplane with maximum margin in the feature space [8]-[11]. Therefore, the SVM estimating function takes the form

$$f(\mathbf{x}) = \langle \mathbf{w} \bullet \Phi(\mathbf{x}) \rangle + b\,, \tag{1}$$

where $\mathbf{w}$ is a weight vector, b is a bias and Φ denotes a nonlinear transformation from input space to high-dimensional feature space. The goal is to find a function $f(\mathbf{x})$ that has an ε-deviation from the actually target y_i for all training data set and as flat as possible, as shown in Fig. 2. The training samples which have an ε-deviation called "support vector". Slack variables, ξ_i, ξ_i^* are measures of the cost of errors on upper and lower constrains. The samples which locate closer than ε-deviation have null cost. We have to solve the following optimization problem

$$\min_{w,\xi_i,\xi_i^*} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{\ell}(\xi_i - \xi_i^*) \tag{2}$$

subject to

$$\left.\begin{array}{c} y_i - \langle \mathbf{w} \bullet \Phi \mathbf{x}_i \rangle - b \le \varepsilon + \xi_i \\ \langle \mathbf{w} \bullet \Phi \mathbf{x}_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{array}\right\} \tag{3}$$

where C is a constant, $\mathbf{x}_i$ is a support vector and $\mathbf{w}$ can be written in terms of data points as

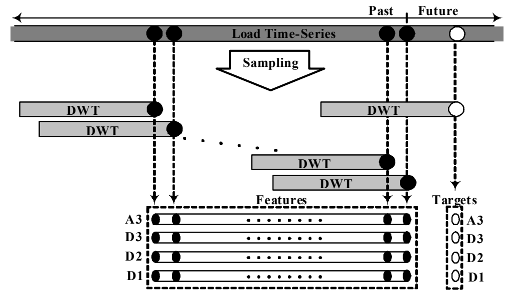

Fig. 3. Features and targets of different frequency components of load time series using DWT.

$$\mathbf{w} = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)\Phi(\mathbf{x}_i)\,. \tag{4}$$

By substituting (4) into (2), the SVM optimum hyperplane equation can be rewritten as

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)\left(\Phi(\mathbf{x}_i) \bullet \Phi(\mathbf{x})\right) + b \\ &= \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) + b \end{aligned} \tag{5}$$

where $K(\mathbf{x}_i, \mathbf{x})$ is a kernel mapping function between sampling $\mathbf{x}$ and support vector $\mathbf{x}_i$.

III. Proposed Models

A. The DWT-Based Models

Wavelet transform uses a mother wavelet as a prototype function. In this work, the DWT is applied to divide load time series into many frequency components. Each component is then predicted with different features and forecast tool parameters. In our experiments, we use Dubechies2 with 3 levels decomposition as a mother wavelet. Fig. 3 shows the DWT that transforms the load data into the features and the targets of the low and high frequency components. The low frequency component is an approximation of the load time series called A3 whereas many high frequency components are the details of the load time series called D1, D2, and D3. The load time series used in our experiments were collected in every 30 minutes. Therefore, we down sample the series by the factor of 2 to achieve the one-hour sampling rate.

B. Feature Selection

One of the important issues in short-term load forecasting is the feature selection. There are many factors that affect the daily load. The main features are the characteristics of the past and current loads. Additionally, the weather condition also plays an important role. In different areas, the characteristics of weather that highly affect the load patterns are different. For example, in Thailand, the temperature is the most important weather condition for the load prediction.

In this work, we utilize parts of the past and current DWT components of the load time series as our features. The 24-

TABLE I
LIST OF INPUT FEATURES FOR DIFFERENT DWT COMPONENTS

Model	Features	Target
1	$A3_{(0)}$, $A3_{(-24)}$, $A3_{(-144)}$, $L_{(0)}$, $L_{(-24)}$, $T_{(+24)}$	$A3_{(+24)}$
2	$D3_{(0)}$, $D3_{(-24)}$, $D3_{(-144)}$, $L_{(0)}$, $L_{(-24)}$, $T_{(+24)}$	$D3_{(+24)}$
3	$D2_{(0)}$, $D2_{(-24)}$, $D2_{(-144)}$, $L_{(0)}$, $L_{(-24)}$, $T_{(+24)}$	$D2_{(+24)}$
4	$D1_{(0)}$, $D1_{(-24)}$, $DI_{(-144)}$, $T_{(+24)}$	$D1_{(+24)}$

