Korean J. Chem. Eng., 27(4), 1063-1071 (2010)
DOI: 10.1007/s11814-010-0220-9

Heat consumption forecasting using partial least squares, artificial neural network and support vector regression techniques in district heating systems

Tae Chang Park*, Ui Sik Kim*, Lae-Hyun Kim, Byung Wan Jo***, and Yeong Koo Yeo*,†**

*Department of Chemical Engineering, Hanyang University, Seoul 133-791, Korea
**Department of Chemical Engineering, Seoul National University of Technology, Seoul 135-743, Korea
***Department of Civil Engineering, Hanyang University, Seoul 133-791, Korea
(*Received 13 November 2009 • accepted 10 January 2010*)

Abstract−Effective management of district heating networks depends upon the correct forecasting of heat consumption during a certain period. In this work short-term forecasting for the amount of heat consumption is performed first to validate the three forecasting methods: partial least squares (PLS) method, artificial neural network (ANN), and support vector regression (SVR) method. Based on the results of short-term forecasting, one-week ahead forecasting was performed for the Suseo district heating network. Data of heat consumption and ambient temperature during January and February in 2007 and 2008 were employed as training elements. The heat consumption estimated was compared with actual one in the Suseo area to validate the forecasting models.

Key words: Partial Least Squares, Artificial Neural Network, Supporting Vector Regression, Heat Forecasting

INTRODUCTION

A district heating system (DHS) usually consists of energy suppliers and a large number of consumers, district heating pipelines and heat storage facilities in a region. DHS plays an important role in covering the heating demands in downtown and suburban area. DHS's can be characterized by reduction of energy consumption, increase of energy efficiency and decrease of generation of pollutants. Hence the subject of optimal operation of DHS's has significant economical potential. In Korea, DHS's take charge of about 10% of energy consumption based on the total households in 2008. In contrast to other countries, the heat source used in DHS's mainly consists of fossil fuels in Korea resulting in relatively high operating cost. To improve economical efficiency in operations it is recommended to use waste materials as a heat source and to increase energy efficiency by the optimal operation of heat generation systems and heat distribution networks. In the optimal operation the first step is the forecasting of heat consumption in target areas. Economical management of the energy distribution system and planning are highly dependent upon the accurate forecasting. Basic operating functions such as unit commitment, generation of electricity, fuel scheduling and heat distribution can be performed efficiently with an accurate forecast.

Forecasting heat consumption can be classified as spatial estimation and time-dependent prediction. In time-dependent prediction there are long-term, mid-term and short-term predictions. The long-term prediction is generally performed for the period of 10-20 years in the future to be used in the management planning of the whole energy distribution network. The supply price, substitute energy sources and energy market as well as marketing strategies are considered in the long-term prediction as the main elements. The mid-term prediction is generally performed for the period of a few weeks in the future to be used in the planning of fuel supply and maintenance program. The short-term prediction is performed for a few days or several hours in the future to be used in the planning of daily operation and management of the network [1]. In this work the short-term prediction is performed first to forecast heat consumption in a short-term period. The short-term demand forecasting is very important because the heat supplier adjusts electricity generation equipment based on the results of short-term forecasting. The operational status of that equipment is mainly dependent upon the amount of heat demand.

So far various statistical forecasting techniques have been applied to short-term demand forecasting, including time series method [2,3] and regression method [4]. In general, these methods are basically linear models while the heat demand pattern is usually a nonlinear function of exogenous variables. For this reason, supervisory learning methods such as support vector machine (SVM), support vector regression (SVR), artificial neural network (ANN) or partial least squares (PLS) have been widely used recently. In these methods input/output functionality can be found based on actual operational data. A multilayer ANN was used to predict consumption of natural gas based on the weather conditions and past consumption [5]. The ANN was also used effectively to estimate short-term maximum/minimum demand to improve performance of electricity distribution network [6]. One-hour ahead load forecasting using ANN was performed incorporating various day types such as holidays, Saturday and Sunday as input variables [7]. PLS method was successfully used to forecast electricity demand [8].

In this work we employed three black-box modeling techniques (SVR, PLS and ANN) to forecast heat consumption in Suseo district heating network which is interconnected with other capital regional district heating networks. Time, past consumption and ambient temperatures were used as input variables to forecast heat consumption for one week in the future. We compared the performances

†To whom correspondence should be addressed.
E-mail: ykyeo@hanyang.ac.kr

of these three black-box modeling techniques for the prediction of heat consumption in the Suseo network and analyzed the accuracy of each method by comparing forecasting errors.

1. Formulation of Forecasting Methods

1-1. Partial Least Squares (PLS)

The partial least squares (PLS) method is a multiple linear regression model which can be used to model relations between inputs and outputs with correlations and/or restricted number of data points [9]. For this capability PLS has been widely used as a powerful modeling technique for constructing so-called black-box models from experimental or operational data. In the PLS method the high dimensional spaces of the input and output data obtained from a plant are projected onto the low dimensional feature spaces followed by identification of the best relationship between the feature vectors. More specifically, the high dimensional data matrices X and Y are projected onto several key factors (such as T, U, etc.) and linear regression is performed for the relation between these factors. The outer relations for the output block Y and input block X are given by:

$$X = TP^T + E = \sum t_h p_h^T + E \tag{1}$$

$$Y = UQ^T + F = \sum u_h q_h^T + F \tag{2}$$

The coefficient b_h combining each block can be found from the following relation:

$$\hat{u}_h = b_h t_h \tag{3}$$

Usually, the least squares method is used to compute the regression coefficient $b_h(=u_h^T t_h / t_h^T t_h)$. In computation of key factors the nonlinear iterative partial least squares scheme is widely used as a computing algorithm. The nonlinear partial least squares method can be classified into two types: in type I, the inner relations among blocks are nonlinear, and in type II, both the inner and outer relations are nonlinear. The general form of the type I is given by:

$$\hat{u}_h = f(t_h) + R \tag{4}$$

Various models according to the function f were proposed. For example, f can take a quadratic form or artificial neural network can be used.

