Available online at www.sciencedirect.com
SCIENCE DIRECT®
Energy Conversion and Management 46 (2005) 2669–2688
ENERGY CONVERSION & MANAGEMENT
www.elsevier.com/locate/enconman

Support vector machines with simulated annealing algorithms in electricity load forecasting

Ping-Feng Pai [a,*], Wei-Chiang Hong [b]

[a] *Department of Information Management, National Chi Nan University, 1 University Road, Puli, Nantou, Taiwan, ROC*
[b] *School of Management, Da-Yeh University, 112 Shan-Jiau Road, Da-Tusen, Chang-Hua, 51505 Taiwan, ROC*

Received 26 August 2004; received in revised form 27 August 2004; accepted 25 February 2005
Available online 8 April 2005

Abstract

Accurate forecasting of electricity load has been one of the most important issues in the electricity industry. Recently, along with power system privatization and deregulation, accurate forecast of electricity load has received increasing attention. Because of the general nonlinear mapping capabilities of forecasting, artificial neural networks have played a crucial role in forecasting electricity load. Support vector machines (SVMs) have been successfully employed to solve nonlinear regression and time series problems. However, SVMs have rarely been applied to forecast electricity load. This investigation elucidates the feasibility of using SVMs to forecast electricity load. Moreover, simulated annealing (SA) algorithms were employed to choose the parameters of a SVM model. Subsequently, examples of electricity load data from Taiwan were used to illustrate the proposed SVMSA (support vector machines with simulated annealing) model. The empirical results reveal that the proposed model outperforms the other two models, namely the autoregressive integrated moving average (ARIMA) model and the general regression neural networks (GRNN) model. Consequently, the SVMSA model provides a promising alternative for forecasting electricity load.

Keywords: Support vector machines (SVMs); Simulated annealing algorithms (SA); General regression neural networks (GRNN); Autoregressive integrated moving average (ARIMA); Electricity load forecasting

* Corresponding author. Tel.: +886 49 2910 960x4871; fax: +886 49 2915 205.
E-mail address: paipf@ncnu.edu.tw (P.-F. Pai).

0196-8904/$ - see front matter
doi:10.1016/j.enconman.2005.02.004

1. Introduction

1.1. Electricity load management

Along with the recent privatization and deregulation of the electricity industry, the reliance and accuracy of future electricity demand forecasting have received growing attention, particularly in the areas of electricity load planning, energy expenditure/cost economy and secure operation fields, in regional and/or national systems. For electricity load reliance, electricity providers face increasing competition in the demand market and must pay increased attention to electricity quality, including unit commitment, hydrothermal coordination, short-term maintenance, interchange and transaction evaluation, network power flow dispatched optimization and security strategies. On the other hand, inaccurate electricity load forecasting may increase operating costs [1–5]. Bunn and Farmer [4] pointed out that a 1% increase in forecasting error implied a £10 million increase in operating costs. Hence, over estimation of future load results in unnecessary spinning reserve and, furthermore, is not accepted by international energy networks owing to excess supply. In contrast, under estimation of load causes failure in providing sufficient reserve and implies high costs in the peaking unit. Because buying at the last minute from other suppliers is expensive, international electricity production cooperation requires accurate forecasting of the needs of all participants. However, forecasting the electricity load is difficult, primarily due to the various influences, such as climate factors, social activities, and seasonal factors.

1.2. Approaches of electricity load forecasting

During recent decades, numerous investigations have been proposed to improve the accuracy of electricity load forecasting. One such method is a weather insensitive approach that uses historical load data to forecast future electricity load. Generally, it is known as the Box–Jenkins autoregressive integrated moving average (ARIMA) [6–8]. Christianse [9] and Park et al. [10] designed exponential smoothing models by Fourier series transformation for electricity load forecasting. Hence, many researchers considered related factors, such as seasonal temperature and day type, in load forecasting models. Mbamalu and El-Hawary [11] proposed multiplicative autoregressive (AR) models that considered seasonal factors in load forecasting. The analytical results showed that the forecasting accuracy of the proposed models outperformed the univariate AR model. Douglas et al. [12] considered verifying the impacts of temperature on the forecasting model. The authors combined Bayesian estimation with a dynamic linear model for load forecasting. The experimental results demonstrated that the presented model is suitable for forecasting load under imperfect weather information. Sadownik and Barbosa [13] proposed dynamic nonlinear models for load forecasting. The main disadvantage of these methods is that they become time consuming to compute as the number of variables increases. To achieve accurate load forecasting, the load forecasting model employed state space and Kalman filtering technology, which were developed to reduce the difference between the actual and predicted loads. The state space and Kalman filtering technology introduced a periodic component of the load as a random process and requires historical data covering a period exceeding 3–10 years to establish the periodic load variation for estimating

the dependent variables (load or temperature) of the power system [14,15]. Moghram and Rahman [16] devised a model based on the state space and Kalman filtering technology, and verified that the proposed model outperformed four other forecasting methods. Similarly, Park et al. [10] proposed a load forecasting model based on the state space and Kalman filtering technology and also showed that their model outperformed other methods. The disadvantage of these methods is the difficultly of avoiding observation noise in the forecasting.

The regression approach is another popular model for forecasting electricity load. Regression models construct the cause–effect relationships between electricity load and the independent variables. The most popular models are linear regression, proposed by Asbury [17], which considers the "weather" variable in the forecasting model. Meanwhile, Papalexopoulos and Hesterberg [18] added "holiday" and "temperature" to the model. The proposed model used the weighted least square method to obtain robust parameter estimation. Furthermore, Soliman et al. [19] designed a multivariate linear regression model in load forecasting. This model includes temperature, wind cooling and humidity factors. The empirical results indicated that the proposed model outperformed the harmonic model and the hybrid model. In these models, the dependent variables were generally decomposed into weather insensitive and weather sensitive components [4,10,20]. These models were assumed to be linear and are computationally intensive. However, these independent variables were not justified for use because the terms are known to be nonlinear.

