Support vector machines with genetic algorithms in forecasting electricity load

Wei-Chiang Hong
Ph.D. Program in Management, School of Management, Da-Yeh University
112 Shan-Jiau Road, Da-Tusen, Chang-Hua, 51505, Taiwan
email: d9230006@mail.dyu.edu.tw
Ping-Feng Pai*
Department of Industrial Engineering and Technology Management, Da-Yeh University
112 Shan-Jiau Road, Da-Tusen, Chang-Hua, 51505, Taiwan
email: paipf@mail.dyu.edu.tw

* Corresponding author

***Abstract*--Support vector machines (SVMs) have been successfully used in solving nonlinear regression and time series problems. However, the application of SVMs to load forecasting is very rare. Therefore, the purpose of this paper is to examine the feasibility of SVMs in forecasting electricity load. In addition, the genetic algorithms are applied in the parameter selection of SVM model. Forecasting results compared with other two models, namely autoregressive integrated moving average (ARIMA) and general regression neural networks (GRNN), are provided. The experimental data are borrowed from the Taiwan Power Company. The numerical results indicate that the SVM model with genetic algorithms (SVMG) results in better predictive performance than the other two approaches.**

I. Introduction

A. Electricity demand management

The accurate forecasting for the future electricity demand has become one of the primary management goals, such as load planning, cost economy and secures operation fields, in a regional or a national system. Since the increasingly competitive energy market in many countries, they are required to play a specific role in scheduling functions, such as unit commitment, hydrothermal coordination, short-term maintenance, inter-change and transaction evaluation, network power flow dispatched optimization, security strategies etc.. Along with the power system privatized and deregulated in the recent years, the reliance and accuracy of forecasting have received more attention.

In the meanwhile, the error of load forecasting may increase the operating cost [1,2]. It was suggested that the increase of 1% in forecasting error would imply in a £ 10 million of extra operating costs [1]. Overestimation of future load results in an unnecessary spinning reserve. On the other hand, underestimation of load leads to a failure in providing enough reserve and implies high costs in peaking unit.

However, the load forecasting is a complex task, due to the influencing factors, such as climate factors, social activities, and seasonal factors. Climate factors depend on the temperature and humidity; social factors imply human duty activities including work, school and entertainment affecting the system load; seasonal factors then include seasonal climate change and load growth year after year.

B. Forecasting approaches in electricity loading

There is a widespread bibliography on improving the accuracy of forecasting methods, which had been proposed in the last few decades. One of these methods is a weather-insensitive approach which used historical load data to infer the future load. This approach is based on univariate time sequences and is known as Box-Jenkins Integrated autoregressive moving average (ARIMA) [3-5]. Christianse [6] and Park et al. [7] proposed exponential smoothing models by Fourier series transformation to forecast electricity load. Hence, many researchers considered to put related factors such as seasonal temperature, and day type in load forecasting models. Mbamalu and EI-Hawary [8] presented multiplicative AR models considering the seasonal factors in load forecasting. The results show that the presented model performs better than the univariate AR model in terms of forecasting accuracy. Douglas et al. [9] combined Bayesian estimation with dynamic linear model in forecasting load. The proposed model is particular suitable for predicting load with imperfect weather information. Sadownik and Barbosa [10] proposed dynamic nonlinear models in load forecasting. The disadvantage of these methods is time consuming while the number of variables is increasing.

Regression models build causal-effect relationships between the load and one or more independent variables. The most popular models are linear regression, indicated by Asbury [11]. Papalexopoulos and Hesterberg [12] added the “holiday” and “temperature” in modeling. The proposed model used weight least square method to obtain robust parameter estimation encountering with the hetero-skedasticity. Recently, Haida and Muto [13] applied transformation function to convert the temperature of the previous year in load forecasting. Unfortunately, the proposed model leads to unexpected forecasting errors caused by seasonal transition. Soliman et al. [14] applied a multivariate linear regression model in load forecasting. This model includes temperature, wind cooling/ humidity factors and outperforms the harmonic model as well as the hybrid model. In these models, the dependent variables are usually decomposed into weather-insensitive and weather-sensitive components (Bunn and Farmer [1], Park et al. [7] and Hyde and Hodnett [15]).