1-2. Artificial Neural Network (ANN)

The artificial neural network (ANN) is a powerful mathematical tool originally inspired by the way the human brain processes information. ANN's are composed of simple elements operating in parallel. These elements are stimulated by biological uneasy systems. As in nature, the network function is determined largely by the connections between elements. We can train an ANN to perform a particular function by adjusting the values of the connections (weights) between elements. The basic unit of an ANN is the artificial neuron. The neuron receives information through a number of input nodes, processes it internally, and puts out a response. The neurons are organized in a way that defines the network structure. The most concerned structure is the multilayer perceptron type in which the neurons are organized in layers. The neurons in each layer may share the same inputs, but not connected to each other. Fig. 1 shows an example of a three-layer network. Each layer has a specified number of nodes and the interconnections are only between neurons of adjacent layers. Each neuron belonging to a layer is connected to all the neurons of adjacent layers. The layer between the input neurons and the output layer is called the hidden layer (see Fig. 1). First, the input values are linearly combined, and then the result is used as the argument of a nonlinear activation function. The combination uses the weights (parameters) attributed to each connection, and a constant bias term. The estimation of the weights is called the training of the network. The neuron output is given by:

$$y = f[(\sum w \cdot x) - \theta] \tag{5}$$

The activation function f represents an activated state when the result of the summation of the products between weights and input values has a positive value. If the result of the summation is negative, f represents an inactivated state. θ represents the value of the characteristic neuron offset (bias) which should be overcome by the result of the summation in order for the state to be in activated state. The most common choice for the activation function in multilayer networks is unipolar sigmoid and bipolar sigmoid functions. The output of the unipolar sigmoid function is limited between 0 and +1 as can be seen in Eq. (6). The output of the bipolar sigmoid function is limited between −1 and +1 as can be seen in Eq. (7).

$$f(x) = \frac{1}{1+e^{-\lambda x}} \tag{6}$$

$$f(x) = \frac{2}{1+e^{-\lambda x}} - 1 \tag{7}$$

The most used training algorithm in load forecasting is back-propagation algorithm. The simplest implementation of back-propagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly, the negative of the gradient, so that the output approaches the target value. The value function J is specified as the summation of the squares of the difference between the desired output and the output from the network as given by Eq. (8).

$$J = \frac{1}{2}\sum_k (\text{desired}_k - \text{out}_k)^2 \tag{8}$$

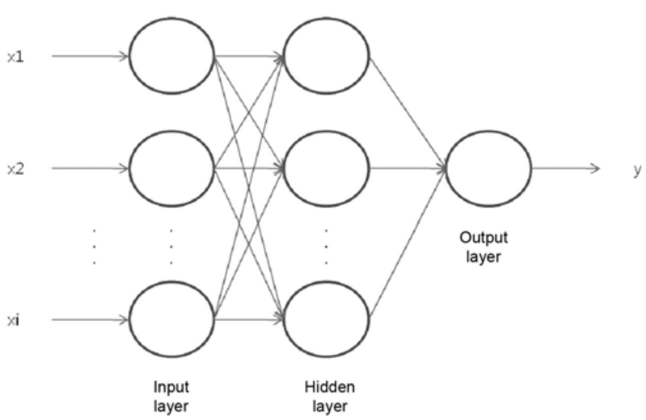

Fig. 1. A three-layer artificial neural network.

1-3. Support Vector Regression (SVR)

The support vector machine (SVM) has been known as a very powerful modeling algorithm used to solve classification problems. Because the formulation of SVMs is based on structural risk minimization rather than empirical risk minimization, which is employed

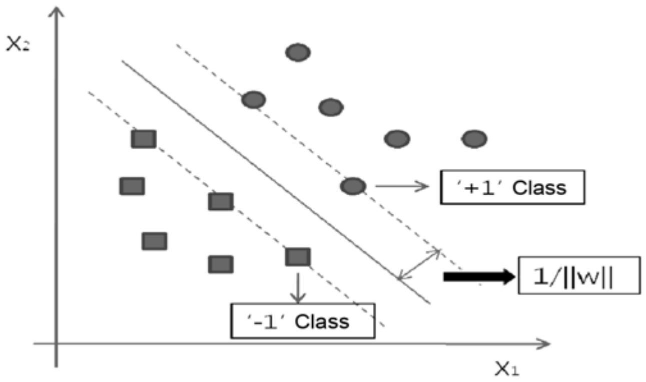

Fig. 2. A decision function of SVM.

by other conventional black-box modeling algorithms such as ANN and PLS, the SVMs typically show better performance than the conventional algorithms. In the SVM, a hyperplane dividing the multi-dimensional learning data into two groups is identified first. The hyperplane is used as a determining function to predict the group to which the unknown data should belong [10]. For given SVM learning data, the hyperplane is found so that two classes are away from each other as long as possible according to the label y_i (see Fig. 2). In this case the hyperplane $f(x)=w^Tx+b$ becomes the determining function. If the value of f(x) is greater than 0 for x with unknown label, it is classified as +1, while it is classified as −1 class when f(x) is less than 0. In this case, the value of $2/\|w\|$ is called a margin which represents the minimum distance between two classes separated by the hyperplane. Basically, an SVM uses the hyperplane that maximizes the margin as the determining function.