1.3. Artificial neural networks in load forecasting

Recently, many works have attempted to apply artificial intelligence techniques for enhancing electricity load forecasting performance. Rahman and Bhatnagar [21] presented a knowledge based expert system (KBES) approach for electricity load forecasting. The KBES model constructed new rules based on received information, including daily temperature, day type, load from the previous day and so on. This approach was derived from training rules and transformed the information into mathematical equations [22,23]. Park et al. [24] established a three layer back propagation neural network to implement daily load forecasting problems. The inputs comprise three temperature indices: average, peak and lowest loads, while the output is the peak load. The proposed model outperformed both the regression model and the time series model in terms of mean absolute percent error (MAPE). Moreover, Ho et al. [25] developed an adaptive learning algorithm for forecasting the electricity load in Taiwan. The numerical results demonstrated that the proposed algorithm converges faster than the traditional back propagation learning method. Novak [26] applied radial basis function (RBF) neural networks to forecast electricity load. The analytical results indicated that the RBF network is at least 11 times faster and more reliable than the back propagation neural networks. Furthermore, Darbellay and Slama [27] applied ANNs to predict the Czech electricity load. The experimental results showed that the proposed ANN model outperformed the ARIMA model in terms of normalized mean square error. Abdel-Aal [28] proposed an abductive network to conduct one hour ahead load forecasts for a five year period. The proposed model achieved extremely promising results based on the measurement of mean absolute percent error. Hsu and Chen [29] employed back propagation neural networks to forecast the regional load in Taiwan. The experiment results showed that the artificial neural network approach outperformed the regression models.

1.4. Support vector machines in forecasting

The support vector machines (SVMs) implement the structural risk minimization (SRM) principle rather than the empirical risk minimization principle implemented by most of the traditional neural network models. The most important concept of SRM is the application of minimizing an upper bound to the generalization error instead of minimizing the training error. Based on this principle, SVMs achieve an optimum networks structure. In addition, the SVMs will be 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 applied for pattern recognition problems. However, along with the introduction of Vapnik's ε insensitive loss function, SVMs also have been extended to solve nonlinear regression estimation problems. Therefore, SVMs are successful in time series forecasting. Cao [30] used the SVMs experts for time series forecasting. The generalized SVMs experts contained a two stage neural network architecture. The numerical results indicated that the SVMs experts are capable of outperforming the single SVMs models in terms of generalization comparison. Cao and Gu [31] proposed a dynamic SVMs model to deal with non-stationary time series problems. Experimental results showed that the dynamic SVMs model outperform standard SVMs in forecasting non-stationary time series. Meanwhile, Tay and Cao [32] presented C-ascending SVMs to model non-stationary financial time series. Experimental results showed that the C-ascending SVMs with actually ordered sample data consistently perform better than standard SVMs. Tay and Cao [33] 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. [34] applied SVMs in predicting air quality parameters with different time series. The experimental results indicated that SVMs outperform the conventional radial basis function networks.

This investigation attempts to examine the improvement in forecasting accuracy achieved using the SVMSA model. Two other forecasting approaches, namely the GRNN model and the ARIMA model, consequently were employed to compare the forecasting accuracy of electricity load.

2. Forecasting models

In this study, three models, namely the autoregressive integrated moving average (ARIMA) model, the general regression neural network (GRNN) model and the support vector machines with simulated annealing (SVMSA) model, were employed to forecast the Taiwanese electricity load. The introduction of the three models is as follows.

Introduced by Box and Jenkins [6], 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:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_{t-1}\theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \quad (1)$$

where y_t is the actual value and ε_t is the random error at time t; ϕ_i and θ_j are the coefficients; and p and q are integers. Eqs. (1) are often referred to as autoregressive and moving average

polynomials, respectively. In addition, the difference (∇) is used to solve the non-stationary problem, and it is defined as follows:

$$\nabla^d y_t = \nabla^{d-1} y_t - \nabla^{d-1} y_{t-1} \tag{2}$$

Basically, three phases are included in an ARIMA model: model identification, parameter estimation and diagnostic checking. Furthermore, the backward shift operator, B, is defined as follows:

$$B^1 y_t = y_{t-1}, B^2 y_t = y_{t-2}, \ldots, B^p y_t = y_{t-p} \tag{3}$$

$$B^1 \varepsilon_t = \varepsilon_{t-1}, B^2 \varepsilon_t = \varepsilon_{t-2}, \ldots, B^p \varepsilon_t = \varepsilon_{t-p} \tag{4}$$

then $\phi_p(B)$ and $\theta_q(B)$ can be written, respectively, as follows:

$$\phi_p(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_p B^p \tag{5}$$

$$\theta_1(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \cdots - \theta_q B^q \tag{6}$$

Hence, Eq. (1) can be rewritten as Eq. (7),

$$\phi_p(B)\nabla^d y_t = C_0 + \theta_q(B)\varepsilon_t \tag{7}$$

Eq. (7) is denoted as ARIMA(p, d, q) with non-zero constant, C_0. For example, the ARIMA (2, 2, 1) model can be represented as Eq. (8):

$$\phi_2(B)\nabla^2 y_t = C_0 + \theta_1(B)\varepsilon_t \tag{8}$$

In general, the values of p, d and q then need to be estimated by an autocorrelation function (ACF) and a partial autocorrelation function (PACF) of the differenced series.

The general regression neural network (GRNN) model, proposed by Specht [35], can approximate any arbitrary function from historical data. The foundation of the GRNN operation is based on the theory of kernel regression. The procedure of the GRNN model can be equivalently represented as follows:

$$E[N|M] = \frac{\int_{-\infty}^{\infty} N f(M,N)\mathrm{d}N}{\int_{-\infty}^{\infty} f(M,N)\mathrm{d}N} \tag{9}$$

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

The GRNN primarily has four layers (Fig. 1). Each layer is assigned a specific computational function within which the nonlinear regression function, Eq. (10), is performed. The first layer of the network is to receive information. The input neurons then feed the data to the second layer. The primary task of the second layer is to memorize the relationship between the input neuron and its proper response. Therefore, the neurons in the second layer are also called pattern neurons. A multivariate Gaussian function of θ_i is given in Eq. (10), and the data from the input neurons are used to compute an output θ_i by a typical pattern neuron i as

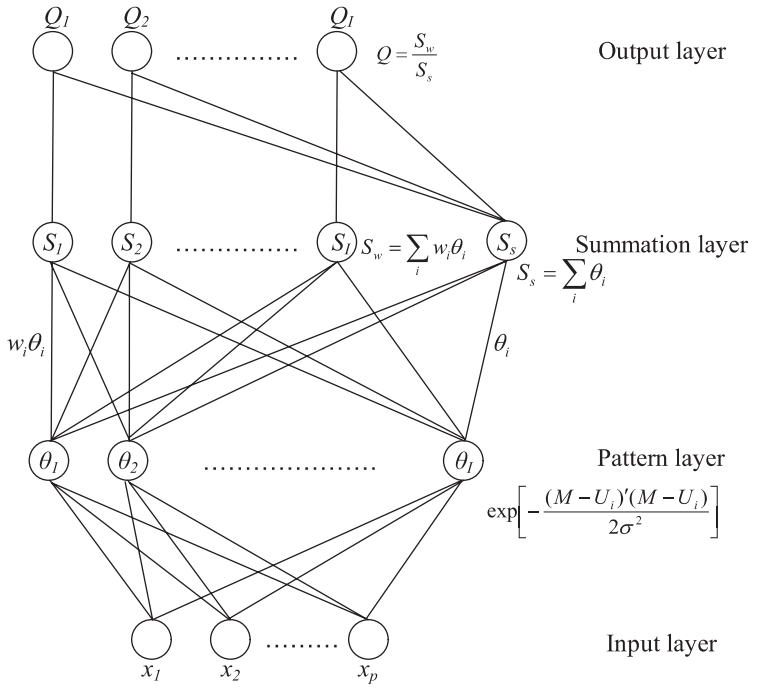

Fig. 1. The architecture of the GRNN model.