C. Artificial neural networks in load forecasting

In the recent decades, lots of researches had tried to apply the artificial neural networks (ANN) to improve the load forecasting accuracy. Park et al. [16] proposed a 3-layer backpropagation neural network to deal with daily load forecasting problems. The inputs include three indices of temperature: average, peak and lowest loads. The outputs are peak loads. The presented model provides more accurate forecasting results than the regression model and the time series model. Lee et at. [17] proposed an ANN model to forecast the electricity loads of weekday and weekend-day. This model does not yield less relative error than other approaches. An adaptive learning algorithm was proposed by Ho et al. [18] to forecast Taiwan electricity load in 1987. The numerical results indicate that the proposed algorithms converge faster than the traditional backpropagation learning method. Novak [19] used the radial basis function (RBF) neural networks to forecast power loading. The results show that RBF is at least 11 times faster and more reliable than the back-propagation neural networks. Darbellay and Slama [20] used an ANN to predict the electricity loading in Czech. The experimental results indicate that the proposed ANN model outperform an ARIMA with a lower normalized mean square error. Abdel-Aal [21] proposed an Abductive network to forecast hourly loads for five years by using hourly temperature and hourly load data. The results of the proposed model are very promising.

D. Support vector machines in forecasting

Unlike most of the traditional neural network models which implement the empirical risk minimization principle, the support vector machines (SVMs) implement the structural risk minimization principle which seeks to minimize an upper bound of the generalization error rather than minimize the training error. Based on this principle, SVMs achieve an optimum networks structure. The SVMs is equivalent to solving a linear constrained quadratic programming problem so that the solution of SVMs is always unique and globally optimal. Originally, SVMs have been developed for pattern recognition problems. Recently, with the introduction of Vapnik's e -insensitive loss function, support vector machines have been extended to solve nonlinear regression estimation problems. The SVMs are successfully in time series forecasting. Cao [22] used the SVMs experts for time series forecasting. A two-stage neural network architecture is contained in the generalized SVMs experts. The numerical results indicated that the SVMs experts are able to achieve the better generalization in comparison with the single SVMs models. In 2002, Cao and Gu [23] presented a dynamic SVMs model to deal with non-stationary time series problems. Experiment results showed that the DSVMs outperform standard SVMs in forecasting non-stationary time series. In the same year, Tay and Cao [24] proposed a C-ascending SVMs, to model non-stationary financial time series. Experiment results showed that the C-ascending SVMs with the actually ordered sample data consistently perform better than the standard SVMs. Tay and Cao [25] used SVMs in forecasting financial time series. The numerical results indicated that the SVMs are superior to the multi-layer back-propagation neural network in financial time series forecasting. Lu et al. [26] used support vector machines in predicting air quality parameters with different time series. The experimental results showed that the SVMs outperform the conventional Radial Basis Function networks.

According to the brief review of literature related to forecasting methods, this paper attempts to verify the forecasting accuracy improvement of SVM model. Therefore, three forecasting models based on different approaches are constructed by using Taiwan Power Company data set. Then, a forecast accuracies comparison is implementing in turn. The rest of the paper is organized as follows. The algorithms of three employed approaches are introduced in section 2. A numerical example modeling results describe in section 3. Conclusions are given in section 4.

II. METHODOLOGY

In this paper, three models are employed in Taiwan load forecasting. These approaches are autoregressive integrated moving average (ARIMA), general regression neural network (GRNN), and SVMs. The follows are a brief of the three forecasting approaches.

A. ARIMA model

Developed by Box and Jenkins [3], the ARIMA model has been one of the most popular approaches in forecasting. In an ARIMA model, the future value of a variable is supposed to be a linear combination of past values and past errors, expressed as follows:

$$y_t = q_0 + f_1 y_{t-1} + f_2 y_{t-2} + \cdots + f_p y_{t-p}$$
$$+ e_t - q_1 e_{t-1} - q_2 e_{t-2} - \cdots - q_q e_{t-q} \qquad (1)$$

where y_t is the actual value and ε_t is the random error at time t; f_i and q_i are the coefficients; p and q are integers and often referred to as autoregressive and moving average polynomials, respectively. In addition, the difference operator ∇ is considered to solve the non-stationary problem, and defined as follows:

$$\nabla y_t = y_t - y_{t-1}, \nabla^2 y_t = \nabla y_t - \nabla y_{t-1} \qquad (2)$$

Basically, three phases are included in this approach: model identification, parameter estimation and diagnostic checking. The ARIMA (2,2,1) model is specified in this paper. It can be represented as follows.