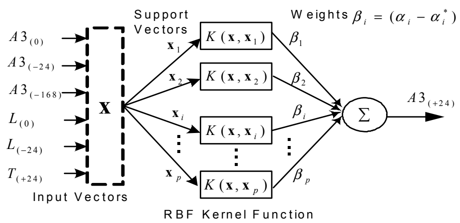

Fig. 4. A3 forcasting using SVM model.

hour ahead predicted temperature is also included in our feature vector. To be more specific, the features are shown in Table I in which L and T denote the electrical load and the temperature features, respectively. A3, D3, D2, and D1 are the DWT components of the load time series. The subscripts indicate the time instance of the feature, i.e., (0), (−24), (−144), and (+24) indicate the present time, previous 24 hours, previous 144 hours, and the next 24 hours, respectively.

C. Forecast Models

We combine the DWT and the SVM to achieve a short-term electrical load forecasting system. As mentioned in the previous section, the DWT is applied to extract the features. The SVM is applied as a forecasting tool. Particularly, we apply the SVM for regression problems called the support vector regression (SVR) in our system.

To decide the number of models in this particular problem, the load data is grouped according to the similarity of the load patterns. We decide not to classify on seasonal approach because there is fuzziness in the season partition for any given day. We classify the forecast load according to their time and day type. The reason is that the patterns of the load on different days of the week are different. Moreover, the load patterns at different time instants of a day are also different. In addition, the normal and abnormal days such as holidays also have the different load demands. However, this work does not forecast the load on the holidays because the different holidays have the different load demands that have to be considered more carefully [6]. Therefore, the forecasted load data are grouped into 7×48 = 336 groups for the 7 calendar days in a week and 48 data points collected every half-hour in a day.

In this work, we forecast the DWT components, i.e., A3, D3, D2 and D1, separately. The final forecasted load is the summation of these 4 components. Fig. 4 shows the structure of the SVM used to forecast A3. From the extensive

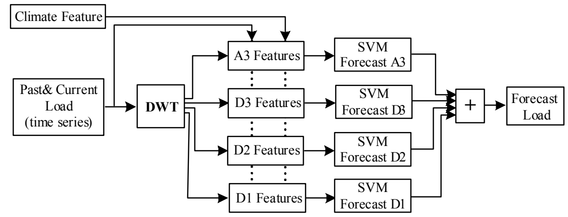

Fig. 5. WTSVM load forcasting model.

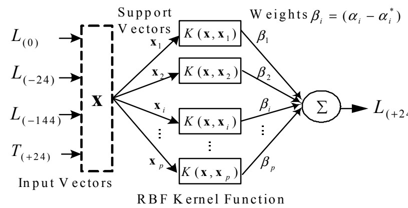

Fig. 6. SVM load forcasting model.

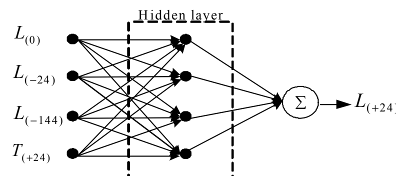

Fig. 7. NN load forcasting model.

experiments, we tried several sets of parameters and found that the parameters that yielded the best performance are C=10, ε=0.001, and σ=7. Support vectors (x_i) are obtained from the training stage. In the testing stage, they are used for kernel functions to induce the data from input space up to high dimensional feature space, i.e.,

$$K(\mathbf{x},\mathbf{x}_i) = \exp(-\|\mathbf{x}-\mathbf{x}_i\|^2 / 2\sigma^2) . \qquad (6)$$

In the same manner, the components D3, D2, and D1 are predicted based on similar SVM models except the radial basis function parameters. Specifically, we use σ = 4, 5, and 15, for D3, D2, and D1, respectively.

Fig. 5 shows the overall picture of our proposed STLF system using the DWT and the SVM models. For the convenience in the comparison in the next section, we denote it as the WTSVM model.

D. Model Comparison

To investigate the performance of the proposed models, we construct other three different schemes, i.e., SVM, NN, and WTNN models. The details of these 3 models are as follows:

1) SVM model: This model uses the SVM as a forecasting tool. The parameters of the SVM are the same as the former SVM except the RBF kernel parameter in which this model uses σ=4. Fig. 6 shows the SVM forecasting model.