$$\theta_i = \exp\left[\frac{-(M-U_i)'(M-U_i)}{2\sigma^2}\right] \tag{10}$$

where U_i is a specific training vector represented by pattern neuron i and σ is the smoothing parameter. In the third layer, the neurons, namely the summation neurons, receive the outputs of the pattern neurons. The outputs from all pattern neurons are augmented. Basically, two summations, the simple summation and the weighted summation, are conducted in the neurons of the third layer. The simple summation and the weighted summation operations can be represented as Eqs. (11) and (12), respectively:

$$S_s = \sum_i \theta_i \tag{11}$$

$$S_w = \sum_i w_i \theta_i \tag{12}$$

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

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

$$Q = \frac{S_s}{S_w} \tag{13}$$

Following the introduction of the ε insensitive loss function, SVMs have been used to solve nonlinear regression problems [36,37]. The basic concept of the SVM regression is to map nonlinearly the original data x into a higher dimensional feature space. Hence, given a set of data $G = \{(x_i, a_i)\}_{i=1}^{N}$ (where x_i is the input vector; a_i is the actual value and N is the total number of data patterns), the SVM regression function is

$$y = f(x) = w_i \psi_i(x) + b \tag{14}$$

where $\psi_i(x)$ is called the feature that is nonlinearly mapped from the input space **x**. The w_i and b are coefficients that are estimated by minimizing the regularized risk function

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N}\Theta_{\varepsilon}(d_i, y_i) + \frac{1}{2}\|w\|^2 \tag{15}$$

where

$$\Theta_{\varepsilon}(a, y) = \begin{cases} 0 & |d - y| \leqslant \varepsilon \\ |d - y| - \varepsilon & \text{otherwise} \end{cases} \tag{16}$$

and C and ε are prescribed parameters. In Eq. (15), $\Theta_{\varepsilon}(d, y)$ is called the ε insensitive loss function. The loss equals zero if the forecasted value is within the ε tube (Eq. (16) and Fig. 2). The second term, $\frac{1}{2}\|w\|^2$, measures the flatness of the function. Therefore, C is considered to specify the trade off between the empirical risk and the model flatness. Both C and ε are user determined parameters. Two positive slack variables ξ and ξ^*, which represent the distance from the actual values to the corresponding boundary values of ε tube (Fig. 2), are introduced. Then, Eq. (15) is transformed into the following constrained form: Minimize

$$R(w, \xi, \xi^*) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right) \tag{17}$$

with the constraints,

$$w_i\psi(x_i) + b_i - d_i \leqslant \varepsilon + \xi_i^*, \quad i = 1, 2 \ldots, N$$

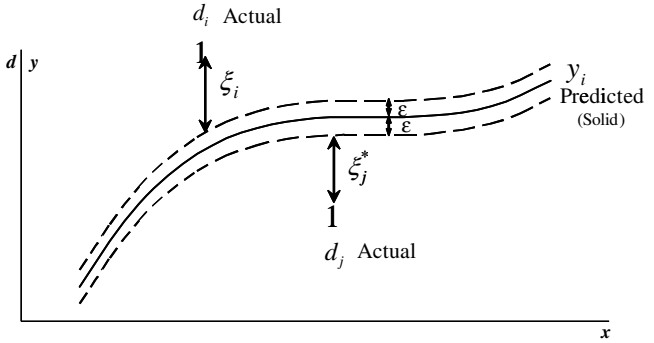

Fig. 2. Parameters used in support vector regression [38].

$$d_i - w_i\psi(x_i) - b_i \leqslant \varepsilon + \xi_i, \quad i = 1, 2 \ldots, N$$

$$\xi_i, \xi_i^* \geqslant 0, \quad i = 1, 2 \ldots, N$$

This constrained optimization problem is solved using the following primal Lagrangian form:

$$L(w_i, b, \xi, \xi^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right) - \sum_{i=1}^{N}\alpha_i[w_i\psi(x_i) + b - d_i + \varepsilon + \xi_i] - \sum_{i=1}^{N}\alpha_i^*[d_i - w_i\psi(x_i) - b + \varepsilon + \xi_i^*] - \sum_{i=1}^{N}(\beta_i\xi_i + \beta_i^*\xi_i^*) \tag{18}$$

Eq. (18) is minimized with respect to the primal variables w_i, b, ξ and ξ^*, and maximized with respect to the non-negative Lagrangian multipliers α_i, α_i^*, β_i and β_i^*. Finally, the Karush–Kuhn–Tucker conditions [39], which is applied to obtain an optimal solution of Eq. (18), are applied to the regression, and Eq. (17), thus, yields the dual Lagrangian,

$$\vartheta(\alpha_i, \alpha_i^*) = \sum_{i=1}^{N} a_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \tag{19}$$

subject to the constraints,

$$\sum\nolimits_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0$$

$$0 \leqslant \alpha_i \leqslant C, \quad i = 1, 2, \ldots, N$$

$$0 \leqslant \alpha_i^* \leqslant C, \quad i = 1, 2, \ldots, N$$

The Lagrange multipliers in Eq. (19) satisfy the equality $\alpha_i^*\alpha_i^* = 0$. The Lagrange multipliers α_i and α_i^* are calculated and an optimal desired weight vector of the regression hyperplane is

$$w^* = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x, x_i) \tag{20}$$

Hence, the regression function is Eq. (21):

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x, x_i) + b \tag{21}$$

Here, $K(x, x_i)$ is called the kernel function. The value of the kernel is equal to the inner product of the two vectors x and x_i in the feature space $\psi(x)$ and $\psi(x_i)$, i.e. $K(x, x_i) = \psi(x) * \psi(x_i)$. Any function that satisfies Mercer's condition [40,41] can be used as the kernel function. The matrix K, $\mathbf{K} = (K(x, x_j))_{j=1}^{N}$, satisfies Mercer's condition when all its eigenvalues are greater than or equal to zero. In this work, the Gaussian function, $\exp\left(-\frac{1}{2} * \left(\frac{\|x_i - x_j\|}{\sigma}\right)^2\right)$, is used in the SVMs.