$$f_2(B)\nabla^2 y_t = q_0 + q_1(B) e_t \qquad (3)$$

where B is the backward shift operator and defined as follows:

$$B^1 y_t = y_{t-1}, B^2 y_t = y_{t-2}, \ldots\ldots, B^p y_t = y_{t-p} \qquad (4)$$

$$B^1 e_t = e_{t-1}, B^2 e_t = e_{t-2}, \ldots\ldots, B^p e_t = e_{t-p} \quad (5)$$

then $f_2(B)$ and $q_1(B)$ can be written as follows respectively.

$$f_2(B) = 1 - f_1 B^1 - f_2 B^2 \quad (6)$$

$$q_1(B) = 1 - q_1 B^1 \quad (7)$$

B. GRNN model

Developed by Specht [27], the general regression neural network (GRNN) model is capable of approximating any arbitrary function from historical data. The foundation of GRNN operation is based on the theory of kernel regression. The GRNN logic can be summarized in an equivalent nonlinear regression formula

$$E[N|M] = \frac{\int_{-\infty}^{\infty} N f(M,N) dN}{\int_{-\infty}^{\infty} f(M,N) dN} \quad (8)$$

where N is the predicted value of GRNN, M the input vector ($M_1, M_2, \cdots, M_n$) which consists of n predictor variables, $E[N|M]$ the expected value of the output *N* given an input vector *M*, and *f(M, N)* the joint probability density function of *M* and *N*.

The GRNN primarily has four layers of processing unit. Each layer of processing units is assigned with a specific computational function when nonlinear regression is performed. The first layer of the network is responsible for the reception of information. There is a unique input neuron for each predictor variable in the input vector *M*. The input neurons then feed the data to the second layer. The number of neurons in the second layer is equal to the number of cases in the training set. Therefore, the neurons in the second layer are called pattern neurons. A pattern neuron is employed to process the data in a systematic way so that the relationship between the input and the proper response is "memorized'. A multivariate Gaussian function of

$$q_i = \exp[-(M - U_i)'(M - U_i)/2s^2] \quad (9)$$

and the data from the input neurons are used to compute an output q_i by a typical pattern neuron *i*, where U_i is a specific training vector represented by pattern neuron *i*, and s is the smoothing parameter. Eq. (8) is extended by Cacoullos [28] and adopted by Specht [27] in his GRNN design.

The neurons of the third layer, namely the summation neurons, receive the outputs of the pattern neurons. In the third layer, the outputs from all pattern neurons are augmented. Basically, two kinds of summations, simple summations and weighted summations, are conducted in neurons of the third layer. The simple summation and the weighted summation operations can be represented as Eq. (10) and Eq. (11) respectively.

$$S_s = \sum_i q_i \quad (10)$$

$$S_w = \sum_i w_i q_i \quad (11)$$

where w_i is the pattern neuron *i* connected to third layer of weights.

The summations of neurons in third layer are then fed into the fourth layer. The GRNN regression output Q is calculated as follows.

$$Q = S_w / S_s \quad (12)$$

C. SVMG model

The support vector machines were proposed by Vapnik [29]. Based on the structured risk minimization principle, the SVMs seek to minimize an upper bound of the generalization error instead of the empirical error in the other neural networks. In addition, the SVM models generate the regress function by applying a set of high dimensional nonlinear functions. The nonlinear function is formulated as follows.

$$y = w_i f(x) + b \quad (13)$$

where $f(x)$ is called feature which is nonlinear mapped from the input space x. The w_i and *b* are coefficients which are estimated by minimizing the regularized risk function

$$R(C) = C \frac{1}{N} \sum_{i=1}^{N} L_e(d_i, y_i) + \frac{1}{2}\|w\|^2 \quad (14)$$

It is known that the regression estimation function is the one that minimizes the function Eq. (14) with the following e-insensitive loss function,

$$L_e(d,y) = \begin{cases} 0 & |d-y| \le e \\ |d-y| - e & otherwise \end{cases} \quad (15)$$

where both *C* and e are prescribed parameters. d_i is the actual value at period *i*. y_i is the estimation value at period *i*. The first term $L_e(d,y)$ function indicates the fact that it does not penalize errors below e. The first term $\frac{C}{N}\sum_{i=1}^{N} L_e(d_i, y_i)$ is the empirical error. The second term, $\frac{\|w\|^2}{2}$, is used as a measure of function flatness. *C* is used as the trade-off between the empirical risk (first term) and the model flatness (second term). Two positive slack variables z and z^*, which represent the distance from actual values to the corresponding boundary values of e-tube (Eq.(15)), are introduced. Then, Eq. (14) is transformed to the following constrained function:

Minimize

$$R(w, z, z^*) = \frac{1}{2} w w^T + C \sum_{i=1}^{N} (z_i + z_i^*) \quad (16)$$

Subjected to:

$w f(x_i) + b_i - d_i \le e + z_i$, i=1,2 …, N

$d_i - w f(x_i) - b_i \le e + z_i$, i=1,2 …, N

$z_i, z_i^* \ge 0$, i=1,2 …, N

Finally, by applying Karush-Kuhn-Tucker conditions for regression, Eq (16) results in a dual Lagrangian form as Eq (17).

$$R(a_i - a_i^*) = \sum_{i=1}^{N} d_i (a_i - a_i^*) - e\sum_{i=1}^{N}(a_i + a_i^*)$$

$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{N}(a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) \quad (17)$$

with the constraints:

$$\sum_{i=1}^{N}(a_i - a_i^*) = 0$$

$$0 \le a_i \le C, \quad i = 1,2,\cdots,N$$

$$0 \le a_i^* \le C, \quad i = 1,2,\cdots,N$$

In Eq (17), a_i and a_i^* are called Lagrange multipliers. They satisfy the equalities, $a_i * a_i^* = 0$

$$f(x, a, a^*) = \sum_{i=1}^{l}(a_i - a_i^*)K(x, x_i) + b \quad (18)$$

Here, $K(x,x_i)$ is called the kernel function. The value of the kernel is equal to the inner product of two vectors x and x_i in the feature space $f(x)$ and $f(x_i)$, i.e., $K(x,x_i)= f(x) * f(x_i)$. Any function that satisfies Mercer's condition by Vapink [29] can be used as the Kernel function. The Gaussian kernel function $K(x,x_i) = \exp\left(-\|x - x_i\|^2 / 2s^2\right)$ is specified in this study. In this study, the SVMs were employed to estimate the nonlinear behavior of the forecasting data set because Gaussian kernels tend to give good performance under general smoothness assumptions.

Inspired by the natural evolution process, Holland [30] proposed genetic algorithms (GAs), an organized random search technique which imitates the biological evolution process. Such algorithms are based on the principle of the survival of the fittest, which attempts to retain genetic information from generation to generation. The major advantages of GAs are the capabilities for finding optimal or near optimal solutions with relatively modest computational requirements. The procedure of GAs briefly described as follows. Step 1: Initialization. Construct randomly the initial population of chromosomes. Step 2: Evaluating fitness. Evaluate the fitness of each chromosome. In this study, the negative value normalized root mean square error measure (-NRMSE) is used as the fitness function and shown as follows.

$$Fitness\ function = -\sqrt{\frac{\sum_{i=1}^{n}(a_i - f_i)^2}{\sum_{i=1}^{n} a_i^2}} \quad (19)$$

where a_i and f_i represent the actual and forecast values correspondingly, and n is the number of forecasting period. Step 3: Selection. Select mating pair, #1 parent and #2 parent, for reproduction. Step 4: Crossover and mutation. Create new offspring by crossover and mutation operations. Step 5: Next generation. Form a population for the next generation. Step 6: Stop condition. If the number of generation is equal to a given scale, then the best chromosomes are presented as a solution, otherwise go back to Step 2.

Figure 1 shows the framework of the proposed SVMG model. GAs are used to search better combination of three parameters in SVMs so that a smaller NRMSE is obtained in each iteration of forecasting.

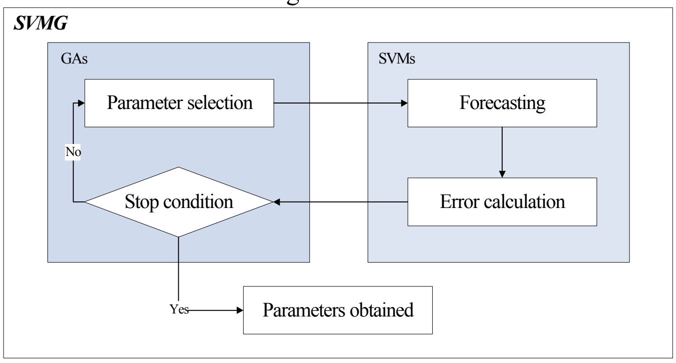

Figure 1 Framework of SVMG

D. Indices of performance evaluation

In this study, two indices, namely MAPE (mean absolute percent error) and NRMSE (normalized root mean square error measure), are used as forecasting accuracy measures.