The problem of estimation of w and b in the determining function $f(x)=w^Tx+b$ of a SVM can be formulated as a 2^{nd}-order programming problem (SP_0) given by:

$$\begin{aligned} &\min w^T w \\ (SP_0)&: \text{s, t, } y(w^T x^i + b) \geq 1, \text{ for } i=1, \ldots, l \end{aligned} \tag{9}$$

The learning data with different labels may not be separated clearly by the hyperplane. To take into account this overlapping case, the relaxed model (SP) may be used instead of the original model (SP_0).

$$\begin{aligned} &\min w^T w + c\sum_i^l \varepsilon_i \\ (SP)&: \text{s, t, } y_i\,(w^T x^i + b) \geq 1-\varepsilon_i, \ \forall\, i \\ &\quad \varepsilon_i \geq 0, \ \forall\, i \end{aligned} \tag{10}$$

$$\begin{aligned} &\min e^T\alpha + \frac{1}{2}\alpha^T Q\alpha \\ (SD)&: \text{s, t, } y^T\alpha = 0 \\ &\quad 0 \leq \alpha_i \leq C, \ \forall\, i \end{aligned} \tag{11}$$

The second term $\sum\varepsilon_i$ represents the amount of the error which represents the degree of mismatch between the learning data and the determining function obtained from experienced errors. C is the relative weight between the margin and the experienced error. In short, the objective in (SP) identifies the determining function f(x) which maximizes the margin and minimizes the experienced error.

$$u = \sum_i c_i^* y_i x^1$$

$$f(x) = w^T x + b = \sum_i \alpha_i^* y_i (x^i)^T x + b \tag{12}$$

Transformation to a specific pace is performed by a kernel function. The most widely used kernel functions include RBF (radial basis function) and polynomial kernel. The RBF kernel is given by:

$$k(x^i, x^j) = \exp\left(-\frac{\|x^i - x^j\|^2}{\sigma^2}\right) \tag{13}$$

And the polynomial kernel is given by:

$$k(x^i, x^j) = ((x^i)^T x^j)^P \tag{14}$$

SVM is used to classify the learning data into '+1' class and '−1' class (see Fig. 2). The support vector regression (SVR) can be considered as generalization of SVM so that arbitrary real values might be estimated. In the SVR ε-insensitive loss function (Eq. (15)) is used instead of $\sum\varepsilon_i$.

$$L_\varepsilon(x, y, f) = \max(0, |y - f(x)| - \varepsilon) = \max(0, |y - (w^T x + b)| - \varepsilon) \tag{15}$$

The value of the ε-insensitive loss function depends on the size of the estimation error defined by the difference between the actual value y and the estimated value $f(x)=w^Tx+b$. If the estimation error is less than ε, we have $L_\varepsilon(x, y, f)=0$. Otherwise the loss function takes the value of the difference between the absolute error and ε as given by $L_\varepsilon(x, y, f)=|y-f(x)|-\varepsilon$ (see Fig. 3). In SVR, the margin is maximized while maintaining y and $f(x)=w^Tx+b$ values within ε, which can be described as an optimization model given by:

$$\begin{aligned} &\min w^T w + C\sum_i^l(\varepsilon_i^+ \varepsilon_i^-) \\ (RP)&: \text{s, t, } y_i(w^T x^i + b) - y_i \geq \varepsilon + \varepsilon_i^+, \ \forall\, i \\ &\quad y_i - (w^T x^i + b) \geq \varepsilon + \varepsilon_i^-, \ \forall\, i \\ &\quad \varepsilon_i^+, \varepsilon_i^- \geq 0, \ \forall\, i \end{aligned} \tag{16}$$

$$\begin{aligned} &\max \sum_i y_i(\alpha_i^- - \alpha_i^+) - \varepsilon\sum_i(\alpha_i^- + \alpha_i^+) - \frac{1}{2}\sum_{i,j}(\alpha_i^- - \alpha_i^+)(\alpha_i^- - \alpha_i^+)Q_{i,j} \\ (RD)&: \text{s, t, } \sum_i(\alpha_i^- - \alpha_i^+) = 0 \\ &\quad 0 \leq \alpha_i^+, \alpha_i^- \leq C, \ \forall\, i \end{aligned} \tag{17}$$

2. Application

Data measured during January in 2007 and 2008, during February in 2007 and those measured from February 1 to February 17 were adopted as training data set. As for forecasting period, we chose one

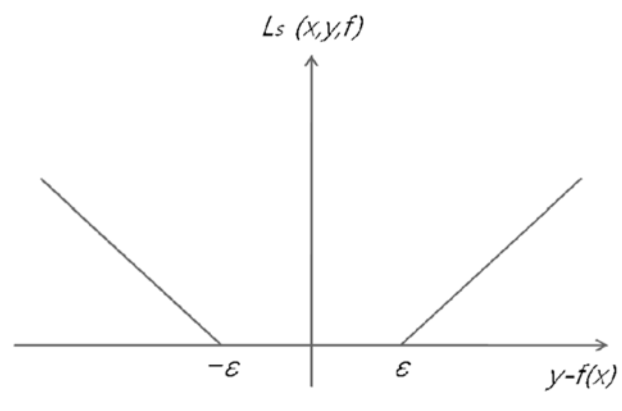

Fig. 3. ε-Insensitive loss function.