The selection of the three positive parameters, σ, ε and C, of a SVM model is important to the accuracy of the forecasting. However, structural methods for confirming the selection of parameters efficiently are lacking. Therefore, a simulated annealing algorithm (SA) is used in the proposed SVM model to optimize the parameter selection.

3. Simulated annealing algorithms for parameter selection of SVM model

The simulated annealing algorithm is an optimization technique, analogous to the annealing process of material physics. Boltzmann [42] pointed out that if the system is in thermal equilibrium at a temperature T, then the probability $P_T(s)$ of the system being in a given state s is given by the Boltzmann distribution:

$$P_T(s) = \frac{\exp(-E(s)/kT)}{\sum_{w \in S} \exp(-E(w)/kT)} \tag{22}$$

where $E(s)$ denotes the energy of state s; k represents the Boltzmann constant and S is the set of all possible states. However, Eq. (22) does not contain information on how a fluid reaches thermal equilibrium at a given temperature. Metropolis et al. [43] developed an algorithm that simulates the process of Boltzmann. The Metropolis algorithm is summarized as follows. When the system is in the original state s_{old} with energy $E(s_{\text{old}})$, a randomly selected atom is perturbed, resulting in a state s_{new} with energy $E(s_{\text{new}})$. This new state is either accepted or rejected depending on the Metropolis criterion: if $E(s_{\text{new}}) \leqslant E(s_{\text{old}})$ then the new sate is automatically accepted. In contrast, if $E(s_{\text{new}}) > E(s_{\text{old}})$, then the probability of accepting the new state is given by the following probability function:

$$P(\text{ accept } \quad s_{\text{new}}) = \exp\left(-\frac{E(s_{\text{old}}) - E(s_{\text{new}})}{kT}\right) \tag{23}$$

Based on the study of Boltzmann and Metropolis, Kirkpatrick et al. [44] claimed that the Metropolis approach is conducted for each temperature on the annealing schedule until thermal equilibrium is reached. Additionally, a prerequisite for applying the SA algorithm is that a given set of the multiple variables defines a unique system state for which the objective function can be calculated. The SA algorithm in our investigation is described as follows:

Step 1 (Initialization). Set upper bounds of the three SVM positive parameters, σ, C and ε. Then, generate and feed the initial values of the three parameters into the SVM model. The absolute value of the forecasting error is defined as the system state (E). Here, the initial state (E_0) is obtained.

Step 2 (Provisional state). Make a random move to change the existing system state to a provisional state. Another set of the three positive parameters are generated in this stage.

Step 3 (Acceptance tests). The following equation is employed to determine the acceptance or rejection of the provisional state [45]:

$$\begin{cases} \text{Accept the provisional state} & \text{if } E(s_{\text{new}}) > E(s_{\text{old}}), \text{ and } p < P(\text{accept} \quad s_{\text{new}}), \quad 0 \leqslant p \leqslant 1. \\ \text{Accept the provisional state} & \text{if } E(s_{\text{new}}) \leqslant E(s_{\text{old}}) \\ \text{Reject the provisional} & \text{otherwise} \end{cases} \tag{24}$$

In Eq. (24), p is a random number to determine the acceptance of the provisional state. If the provisional state is accepted, then set the provisional state as the current state.

Step 4 (Incumbent solutions). If the provisional state is not accepted, then return to Step 2. Furthermore, if the current state is not superior to the system state, then repeat Steps 2 and 3 until the current state is superior to the system state and, finally, set the current state as the new system state. Previous studies [44,45] indicated that the maximum number of loops (N_{sa}) is $100d$ to avoid infinitely repeated loops, where d denotes the problem dimension. In this investigation, the three parameters (σ, C and ε) are used to determine the system states, hence, N_{sa} is set to 300.

Step 5 (Temperature reduction). After the new system state is obtained, reduce the temperature. The new temperature reduction is obtained by Eq. (25):

$$\text{New temperature} = (\text{Current temperature}) \times \rho, \quad \text{where } 0 < \rho < 1. \tag{25}$$

The ρ is set at 0.9 in this study [46]. If the pre-determined temperature is reached, then stop the algorithm, and the latest state is an approximate optimal solution. Otherwise, go to Step 2.

In the investigation, the value of the mean absolute percent error (MAPE), shown as Eq. (26), serves as the criterion for identifying suitable parameters for use in the SVMSA model.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{d_i - y_i}{d_i}\right| \times 100\% \tag{26}$$

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

Fig. 3 illustrates the framework of the proposed SVMSA model. The SA algorithm is used to seek a better combination of the three parameters in the SVMs so that a smaller MAPE is obtained in each forecasting iteration.

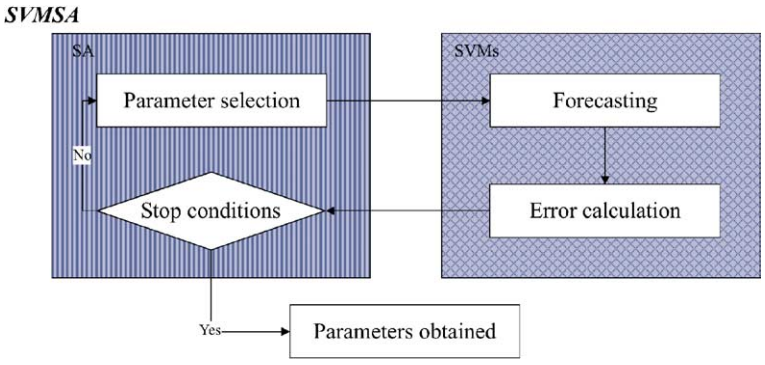

Fig. 3. Framework of SVMSA.