The indices are shown as follows respectively.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{a_i - f_i}{a_i}\right| \times 100\% \quad (20)$$

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n}(a_i - f_i)^2}{\sum_{i=1}^{n} a_i^2}} \quad (21)$$

where n is the number of forecasting periods; a_i is the actual production value at period i; and f_i is the forecasting load value of Taiwan power demand at period i.

III. NUMERICAL EXAMPLE

A. Data set

In this paper, ARIMA, GRNN and SVMs are employed to forecast the future data of the Taiwan electricity demand values. The total load values from 1945 to 2003 are used as experimental data. Totally, there are 59 data of load values of Taiwan electricity demand. It is necessary to implement the forecast accuracy comparison for the three models based on the same modeling periods. Hence, the data are divided into three data sets: the training data set (40 years, from 1945 through 1984), the validation (parameter selection) data set (10 years, from 1985 through 1994), and the testing data set (9 years, from 1995 through 2003). The training data set of ARIMA model is from 1945 through 1984.

B. Parameters selection of three models

In this study, the selection of free parameters of three models plays an essential role in obtaining good forecasting results. For ARIMA models, the parameters selection is conducted by taking the second-order regular difference. By statistical packages, the most suitable model for the 40 year training data set is ARIMA(2,2,1). The equation used is presented as follows.

$$(1+1.6702B^1+0.9245B^2)\nabla^2 y_t = 414.2+(1+0.7552B^1)e_t \quad (22)$$

After the ARIMA model has been constructed, it is important to examine the fitness of the model to a given time series. To confirm this, autocorrection function (ACF) is calculated. Plot of the estimated residual ACF is showed in Figure 2, it is observed that residuals are not autocorrelated.. PACF, showed in Figure 3, is also used to check the residuals and indicates. The residuals are approximately white noise.

The kernel parameters, s , C, and e of the SVMs are determined based on the index, MAPE. The model with the minimum testing MAPE value is selected as the most suitable model for this example. The forecasting results and the parameters for the different SVMG models are illustrated in Table 1.

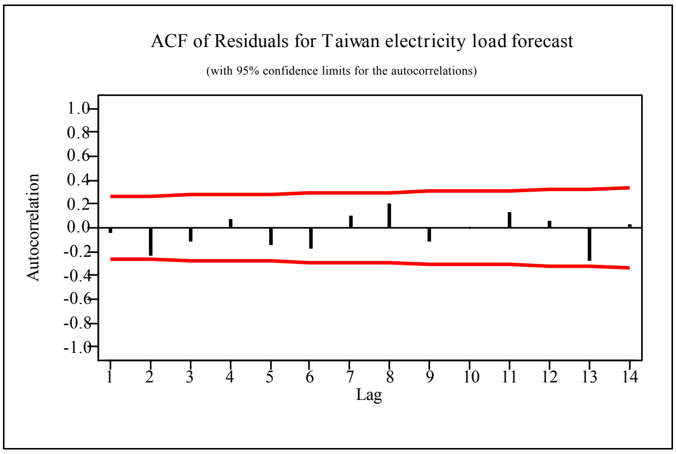

Figure 2 Estimated residual ACF

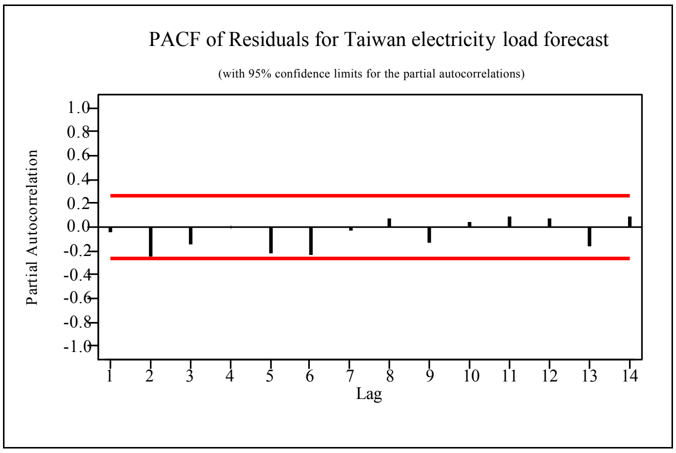

Figure 3 Estimated residual PACF

Table 2 gives the free parameters for different models. Therefore, the suitable parameters of different models are used to forecast Taiwan electricity load values in testing data set.