Table 1. Typical input data for training

Time (h)	T_i (ºC)	T_{i-1} (ºC)	T_{i-2} (ºC)	T_{i-3} (ºC)	D_{i-1} (Gcal/h)	D_{i-2} (Gcal/h)	D_{i-3} (Gcal/h)	D_{i-4} (Gcal/h)	D_{i-5} (Gcal/h)
1	4.5	0.3	4.5	0.6	399	421	475	457	471
2	4.1	0	4.5	0.3	369	399	409	411	474
3	2	0	2.4	0	356	372	390	380	433
4	2.4	−0.8	1.3	0.3	368	403	425	411	454
5	2.8	−1.1	0.6	0.6	400	433	438	434	491
6	2	−1.1	0	1	415	452	427	424	446
7	2	−0.8	0	1	458	493	447	490	453
8	1.3	−0.8	−0.4	2	479	476	487	514	464
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
21	−3.5	−0.8	−3.2	−0.8	598	580	554	572	654
22	−4.6	−1.1	−3.9	−0.1	605	579	566	581	674
23	−5	−1.5	−4.6	0.3	598	674	557	577	666
24	−5.3	−2.2	−4.6	0.1	576	672	555	578	647

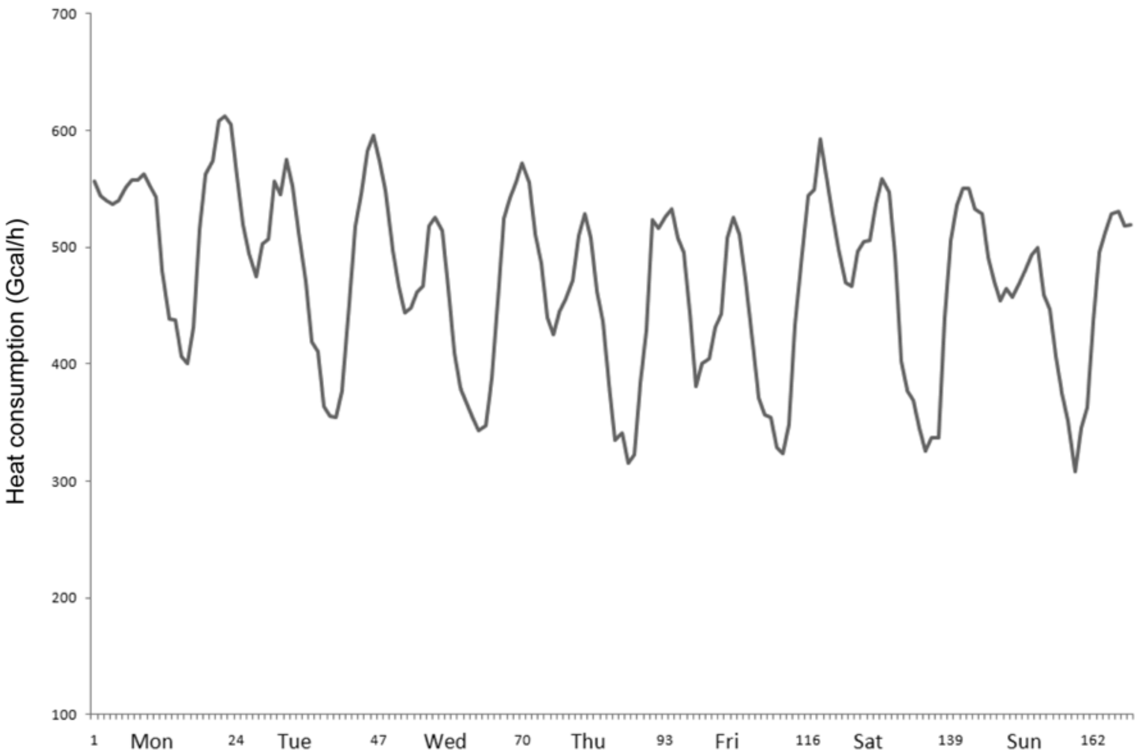

Fig. 4. Hourly load curve for one week.

week from February 18 to February 24, 2008. The input vector for prediction of heat consumption consists of the amount of heat consumption up to 5 days prior to forecasting (D_{i-1}, D_{i-2}, D_{i-3}, D_{i-4}, and D_{i-5}), ambient temperatures up to 3 days prior to forecasting (T_{i-1}, T_{i-2}, and T_{i-3}) and forecasting day (T_i), and time (Time). Table 1 shows typical input data for 97 days, which means the number of data is 2328. Fig. 4 shows heat consumption trend for one week (from January 14 to 21, 2007). As can be seen, heat consumption during afternoon is lower than that during morning, evening and night hours. The overall heat consumption trend during the training period is shown in Fig. 5.

RESULTS AND DISCUSSION

1. One-day-ahead Hourly Forecasting

One-day-ahead hourly forecasting for February 18, 2008 was performed by using PLS, ANN and SVR. The mean absolute percentage error (MAPE) method was used to evaluate accuracy of the results.

$$\mathrm{MAPE} = \frac{|c_{actual} - c_{forecasting}|}{c_{actual}} \times 100 \tag{18}$$