4. A numerical example

4.1. Data set and indices of performance evaluation

This investigation used Taiwanese electricity load data to compare the forecasting performances of the SVMSA, ARIMA and GRNN models. The total load values from 1945 to 2003 serve as experimental data. Totally, 59 load data for the Taiwanese electricity load are available, as listed in Table 1. The forecasting accuracy of the three models is compared based on the same modeling periods. The data are divided into three data sets: the training data set (40 years, from 1945 to 1984), the validation data set (10 years, from 1985 to 1994) and the testing data set (9 years, from 1995 to 2003). The three data sets are listed in Table 2. Schalkoff [47] recommended that the ratio of validation data to training data should be approximately one to four. The

Table 1
Electricity load of Taiwan (from 1945 to 2003) (unit: Wh)

Year	Real load values	Year	Real load values	Year	Real load values
1945	279,724,731	1965	5,672,212,374	1985	47,919,101,822
1946	304,572,759	1966	6,480,780,352	1986	53,812,861,713
1947	432,055,089	1967	7,469,753,968	1987	59,174,751,488
1948	631,346,909	1968	8,762,154,199	1988	65,227,727,033
1949	605,087,927	1969	10,050,753,666	1989	69,251,809,494
1950	717,017,534	1970	11,963,647,557	1990	74,344,947,489
1951	955,316,317	1971	13,836,002,258	1991	80,977,405,302
1952	1,076,022,820	1972	16,080,808,891	1992	85,290,353,628
1953	1,224,826,874	1973	17,938,474,183	1993	92,084,684,170
1954	1,402,548,324	1974	18,880,734,237	1994	98,561,004,240
1955	1,496,666,215	1975	21,216,825,821	1995	105,368,193,346
1956	1,769,908,228	1976	24,744,064,017	1996	111,139,816,430
1957	2,084,424,666	1977	27,607,246,583	1997	118,299,046,354
1958	2,416,369,727	1978	31,864,361,567	1998	128,129,801,081
1959	2,769,844,171	1979	35,320,579,327	1999	131,725,891,697
1960	3,135,727,261	1980	37,947,084,699	2000	142,412,887,053
1961	3,528,129,419	1981	37,447,713,001	2001	143,623,579,667
1962	4,065,485,183	1982	38,155,718,006	2002	151,192,689,981
1963	4,366,678,735	1983	42,318,624,633	2003	159,379,855,467
1964	5,185,327,196	1984	45,826,958,088		

Table 2
Training, validation, and testing data sets of the SVMSA model

Data sets	Periods
Training data	1945–1984
Validation data	1985–1994
Testing data	1995–2003

accuracy of proposed forecasting models is measured by the mean absolute percentage error (MAPE), mean absolute deviation (MAD) and normalized root mean square error measure (NRMSE), as given by Eqs. ()()()(26)–(28), respectively:

$$\mathrm{MAD} = \frac{\sum_{i=1}^{n} |d_i - y_i|}{n} \tag{27}$$

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{i=1}^{n} (d_i - y_i)^2}{\sum_{i=1}^{n} d_i^2}} \tag{28}$$

where n is the number of forecasting periods; d_i is the actual production value at period i; and y_i is the forecasting load value of the Taiwan electricity load at period i.

4.2. Parameter determination of the three models

In this investigation, the free parameters of the three models are essential to obtain good forecasting results. For ARIMA models, the statistical package identified the most suitable model for the training data as ARIMA(2, 2, 1). The ARIMA(2, 2, 1) model can be expressed as follows:

$$(1 + 1.6702B^1 + 0.9245B^2)\nabla^2 y_t = 414.2 + (1 + 0.7552B^1)\varepsilon_t \tag{29}$$

After determining the suitable parameters of the ARIMA model, it is important to examine how closely the proposed model fits a given time series. The autocorrelation function (ACF) was calculated to verify the parameters. Fig. 4 plots the estimated residual ACF and indicates that the residuals are not autocorrelated. PACF, the partial autocorrelation function, displayed in Fig. 5, is also used to check the residuals and indicates that the residuals are not correlated.

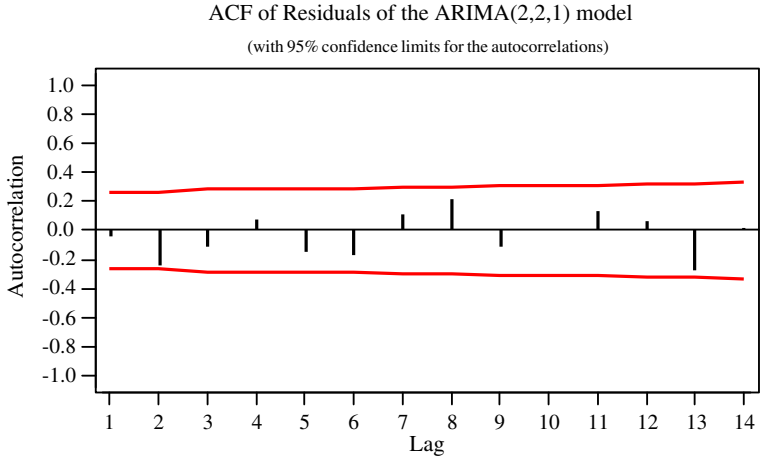

Fig. 4. Estimated residual ACF.

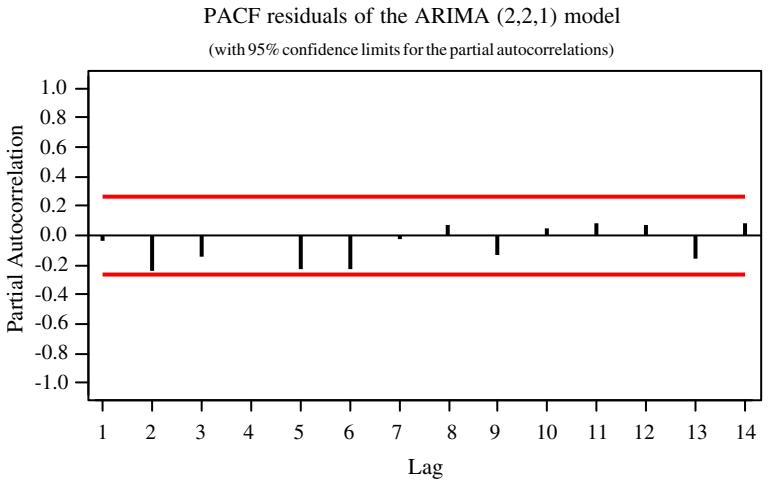

Fig. 5. Estimated residual PACF.

Table 3
Forecasting results and associated parameters of the SVMSA model

Numbers of input data	Parameters			MAPE of testing (%)
	σ	C	ε	
30	**0.2707**	**2.8414×10^{11}**	**39.127**	**1.760**
25	0.5404	4.7322×10^{11}	82.807	3.155
20	0.8000	6.3999×10^{11}	0.584	2.886
15	1.2500	8.3232×10^{11}	98.001	2.977

For the SVMSA model, several types of data rolling were applied during the training stage to conduct the rolling based forecasting procedure. Different numbers of the electricity load in a time series were fed into the SVMSA model to forecast the electricity load during the next year. The kernel parameters σ, C and ε in the SVMSA model with the minimum testing MAPE value were determined as the most suitable model for the present example. Table 3 lists the forecasting results and the suitable parameters for the different SVMSA models. The SVMSA model performs best when using 30 input data.