2) NN model: The NN model uses the feed forward back propagation neural network as a forecasting tool. The

TABLE II
SUMMARY OF TWO-YEAR TEMPERATURE AND LOAD DATA

Year	Min	Max	Mean	Median	Std.
	Load (MW)				
2004	124.60	536.20	356.09	356.55	85.81
2005	111.20	555.10	339.39	340.10	80.53
	Temperature (°C)				
2004	14.50	43.90	31.10	30.40	3.96
2005	20.30	42.70	31.07	30.40	3.86

structure of the NN is shown in Fig. 7. The features used in this model are the same as the SVM model.

3) WTNN model: The structure of this model consists of the DWT and the NN. This model has a similar structure to that of the WTSVM model except the forecasting tool in which it uses the NN instead of the SVM.

Therefore, this work compares 4 different models. The objectives of the comparison are two-fold. Firstly, to investigate the role of the DWT in the forecasting, we compare the forecasting results of the models with and without the DWT. To achieve this objective, the performance of the SVM model is compared to that of the WTSVM model while the performance of the NN model is compared to that of the WTNN model. Secondly, to investigate the performance of the SVM models, we compare the forecasting results of the SVM and that of the NN which is a popular forecasting tool to solve the nonlinear function approximation. To achieve this objective, the accuracy of the SVM model is compared to that of the NN model while the accuracy of the WTSVM model is compared to that of the WTNN model.

IV. DATA SETS AND VALIDATION DESCRIPTION

A. Data Sets

The data sets used in our experiments consist of the half-hourly electrical load series and temperature data from Bangkok-Noi area in Bangkok, Thailand, over the two-year period from January 1, 2004 to December 31, 2005. Table II shows the information of these load series and temperature data.

B. Validation Description

The entire data set is separated into 2 sets, i.e., a training set and a test set. The training set is the one-year data from January 1, 2004 to December 31, 2004. It is used in the evolution the proposed algorithm, i.e., selection of input features and the SVM parameters. The test set is the other one-year data from January 1, 2005 to December 31, 2005. It is used as a blind test set to evaluate the performance of the model.

In the training stage, the k-fold cross validation has been applied to evaluate the performance of the proposed method. It can detect and prevent over-fitting in a model in which the training model fits in the training data but does not fit in the test data and then produces poor forecasted values.

In the k-fold cross validation, the training data are broken into k subsets in the first stage. Then a regression model is trained on the union of $k-1$ subsets, and evaluated on the remaining subset. The process is repeated k times, using each of subset as the test set once. Finally, the results from the test subsets are combined to get an overall estimate of the effectiveness of the training procedure. Selection of parameter k depends on the data set. In our experiments, we use k=5. Therefore, all the forecasting tools' parameters and the features in the previous sections are obtained based on the 5-fold cross validation.

TABLE III
MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) FROM THE WTSVM MODEL

Data	*DWT Components*				*Load*
	A3	*D3*	*D2*	*D1*	
Training set	2.26	102.94	147.41	259.74	4.51
Test set	3.13	307.14	184.96	281.62	6.01

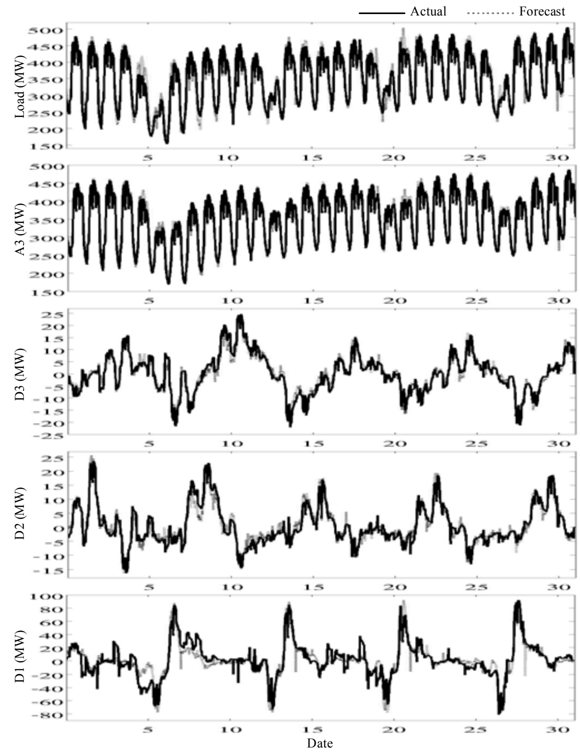

Fig. 8. Actual and forcasted values of load and DWT components in March 2005 from the WTSVM model.