In Eq. (18) *Cactual* and *Cforecasting* represent actual and fore-

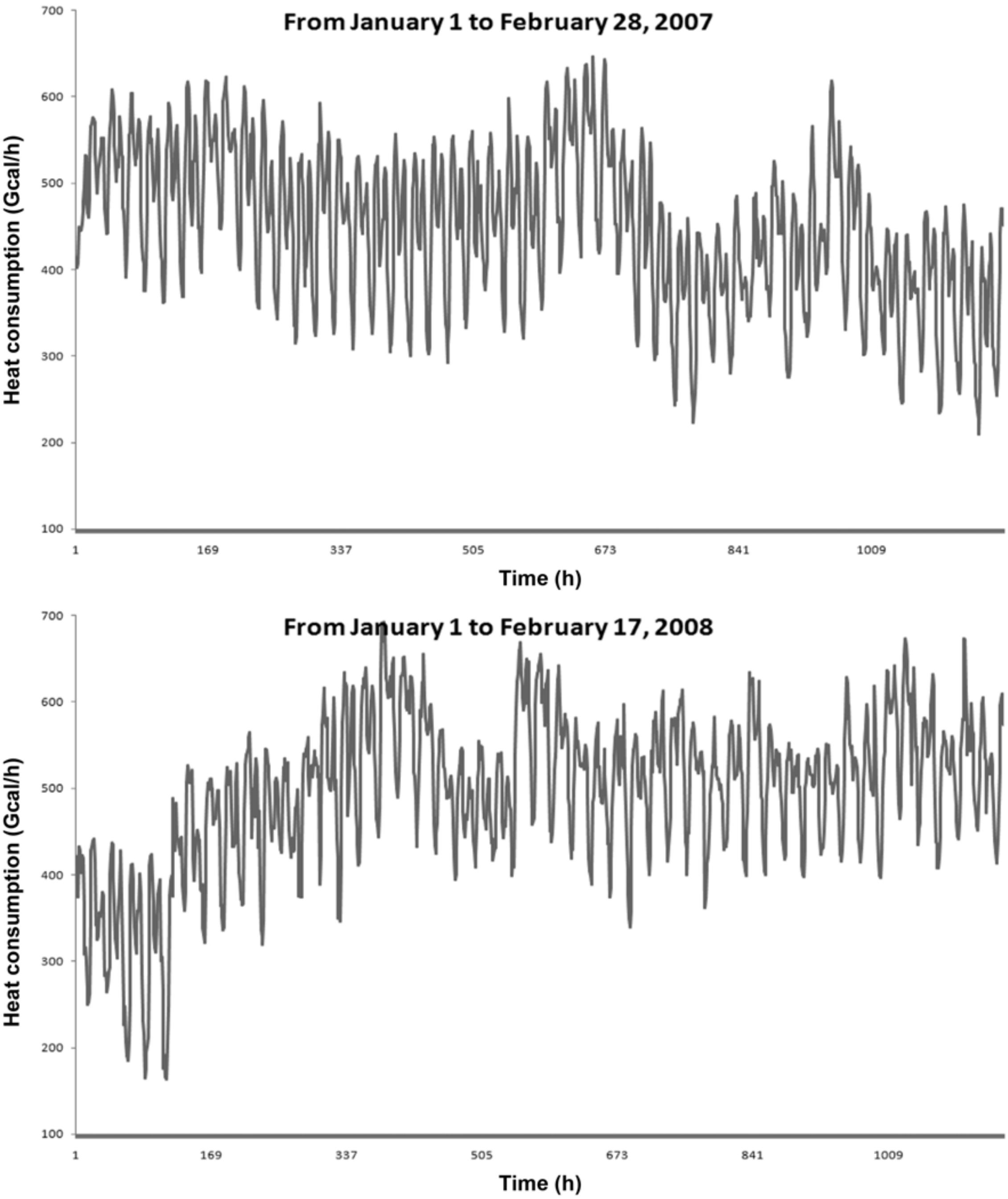

Fig. 5. Historical data curve for training.

casted heat consumption, respectively. Hourly forecasting errors are shown in Table 2. The overall average error of PLS is 3.87% while that of ANN and SVR is 6.54% and 4.95%, respectively. The maximum error of SVR is 9.82%, which is lower than that of PLS (16.47%) and ANN (13.20%). In terms of the overall error, we can say that PLS exhibits better forecasting performance than ANN or SVR.

2. Two-days-ahead Hourly Forecasting

The three trained models (PLS, ANN and SVR) were used to perform two-days-ahead hourly forecasting for February 19, 2008. Table 3 shows errors in two-days-ahead forecasting. The overall average error of SVR is 5.55%, while that of PLS and ANN is 6.57% and 10.09%, respectively. The maximum error of SVR is 14.60%, which is lower than that of PLS (21.07%) and ANN (20.18%). We can see that SVR exhibits better overall forecasting performance than ANN or SVR in two-days-ahead forecasting.

3. Three-days-ahead Hourly Forecasting

Again, the trained three models (PLS, ANN and SVR) were used to perform three-days-ahead hourly forecasting for February 20, 2008. Table 4 shows errors in three-days-ahead forecasting. The overall average error of SVR is 6.32%, while that of PLS and ANN is 11.41% and 10.29%, respectively. The maximum error of SVR is 19.64%, which is lower than that of PLS (34.41%) and ANN (23.54%). We can see that SVR again exhibits better overall forecasting performance than ANN or SVR in three-days-ahead forecasting.

Table 2. One-day-ahead hourly forecasting using three forecasting methods (2008.2.18)

Time (h)	Actual (Gcal/h)	PLS (Gcal/h)	Error (%)	ANN (Gcal/h)	Error (%)	SVR (Gcal/h)	Error (%)
1	551	536.95	2.55	521.66	5.32	576.86	4.69
2	517	531.40	2.79	505.71	2.18	567.78	9.82
3	513	513.76	0.15	475.80	7.25	539.45	5.16
4	521	540.62	3.77	490.09	5.93	571.26	9.65
5	521	540.78	3.80	485.14	6.88	568.65	9.15
6	535	530.88	0.77	491.53	8.12	556.12	3.95
7	549	548.14	0.16	495.84	9.68	567.80	3.42
8	544	563.79	3.64	505.67	7.05	581.35	6.87
9	511	546.06	6.86	499.42	2.27	559.88	9.57
10	502	502.56	3.70	480.01	4.38	529.41	5.46
11	504	485.61	3.65	443.56	11.99	483.76	4.02
12	463	464.79	0.39	427.84	7.59	449.69	2.87
13	447	449.90	0.65	422.59	5.46	425.20	4.88
14	435	429.37	1.29	412.68	5.13	404.01	7.12
15	402	429.63	6.87	415.40	3.33	399.81	0.54
16	399	430.09	7.79	410.61	2.91	398.03	0.24
17	390	454.24	16.47	414.14	6.19	425.37	9.07
18	489	487.22	0.36	424.47	13.20	462.21	5.48
19	511	521.13	1.98	448.12	12.31	502.26	1.71
20	536	551.57	2.90	521.28	2.75	542.26	1.17
21	552	570.68	3.38	520.92	5.63	556.15	0.75
22	551	584.92	6.16	525.42	4.64	573.55	4.09
23	566	595.86	5.28	521.23	7.91	587.85	3.86
24	549	590.16	7.50	500.06	8.91	578.31	5.34