Fig. 6 shows the MAPE values of the GRNN with various σ. Clearly, when σ exceeds 0.41, the value of MAPE subsequently also increases. Therefore, the limit of σ is 0.41. In this study, the value of σ was set at 0.04. Table 4 lists the free parameters of the different models. These suitable parameters of the different models were used for forecasting the Taiwan electricity load in the testing data set.

4.3. Forecasting results

In this study, MAPE, MAD and NRMSE, which are given by Eqs. (26)–(28), were used as measurements of the forecasting accuracy. Table 5 lists the MAPE, MAD and NRMSE values of the

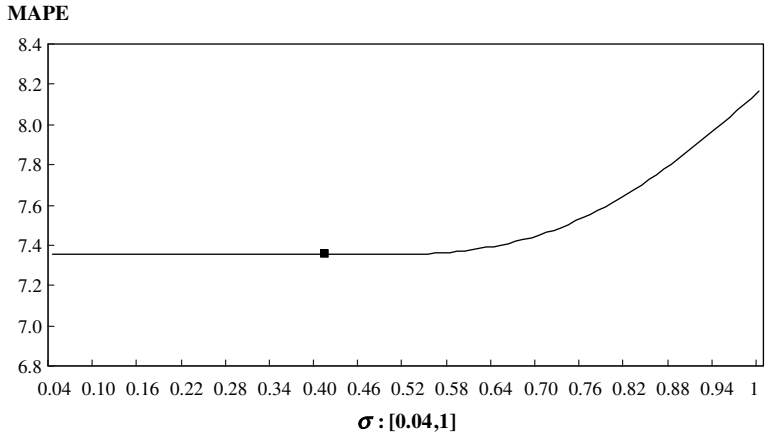

Fig. 6. MAPE with various σ values of GRNN model.

Table 4
Suitable values of parameters for different models

Models	Suitable parameter combinations
ARIMA	$p = 2$, $d = 2$, $q = 1$
GRNN	$\sigma = 0.04$
SVMS	$\sigma = 0.2707$, $C = 2.8414 \times 10^{11}$, $\varepsilon = 39.127$

Table 5
Forecasting results (unit: 10^6 Wh)

Years	Actual load	ARIMA(2, 2, 1)	GRNN ($\sigma = 0.04$)	SVMSA(σ, C, ε) (0.2707, 2.8414×10^{11}, 39.127)
1995	105,368	96,531	98,561	105,710
1996	111,140	101,562	105,370	112,400
1997	118,299	107,135	111,140	118,490
1998	128,130	112,554	118,300	123,400
1999	131,726	118,143	128,130	131,240
2000	142,413	124,004	131,730	135,570
2001	143,624	129,669	142,410	143,840
2002	151,193	135,824	143,620	145,150
2003	159,380	141,755	151,190	157,460
MAPE		10.31%	5.18%	1.76%
MAD		13,788	6,758	2,448
NRMSE		0.105997	0.054732	0.026357

various forecasting models. The proposed SVMSA model has smaller MAPE, MAD and NRMSE values than the ARIMA(2, 2, 1) and GRNN models based on the same forecasting period. Fig. 7 illustrates the real values and the forecasting values of the different models. The sensitivity analysis of the SVMSA model to the three parameters is provided, as displayed in Figs. 8–10. In the analysis process, two SVMSA parameters were fixed to study the change of MAPE

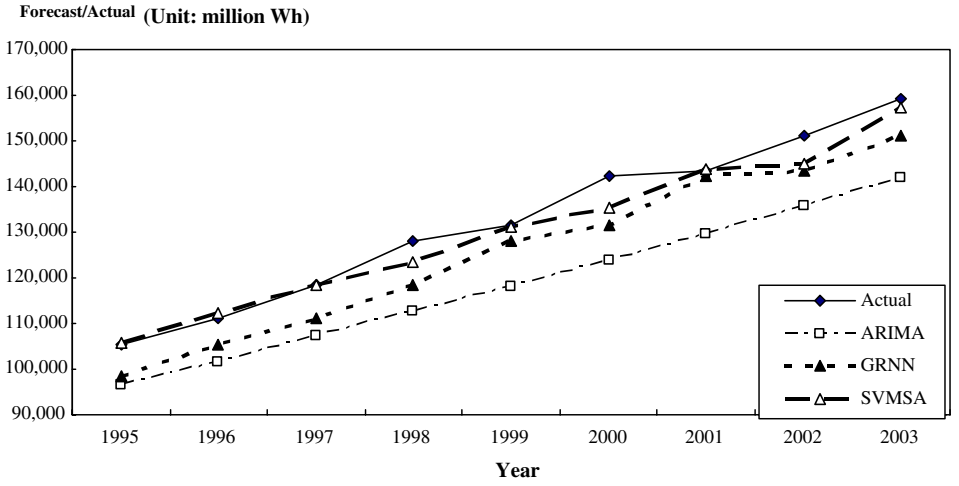

Fig. 7. Forecasting values for different models.

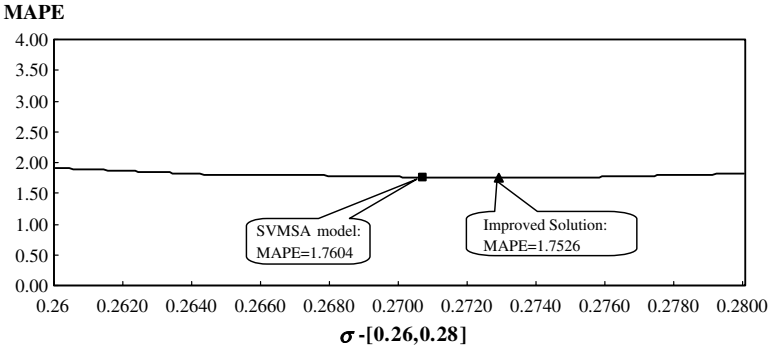

Fig. 8. Results of various σ in which $C = 284140$ and $\varepsilon = 39.127$.