Table 1 Forecasting Results and Parameters of SVMG model

Numbers of input data	Parameters s	C	e	MAPE of testing (%)
30	0.4651	4.1716×10^{11}	45.602	3.80
25	**0.6627**	$\mathbf{6.7422\times10^{11}}$	**53.705**	**2.96**
20	1.4569	8.1699×10^{11}	72.121	5.06
15	1.0024	2.9579×10^{11}	98.158	4.98

Table 2 Suitable values of parameters for different models

Models	Suitable parameter combinations
ARIMA	p=2, d=2, q=1
GRNN	s =0.04
SVMG	s =0.6627, C=6.7422×10^{11}, e =53.705

C. Forecasting results

Figure 4 shows real values as well as forecasting values of different models. Performance of three forecasting models is listed in Table3. The numerical results indicate that the SVMs are superior to the ARIMA(2,2,1) and GRNN models in terms of forecasting accuracy.

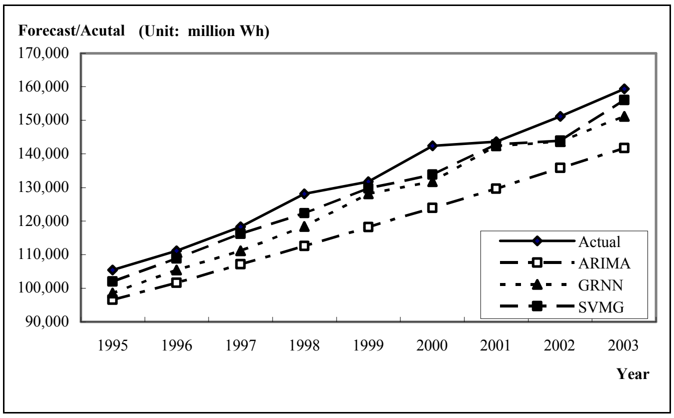

Figure 4 Forecasting values for different models

Table 3 Forecasting results based on forecasting indices

	MAPE	NRMSE
ARIMA(2,2,1)	10.31%	0.105997
GRNN (spread=0.04)	5.18%	0.054732
SVMG(0.6627, 6.7422×10^{11}, 53.705)	2.96%	0.035016

IV. Conclusions

The electricity supply is the most important infrastructure to assist social activities operation for a regional or a national economic system. Especially, for an export-oriented economy, like Taiwan, more accurate load forecasting is able to provide more advantages in saving and distributing limited resource efficiently. From the historical data, the Taiwan electricity demand values show a strong growth trends. In this paper, we employed a novel forecasting technique, SVMG, to examine its feasibility in forecasting annual electricity loads. Two other forecasting approaches, ARIMA and GRNN, are used to compare the forecasting performance. Experiment results indicate that the proposed SVMG model outperforms the other approaches in terms of forecasting accuracy.

REFERENCES

[1] D.W.Bunn and E.D.Farmer. *Comparative models for Electrical Load Forecasting*, John Wiley & Sons, New York, 1985.

[2] D.W.Bunn. Forecasting loads and prices in competitive power markets, *Proceedings of the IEEE*, Vol. 88, pp. 163-169, 2000.

[3] G.E.P.Box and G.M.Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, CA, 1970.

[4] S.Vemuri, D.Hill and R.Balasubramanian. Load forecasting using stochastic models, *in Proceeding 8th Power Industrial Computing Application Conference*, pp. 31-37, 1973.

[5] J.F.Chen, W.M.Wang and C.M.huang. Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting, *Electric Power Systems Research*, Vol. 34, pp. 187-196, 1995.

[6] W.R.Christianse. Short term load forecasting using general exponential smoothing, *IEEE Transactions on Power Apparatus and System*, PAS-90, pp. 900-911, 1971.

[7] J.H.Park, Y.M.Park and K.Y.Lee. Composite modeling for adaptive short-term load forecasting, *IEEE Transactions on Power Systems*, Vol. 6, pp. 450-457, 1991.