V. EXPERIMENTAL RESULTS

In our experiments, we perform the one-day ahead electrical load forecasting. The experiments are extensive because we construct each model for a particular time instant (every half hour) in a particular calendar day. Additionally, we use the 5-fold cross validation and try several model parameters. The accuracy is reported in term of the mean absolute percentage error (MAPE) which is defined as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{ActualLoad_i - ForecastLoad_i}{ActualLoad_i}\right| \times 100\% , \qquad (7)$$

TABLE IV
FORECASTED RESULTS OF THE FOUR MODELS

Data	MAPE			
	NN	*SVM*	*WTNN*	*WTSVM*
Training Set	5.18	4.70	4.81	4.51
Test Set	10.87	6.65	10.67	6.01

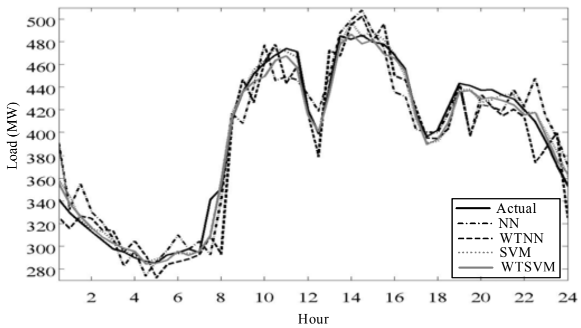

Fig. 9.Comparison of actual and forecasted load from 4 models on Thursday, March 24, 2005.

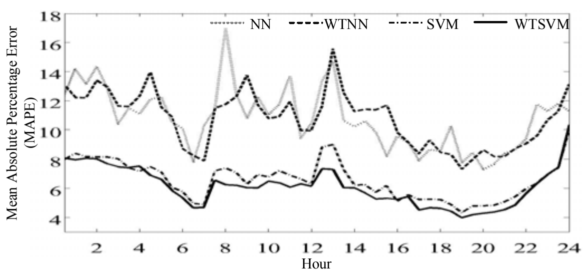

Fig. 10. Comparison of forecasting errors at the time of day.

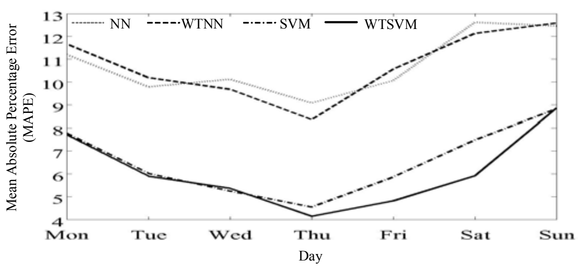

Fig. 11. Comparison of forecasting errors on the day of week.

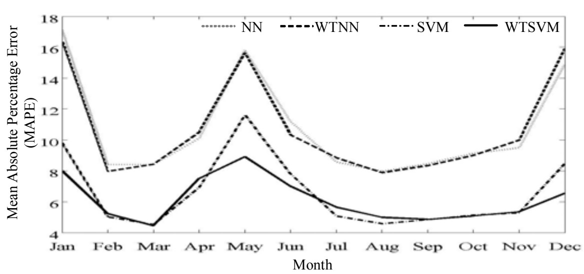

Fig. 12. Comparison of the forecasting errors in the month of year.

where n is the number of the test data.

Table III shows the MAPE of the training and test sets from the WTSVM model. We can see that the MAPE of the forecast of the approximation component A3 is less than that of the 3 high frequency components. We also achieve a good forecasting performance of the load, i.e., the MAPE is about 6%. We show the pictorial performance of the forecasting model in Fig. 8. It shows the actual and forecasted load and 4 DWT components in March 2005. It can be seen in the top row that the forecasted load is very close the actual one. This is also applied to the 4 DWT components in the next 4 rows.

The forecasted results from the different forecasting schemes are presented in Table IV. Fig. 9 shows the comparison for an actual and forecast load on Thursday, March 24, 2005. The MAPE of hourly, daily, and monthly for the test data in one year are shown in Fig. 10 to 12. It is noticed that the DWT strategies improve the accuracy of the load forecasting. In addition, the SVM is the better forecasting tool than the NN for this particular problem. Therefore, all the results obtained demonstrate the proposed WTSVM model is the promising choice in short- term load forecasting.

VI. CONCLUSION

We present a short-term load forecasting system using discrete wavelet transform (DWT) and support vector machines (SVMs). The load time series have been separated into high and low frequency components using the DWT. The SVM forecasts each component separately. The forecasted components are summed up to produce a final forecasted load. Each DWT component forecasting has different SVM parameters. The data from Bangkok-Noi area in Bangkok, Thailand, is used to verify the proposed method. The proposed algorithm is compared with the SVM, NN and WTNN models. The experimental results show that using DWT in the feature extraction stage improves the forecasting performances in both NN and SVM. The results also suggest that the SVM yields better forecasting performances than the NN. The forecasting model using the SVM as a forecasting tool and the DWT in the feature extraction stage yields more accuracy in the STLF than the other models.