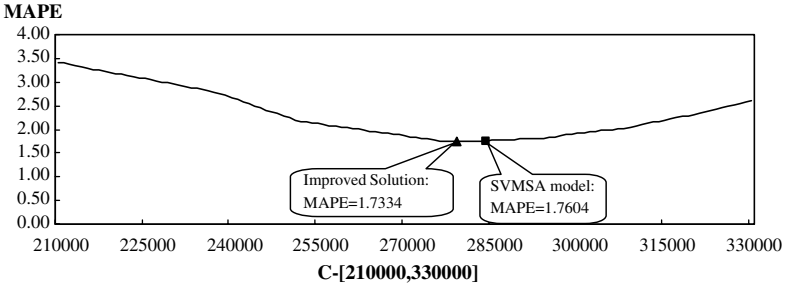

Fig. 9. Results of various C in which $\sigma = 0.2707$ and $\varepsilon = 39.127$.

values when the third parameter varied. Smaller MAPE values were obtained by different combinations of the free parameters. However, the improved solutions are only a little better than the

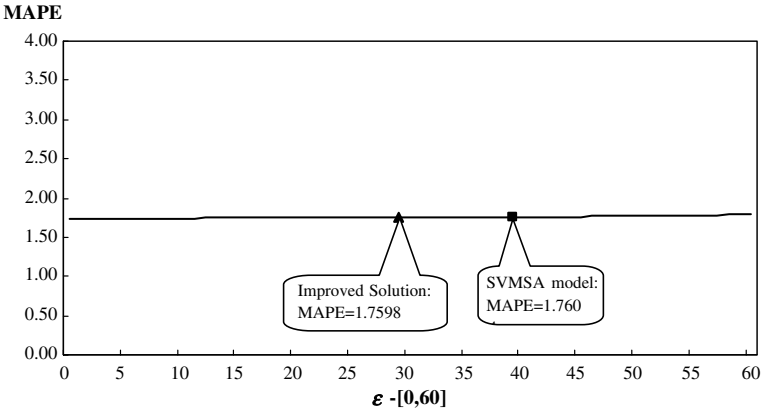

Fig. 10. Results of various ε in which $\sigma = 0.2707$ and $C = 284140$.

Table 6
Differences of MAPE values between the SVMSA solution and improved solutions

The SVMSA solution		Improved solutions		Difference (MAPE1 − MAPE2)
Parameters	MAPE1	Parameters	MAPE2	
$\sigma = 0.2707$	1.76%	$\sigma = 0.2729$; $C = 284140$; $\varepsilon = 39.127$	1.75%	0.01%
$C = 284140$		$\sigma = 0.2707$; C=279000; $\varepsilon = 39.127$	1.73%	0.03%
$\varepsilon = 39.127$		$\sigma = 0.2707$; C=284140; $\varepsilon = 29.000$	1.7598%	0.0002%

original SVMSA solution in terms of MAPE. Table 6 presents the difference between the MAPE values obtained by the proposed SVMSA and the improved solutions. To verify the significance of the accuracy improvement of the SVMSA, two tests, namely the Mann–Whitney U test [48] and the Wilcoxon signed rank test [49], were conducted. The Mann–Whitney U test is an approach for assessing the significance of a difference in the central tendency of two data series. These two data error series are ranked from the smallest value to the largest value. The test statistic U is given by Eq. (30):

$$U = \min\{U_1, U_2\} \tag{30}$$

where

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{31}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \tag{32}$$

The n_1 and n_2 are the sizes of data series one and data series two, respectively. The R_1 and R_2 are the rank sums of data series one and data series two, respectively. Additionally, the Wilcoxon signed rank test is used to detect the significance of a difference in the central tendency of two data series when the size of the two data series is equal [50]. The test statistic W is represented as Eq. (33):

Table 7
Mann–Whitney U test and Wilcoxon signed-rank test

	Mann–Whitney U test		Wilcoxon signed-rank test	
	$\alpha = 0.025$ $U = \mathbf{17}$	$\alpha = 0.05$ $U = \mathbf{21}$	$\alpha = 0.025$ $W = \mathbf{6}$	$\alpha = 0.05$ $W = \mathbf{9}$
SVMSA vs. ARIMA(2, 2, 1)	0	0	0	0
SVMSA vs. GRNN($\sigma = 0.04$)	11	11	0	0

$$W = \min\{S^+, S^-\} \tag{33}$$

where

$$S^+ = \sum_{i=1}^{n} I^+(d_i), \quad S^- = \sum_{i=1}^{n} I^-(d_i) \tag{34}$$

$$I^+(d_i) = \begin{cases} 1 & \text{if } d_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad I^-(d_i) = \begin{cases} 1 & \text{if } d_i < 0 \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

and

$$d_i = (\text{data series one})_i - (\text{data series two})_i \tag{36}$$

Both tests were performed at the 0.025 and 0.05 significance levels in one tail tests. The test results (Table 7) showed that the SVMSA model yields improved forecast results and significantly outperforms the other two forecasting models.

5. Conclusions

For an export oriented economy, like Taiwan, accurate load forecasting can provide advantages through saving and efficiently distributing limited power. From the historical data, the electricity load values for Taiwan exhibit a strong growth trend. This electricity load phenomenon is common in developing countries. Therefore, the forecasting of electricity demand is a rigorous task to avoid over production or under production of electricity load. This work employed a novel forecasting technique, SVMSA, to investigate its feasibility in forecasting annual electricity loads. Two other forecasting approaches, ARIMA and GRNN, were used for comparing the forecasting performance. The experimental results indicate that the SVMSA model outperformed the other approaches in forecasting accuracy. The superior performance of the SVMSA models compared to the ARIMA(2, 2, 1) and GRNN models has several causes. First, the SVMSA model has nonlinear mapping capabilities and, thus, can more easily capture electricity load data patterns than can the ARIMA and GRNN models. Second, the parameter selection in the SVMs significantly influences their forecasting performance. Improper determination of these three parameters will cause either over fitting or under fitting of a SVM model. In this work, the SA algorithm can determine suitable free parameters for forecasting the electricity load. Finally, the SVMSA model performs structural risk minimization rather than minimizing the training errors. Minimizing the

upper bound on the generalization error improves the generalization performance compared to the ARIMA and GRNN models.

This investigation is the first to apply the SVM model with SA to electricity load forecasting. The empirical results obtained in this study demonstrate that the proposed model offers a valid alternative for application in the electricity industry. In the future, climate factors, social activities and seasonal factors can be included in the SVMSA model for forecasting electricity load. Additionally, some other advanced searching techniques for determining suitable parameters can be combined with SVMs to forecast electricity load.

Acknowledgment

This research was conducted with the support of the National Science Council (NSC 93-2213-E-212-001 and NSC 93-2745-H-212-001-URD).