[8] G.A.N.Mbamalu and M.E.El-Hawary. Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation, *IEEE Transactions on Power Systems*, Vol. 8, pp. 343-348, 1993.

[9] A.P.Douglas, A.M.Breipohl, F.N.Lee and R.Adapa. The impact of temperature forecast uncertainty on Bayesian load forecasting, *IEEE Transactions on Power Systems*, Vol. 13, pp. 1507-1513, 1998.

[10] R.Sadownik and E.P.Barbosa. Short-term forecasting of industrial electricity consumption in Brazil, *Journal of Forecasting*, Vol. 18, pp. 215-224, 1999.

[11] C.Asbury. Weather load model for electric demand energy forecasting, *IEEE Transactions Power Apparatus and System*, PAS-94, pp. 1111-1116, 1975.

[12] A.D.Papalexopoulos and T.C.Hesterberg. A regression-based approach to short-term system load forecasting, *IEEE Transactions on Power Systems*, Vol. 5, pp. 1535-1547, 1990.

[13] T.Haida and S.Muto. Regression based peak load forecasting using a transformation technique, *IEEE Transactions on Power Systems*, Vol. 9, pp. 1788-1794, 1994.

[14] S.A.Soliman, S.Persaud, K.El-Nagar and M.E. El-Hawary. Application of least absolute value parameter estimation based on linear programming to short-term load forecasting, *International Journal of Electrical Power and Energy systems*, Vol. 19,pp. 209-216, 1997.

[15] O.Hyde and P.F.Hodnett. An adaptable automated procedure for short-term electricity load forecasting, *IEEE Transactions on Power Systems*, Vol. 12,pp. 84-93, 1997.

[16] D.C.Park, M.A.El-Sharkawi, R.J.Marks II, L.E.Atlas and M.J.Damborg. Electric load forecasting using an artificial neural network, *IEEE Transactions on Power Systems*, Vol. 6, pp. 442-449, 1991.

[17] K.Y.Lee, Y.T.Cha and J.H.Park. Short-term load forecasting using an artificial neural network, *IEEE Transactions on Power Systems*, Vol. 7, pp. 124-132, 1992.

[18] K.L.Ho, Y.Y.Hsu and C.C.Yang. Short-term load forecasting using a multilayer neural network with an adaptive learning algorithm, *IEEE Transactions on Power Systems*, Vol. 7, pp. 141-149, 1992.

[19] B.Novak. Superfast autoconfiguring artificial neural networks and their application to power systems, *Electric Power Systems research*, Vol. 35, pp. 11-16, 1995.

[20] G.A.Darbellay and M.Slama. Forecasting the short-term demand for electricity—do neural networks stand a better chance? *International Journal of Forecasting*, Vol. 16, pp. 71-83, 2000.

[21] R.E.Abdel-Aal. Short-term hourly load forecasting using abductive networks, *IEEE Transactions on Power Systems*, Vol. 19, pp. 164-173, 2004.

[22] L.Cao. Support vector machines experts for time series forecasting, *Neurcomputing*, Vol. 51, pp. 321-339, 2003.

[23] L.Cao and Q.Gu. Dynamic support vector machines for non-stationary time series forecasting, *Intelligent Data Analysis*, Vol. 6, pp. 67-83, 2002.

[24] F.E.H. Tay, L.Cao. Modified support vector machines in financial time series forecasting, *Neurcomputing*, Vol. 48, pp. 847-861, 2002.

[25] F.E.H. Tay and L.Cao. Application of support vector machines in financial time series forecasting. Omega Vol. 29, pp. 309-317, 2001.

[26] W.Lu, W.Wang, A.Y.T.Leung, S.M.Lo, R.K.K.Yuen, Z.Xu and H.Fan. Air pollutant parameter forecasting using support vector machines, In: IJCNN '02 Proceedings of the 2002 International Joint Conference on , Vol. 1, pp. 12-17, 2002.

[27] D.A.Specht. A general regression neural network, *IEEE Transactions on Neural Networks*, Vol. 2, pp. 568-576, 1991.

[28] T.Cacoullos. Estimation of a multivariate density, *Annals of the Institute of Statistical Mathematics*, Vol. 18, pp. 179-189, 1966.

[29] V.Vapnik. *The nature of statistic learning theory*, Springer-Verlag, New York, 1995.

[30] J.H.Holland. Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.