ACKNOWLEDGMENT

The authors would like to thank the metropolitan area control center, metropolitan area operation, department of the Electricity Generating Authority of Thailand (EGAT) for providing the data.

REFERENCES

[1] H. S. Hippert, C. E. Pedreira and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Trans. On Power Syst.*, vol.16, no. 1, pp.44-55, Feb. 2001.

[2] H. Mori and A. Yuihara, "Deterministic annealing clustering for ANN-based short-term load forecasting," *IEEE Trans. on Power Syst.*, vol.16, no.3, pp.545-551, Aug. 2001.

[3] G. C. Liao and T. P. Tsao, "Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting," *IEEE Trans. on Evol. Comp."* vol.10, no.3, pp.330-340, June 2006.

[4] A. J. R. Reis and A. P. A. Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. on Power Syst.*, vol.20, no.1, pp189-198, Feb. 2005.

[5] K. B. Song, S. K. Ha, J. W. Park, D. J. Kweon and K. H. Kim, "Hybrid load forecasting method with analysis of temperature sensitivities," *IEEE Trans. on Power Syst.*, vol.21, no.2, pp.869-876, May 2006.

[6] K. B. Song,Y. S. Baek, D. H. Hong and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Trans. on Power Syst.*, vol.20, no.1, pp.96-101, Feb. 2005.

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

[8] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol.20, no.3, pp.273-297, 1995.

[9] V. Vapnik, "An overview of statistical learning theory," *IEEE Trans on Neural Networks*, vol.10, no.5, pp. 988-999, Sept. 1999.

[10] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol.14, pp.199-222, 2004.

[11] S. Gunn, "Support vector machiens for classification and regression," *ISIS Technical Report ISIS-1-98, Image Speech & Intelligent Systems Research Group*, University of Southapton, May 1998.

[12] U. Boonprasert, N. Theera-Umpon, and C. Rakpenthai, "Support vector regression based adaptive power system stabilizer," in *Proc. 2003 IEEE Intl. Symp. on Circuits and Syst.***,** pp.III-371 - III-374.

[13] N. Theera-Umpon and U. Boonprasert, "Leeway prediction of oceanic disastrous target via support vector regression," *Jnl. of Adv. Comp. Intel. and Intel. Inform.,* vol.8, no.6, pp.602-612, Nov.2004.

[14] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine,* vol.8, no.4, pp.14-38, Oct.1991.

[15] C. Koley, P. Purkait, and S. Chakravorti, "Wavelet-aided SVM tool for impulse fault identification in transformers," *IEEE Trans. On Power Deli.*, vol.21, no.3, pp.1283-1290, July 2006.

[16] H. He and J. A. Starzyk, "A self-organizing learning array system for power quality classification based on wavelet transform," *IEEE Trans. On Power Deli.*, vol.21, no.1, pp.286-295, Jan.2006.

Jonglak Pahasa received M.Eng in Electrical Engineering from Chiang Mai University, Thailand, in 2007. She received her B.Eng in Electrical Engineering from the King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2000. Currently, she is a Lecturer with the Department of Electrical Engineering, Naresuan University Phaoyao, Phayao Thailand. Her research interests include an application of support vector machine and neural network in power system planing and operation.

Nipon Theera-Umpon (S'98-M'00-SM'06) received his Ph.D. from the University of Missouri–Columbia, M.S. degree from the University of Southern California, and B.Eng (Hons.) degree from Chiang Mai University, all in Electrical Engineering, in 2000, 1996, and 1993, respectively. Since 1993, he has been a faculty member in the department of Electrical Engineering, Chiang Mai University, where he is now an associate profesor. He has also served as an associate dean of Engineering for research and international relations, Chiang Mai University, since 2005. From 2000 to 2001, he received a postdoctoral fellowship and was a visiting scholar at the Center for Applied Image Recognition, University of Missouri–Columbia where he conducted a research on automatic land mine detection. His research interests include computational intelligence, mathematical morphology, neural networks, medical imaging, and signal and image processing in many fields of application. He is a vice president of the Thai Engineering in Medicine and Biology Society. He is also a member of the Thai Robotics Society, and the Council of Engineers, Thailand.