References

[1] Gross G, Galiana FD. Short term load forecasting. Proc IEEE 1987;75:1558–73.
[2] Ranaweera DK, Karady GG, Farmer RG. Economic impact analysis of load forecasting. IEEE Trans Power Syst 1997;12:1388–92.
[3] Douglas AP, Breipohl AM, Lee FN, Adapa R. Risk due to load forecast uncertainty in short term power system planning. IEEE Trans Power Syst 1998;13:1493–9.
[4] Bunn DW, Farmer ED. Comparative models for electrical load forecasting. New York: John Wiley; 1985.
[5] Bunn DW. Forecasting loads and prices in competitive power markets. Proc IEEE 2000;88:163–9.
[6] Box GEP, Jenkins GM. Time series analysis, forecasting and control. San Francisco: Holden-Day; 1970.
[7] Vemuri S, Hill D, Balasubramanian R. Load forecasting using stochastic models. In: Proceeding 8th Power Industrial Computing Application Conference, 1973, pp. 31–37.
[8] Chen JF, Wang WM, Huang CM. Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting. Electric Power Syst Res 1995;34:187–96.
[9] Christianse WR. Short term load forecasting using general exponential smoothing. IEEE Trans Power Apparatus Syst 1971;PAS-90:900–11.
[10] Park JH, Park YM, Lee KY. Composite modeling for adaptive short-term load forecasting. IEEE Trans Power Syst 1991;6:450–7.
[11] Mbamalu GAN, El-Hawary ME. Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation. IEEE Trans Power Syst 1993;8:343–8.
[12] Douglas AP, Breipohl AM, Lee FN, Adapa R. The impact of temperature forecast uncertainty on Bayesian load forecasting. IEEE Trans Power Syst 1998;13:1507–13.
[13] Sadownik R, Barbosa EP. Short-term forecasting of industrial electricity consumption in Brazil. J Forecast 1999;18:215–24.
[14] Gelb A. Applied optimal estimation. Massachusetts: The MIT Press; 1974.
[15] Brown RG. Introduction to random signal analysis and Kalman filtering. New York: John Wiley; 1983.
[16] Moghram I, Rahman S. Analysis and evaluation of five short-term load forecasting techniques. IEEE Trans Power Syst 1989;4:1484–91.
[17] Asbury C. Weather load model for electric demand energy forecasting. IEEE Trans Power Apparatus Syst 1975;PAS-94:1111–6.
[18] Papalexopoulos AD, Hesterberg TC. A regression-based approach to short-term system load forecasting. IEEE Trans Power Syst 1990;5:1535–47.

[19] Soliman SA, Persaud S, El-Nagar K, El-Hawary ME. Application of least absolute value parameter estimation based on linear programming to short-term load forecasting. Int J Electrical Power Energy Syst 1997;19:209–16.
[20] Hyde O, Hodnett PF. An adaptable automated procedure for short-term electricity load forecasting. IEEE Trans Power Syst 1997;12:84–93.
[21] Rahman S, Bhatnagar R. An expert system based algorithm for short-term load forecasting. IEEE Trans Power Syst 1988;3:292–9.
[22] Rahman S, Hazim O. A generalized knowledge-based short-term load-forecasting technique. IEEE Trans Power Syst 1993;8:508–14.
[23] Chiu CC, Kao LJ. Combining a neural network with a rule-based expert system approach for short-term power load forecasting in Taiwan. Expert Syst Appl 1997;13:299–305.
[24] Park DC, El-Sharkawi MA, Marks II RJ, Atlas LE, Damborg MJ. Electric load forecasting using an artificial neural network. IEEE Trans Power Syst 1991;6:442–9.
[25] Ho KL, Hsu YY, Yang CC. Short-term load forecasting using a multilayer neural network with an adaptive learning algorithm. IEEE Trans Power Syst 1992;7:141–9.
[26] Novak B. Superfast auto-configuring artificial neural networks and their application to power systems. Electric Power Syst Res 1995;35:11–6.
[27] Darbellay GA, Slama M. Forecasting the short-term demand for electricity—do neural networks stand a better chance. Int J Forecast 2000;16:71–83.
[28] Abdel-Aal RE. Short-term hourly load forecasting using abductive networks. IEEE Trans Power Syst 2004;19:164–73.
[29] Hsu CC, Chen CY. Regional load forecasting in Taiwan: Applications of artificial neural networks. Energy Convers Manage 2003;44:1941–9.
[30] Cao L. Support vector machines experts for time series forecasting. Neurocomputing 2003;51:321–39.
[31] Cao L, Gu Q. Dynamic support vector machines for non-stationary time series forecasting. Intell Data Anal 2002;6:67–83.
[32] Tay FEH, Cao L. Modified support vector machines in financial time series forecasting. Neurocomputing 2002;48:847–61.
[33] Tay FEH, Cao L. Application of support vector machines in financial time series forecasting. Omega 2001;29:309–17.
[34] Lu W, Wang W, Leung AYT, Lo SM, Yuen RKK, Xu Z, Fan H. Air pollutant parameter forecasting using support vector machines. In: IJCNN '02 Proceedings of the 2002 International Joint Conference, vol. 1, 2002. p. 12–7.
[35] Specht DA. A general regression neural network. IEEE Trans Neural Networks 1991;2:568–76.
[36] Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik VN. Support vector regression machines. Adv Neural Informat Process Syst 1997;9:155–61.
[37] Vapnik V, Golowich S, Smola A. Support vector machine for function approximation, regression estimation, and signal processing. Adv Neural Informat Process Syst 1996;9:281–7.
[38] Vojislav K. Learning and soft computing—Support vector machines, Neural Networks and Fuzzy Logic Models. Massachusetts: The MIT Press; 2001.
[39] Peressini AL, Sullivan FE, Uhl JJ. The mathematics of nonlinear programming. New York: Springer; 1988.
[40] Christiani N, Shawe-Taylor J. An introduction to support vector machines. Cambridge: Cambridge University Press; 2000.
[41] Riesz F, Nagy BSz. Functional analysis. New York: Frederick Ungar; 1955.
[42] Cercignani C. The Boltzmann equation and its applications. Berlin: Springer-Verlag; 1988.
[43] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH. Equations of state calculations by fast computing machines. J Chem Phys 1953;21:1087–92.
[44] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science 1983;220:671–80.
[45] Van Laarhoven PJM, Aarts EHL. Simulated annealing: Theory and applications. Dordrecht: Kluwer Academic Publishers; 1987.
[46] Dekkers A, Aarts EHL. Global optimization and simulated annealing. Math Programm 1991;50:367–93.
[47] Schalkoff RJ. Artificial neural networks. New York: McGraw-Hill; 1997.

[48] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947;18:50–60.
[49] Wilcoxon F. Individual comparisons by ranking methods. Biometrics 1945;1:80–3.
[50] Daniel WW. Applied nonparametric statistics. Boston: Houghton Mifflin Co; 1978.