Averaged forecasting error (%):
PLS: 3.87, ANN: 6.54, SVR: 4.95
Maximum forecasting error (%):
PLS: 16.47, ANN: 13.20, SVR: 9.82

*PLS: Partial Least Squares, ANN: Artificial Neural Network, SVR: Support Vector Regression

Table 3. Two-days-ahead hourly forecasting using three forecasting methods (2008.2.19)

Time (h)	Actual (Gcal/h)	PLS (Gcal/h)	Error (%)	ANN (Gcal/h)	Error (%)	SVR (Gcal/h)	Error (%)
1	518	509.58	1.62	491.81	5.06	544.37	5.09
2	518	506.35	2.25	463.25	10.57	539.81	4.21
3	507	492.10	2.94	426.25	15.93	516.14	1.80
4	514	517.19	0.62	434.89	15.39	546.38	6.30
5	511	521.18	1.99	434.77	14.92	549.07	7.45
6	506	528.16	4.38	445.85	11.89	557.65	10.21
7	527	542.53	2.95	466.33	11.51	573.35	8.80
8	531	555.70	4.65	477.70	10.04	583.19	9.83
9	474	530.77	11.98	475.08	0.23	543.22	14.60
10	517	502.83	2.74	438.97	15.09	506.42	2.05
11	497	468.91	5.65	397.26	20.07	458.29	7.79
12	464	449.10	3.21	386.75	16.65	429.31	7.48
13	403	432.58	7.34	366.26	9.12	403.23	0.06
14	395	422.99	7.09	363.55	7.96	393.50	0.38
15	362	421.11	16.33	367.43	1.50	386.51	6.77
16	351	424.95	21.07	367.13	4.60	391.54	11.55
17	375	443.25	18.20	373.41	0.42	406.92	8.51
18	445	473.87	6.49	388.53	12.69	439.47	1.24
19	501	505.70	0.94	399.92	20.18	473.64	5.46
20	511	534.48	4.60	477.53	6.55	510.60	0.08

Table 3. Continued

Time (h)	Actual (Gcal/h)	PLS (Gcal/h)	Error (%)	ANN (Gcal/h)	Error (%)	SVR (Gcal/h)	Error (%)
21	536	555.19	3.58	491.02	8.39	531.21	0.89
22	533	566.61	6.31	490.70	7.94	540.08	1.33
23	529	584.07	10.41	498.44	5.78	562.23	6.28
24	524	578.24	10.35	473.32	9.67	551.04	5.16

Averaged forecasting error (%):
PLS: 6.57, ANN: 10.09, SVR: 5.55
Maximum forecasting error (%):
PLS: 21.07, ANN: 20.18, SVR: 14.60

*PLS: Partial Least Squares, ANN: Artificial Neural Network, SVR: Support Vector Regression

Table 4. Three-days-ahead hourly forecasting using three forecasting methods (2008.2.20)

Time (h)	Actual (Gcal/h)	PLS (Gcal/h)	Error (%)	ANN (Gcal/h)	Error (%)	SVR (Gcal/h)	Error (%)
1	478	485.14	1.49	466.86	2.33	505.78	5.81
2	477	482.12	1.07	446.76	6.34	499.55	4.73
3	448	461.72	3.06	377.86	15.66	460.47	2.78
4	458	476.16	3.96	389.02	15.06	469.22	2.45
5	457	472.88	3.47	380.32	16.78	457.87	0.19
6	461	483.19	4.81	389.47	15.52	477.11	3.50
7	472	499.37	5.80	420.79	10.85	496.10	5.10
8	486	513.33	5.62	436.69	10.15	509.05	4.74
9	489	489.70	0.14	425.09	13.07	471.24	3.63
10	451	478.15	6.02	381.48	15.41	461.35	2.30
11	428	451.70	5.54	364.18	14.91	425.75	0.53
12	383	437.76	14.30	345.70	9.74	406.33	6.09
13	379	418.81	10.50	323.15	14.74	373.51	1.45
14	359	415.61	15.77	328.52	8.49	372.02	3.63
15	304	408.59	34.41	317.42	4.41	355.73	17.02
16	312	410.90	31.70	319.02	2.25	355.79	14.03
17	361	430.28	19.19	293.08	18.81	378.21	4.77
18	409	459.27	12.29	312.70	23.54	412.51	0.86
19	449	490.89	9.33	349.77	22.10	451.80	0.62
20	466	524.80	12.62	471.15	1.11	503.16	7.97
21	477	542.63	13.76	482.60	1.17	519.46	8.90
22	487	558.96	14.78	485.22	0.37	543.84	11.67
23	475	581.10	22.34	486.01	2.32	568.28	19.64
24	476	580.49	21.95	484.41	1.77	567.93	19.31

Averaged forecasting error (%):
PLS: 11.41, ANN: 10.29, SVR: 6.32
Maximum forecasting error (%):
PLS: 34.41, ANN: 23.54, SVR: 19.64

*PLS: Partial Least Squares, ANN: Artificial Neural Network, SVR: Support Vector Regression

4. Forecasting Heat Consumption for One Week

From one- to three-days-ahead forecasting, we can anticipate effective long-term forecasting by those three methods (PLS, SVR and ANN). We extended the forecasting period to one week and performed forecasting by using three methods. Results of heat consumption forecasting for one week based on PLS are shown in Fig. 6 with data on actual heat consumption. Fig. 7 shows results of heat consumption forecasting for one week based on ANN with data on actual heat consumption. In Fig. 8, results of heat consumption forecasting for one week based on SVR are shown with data on actual heat consumption. Table 5 shows daily forecasting errors for each forecasting method. In PLS forecasting, we can see that the forecasting error grows with time, which means that PLS is not suitable for mid-time or long-time forecasting. ANN shows relatively consistent forecasting errors values, which are large compared to other methods. The SVR method shows the best forecasting results in terms of consistency as well as the magnitude of errors. We can see that the largest average error in the SVR forecasting is 8.81%,

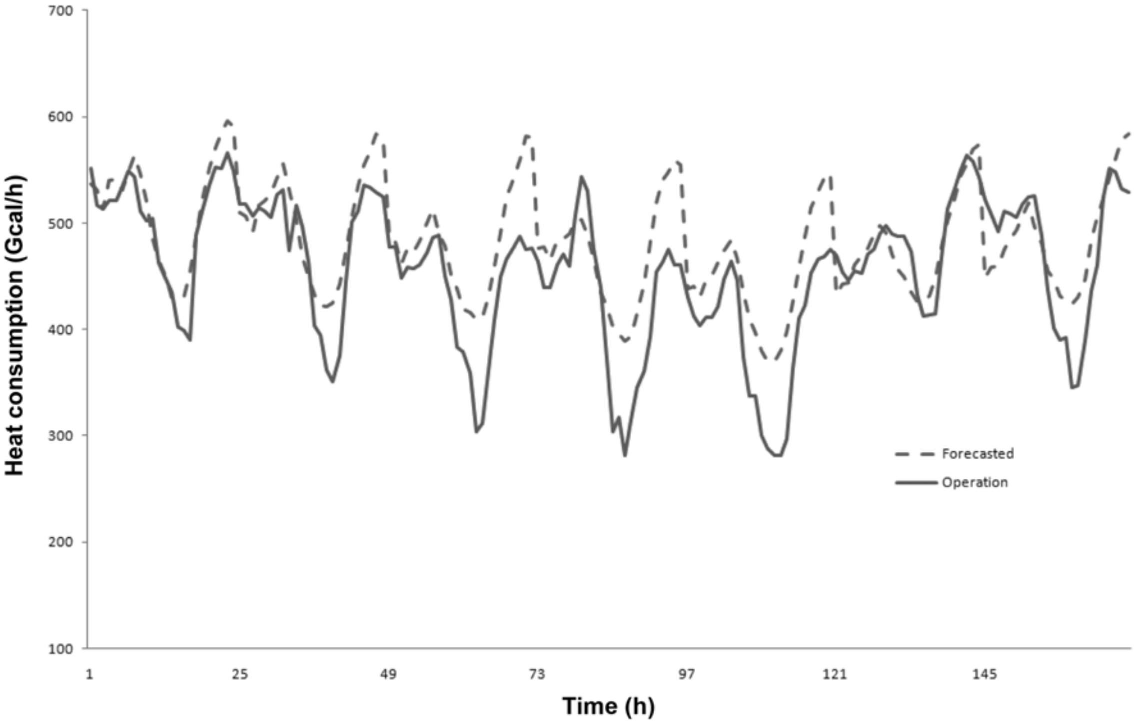

Fig. 6. Forecasting heat consumption hourly using PLS for one week.

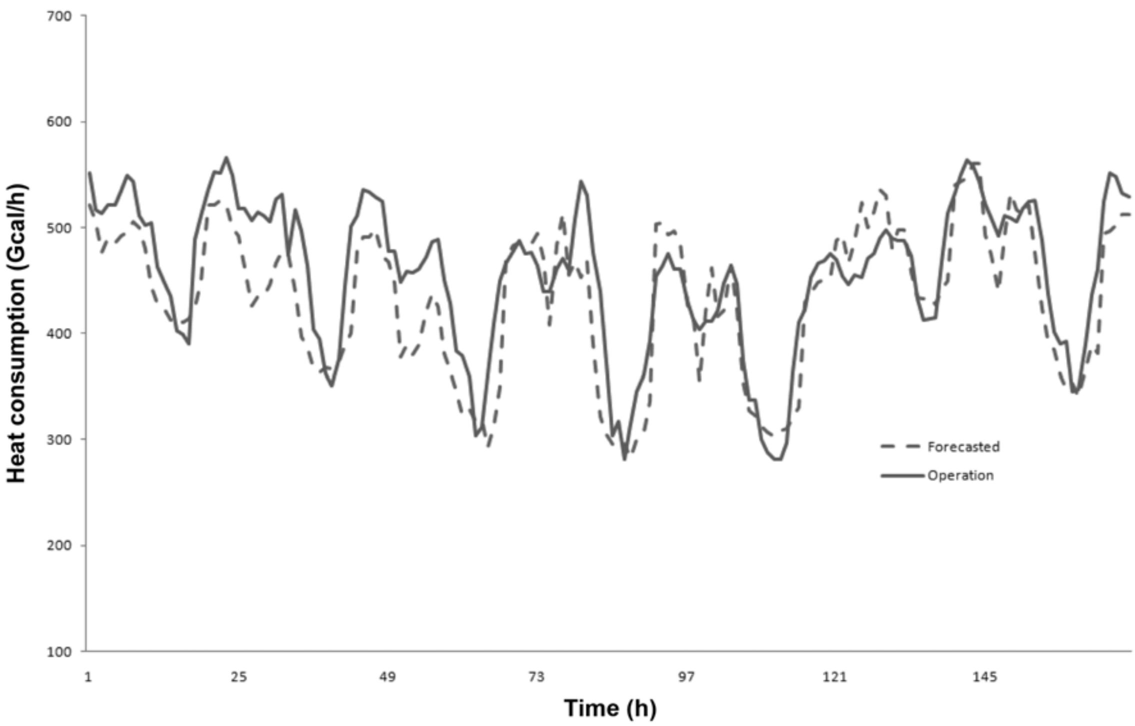

Fig. 7. Forecasting heat consumption hourly using ANN for one week.

which is much lower than that obtained from PLS or ANN.

CONCLUSION

Forecasting heat consumption was performed for the Suseo district heating system by using three forecasting methods: PLS, ANN and SVR. Data on past heat consumption and ambient temperature during January and February in 2007 and 2008 were used as training elements in the three forecasting methods. One week (from February 18 to 24) was chosen as the forecasting period. The results forecasted hourly were compared with actual heat consumption data. The SVR showed the best forecasting performance in terms of over-

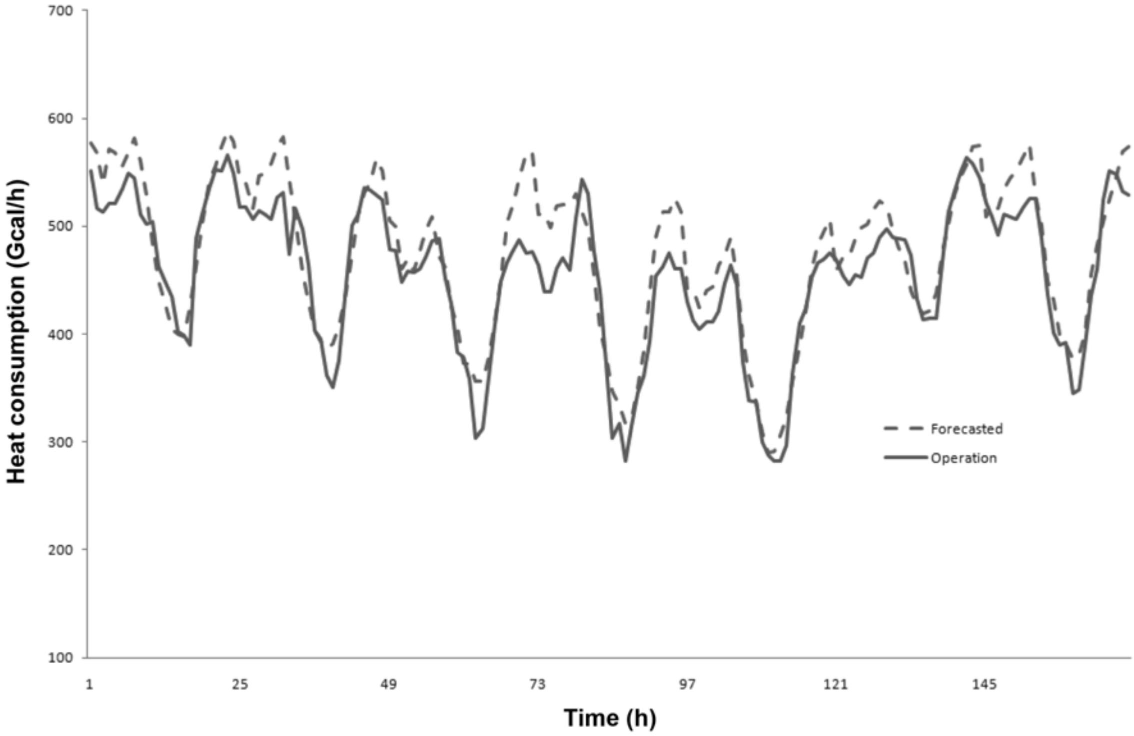

Fig. 8. Forecasting heat consumption hourly using SVR for one week.

Table 5. Forecasting error for one week in February

February		18	19	20	21	22	23	24
PLS	Ave. (%)	3.87	6.57	11.41	14.08	15.48	3.42	8.16
	Max. (%)	16.47	21.07	34.41	37.80	34.70	8.13	23.64
ANN	Ave. (%)	6.54	10.09	10.29	9.97	5.58	4.94	6.70
	Max. (%)	13.20	20.18	23.54	26.65	19.34	15.50	17.31
SVR	Ave. (%)	4.95	5.55	6.32	8.81	4.82	3.98	5.01
	Max. (%)	9.82	14.60	19.64	16.51	10.27	10.13	9.61

*Ave.: Averaged forecasting error, Max.: Maximum forecasting error

all average error as well as consumption trend. This means that SVR can be effectively used in the management of a district heating system in which the number and type of operating facilities should be determined. Effects of the structure, number and type of input data on the performance of SVR method are yet to be investigated.

ACKNOWLEDGEMENTS

This work was sponsored by the Ministry of Knowledge Economy, Republic of Korea, as a part of the research project titled "Constitution of energy network using District heating energy" (Project No: 2007-E-ID25-P-02-0-000). The authors wish to thank them for their support.

REFERENCES

1. B. Jie, *Short-term load forecasting based on neural network and moving average*, Iowa State University (2002).
2. M. Ibrahim and R. Saifure, *IEEE Transactions on Power Systems*, **4**(4) (1989).
3. K. Lru, S. Subbarayan, R. R. Shoults, M. T. Manry, C. Kwan, F. L. Iewis and J. Naccarino, *IEEE Transactions on Power Systems*, **11**(2) (1996).
4. D. Alex and C. Timothy, *IEEE Transactions on Power Systems*, **5**, 1535 (1990).
5. I. Dejan, *FME Transactions*, **34**, 165 (2006).
6. T. Yalcinoz and U. Eminoglu, *Energy Conver. Manage.*, **46**, 1393 (2005).
7. A. J. Al-Shareef, E. A. Mohamed and E. Al-Judaibi, *One hour ahead load forecasting using artificial neural network for the western area of saudi arabia*, *International J. Electr. Systems Sci. Eng.* (2006).
8. Z. Ran, *Power system load forecasting using partial least square method*, 40th Southeastern Symposium, 169 (2008).
9. K. G. Min, C. H. Han and K. S. Chang, *Korean J. Chem. Eng.*, **5**(2), 2437 (1999).
10. C. K. Park, *Korean Operations Res. Manage. Soc.*, **23**(2), 75 (2006).