Korean J. Chem. Eng., **28**(7), 1505-1510 (2011)
DOI: 10.1007/s11814-011-0017-5

A new forecasting method for system marginal prices based on power supply and demand

Kee Jun Lee*, Tae Hwan Lee*, Lae-Hyun Kim, and Yeong Koo Yeo*,†**

*Department of Chemical Engineering, Hanyang University, Seoul 133-791, Korea
**Department of Chemical Engineering, Seoul National University of Science and Technology, Seoul 135-743, Korea
(*Received 16 October 2010 • accepted 19 January 2011*)

Abstract−A new forecasting scheme is presented of short-term system marginal price (SMP) using past data on power supply and demand as well as past cost data. The forecasting of SMP is one of the most significant factors in an electricity market in which power is produced by generators, transmitted by transmission companies, and distributed by suppliers according to new trading agreements. The accurate forecasting of SMP can significantly influence the profit of market participants. In this paper, a new methodology for day-ahead SMP forecasting is proposed using the law of supply and demand on power. The salient feature of the proposed approach is that it exhibits excellent predicting performance in short-term forecasting.

Key words: System Marginal Price (SMP), Price Forecasting, Short-term Forecasting

INTRODUCTION

Deregulation of the electric power industry raises many challenging issues. Forecasting electricity prices and quantities in daily power markets is the most essential task and basic to any decision making. Forecasting electricity loads has reached a comfortable state of performance. In electricity markets, companies that are involved in trading make extensive use of price prediction techniques either to bid or to hedge against volatility. Having reliable daily price forecast information, producers or energy service companies are able to delineate good bilateral contracts and make better financial decisions.

The Korea Power Exchange (KPX) is a competitive electricity market in which electricity is produced by generators, transmitted by a dispatcher and distributed by suppliers under new trading agreements. All electricity is sold to and bought from KPX. The system marginal price (SMP) is a major portion of the payment that generators obtain from KPX for their scheduled generating units. From a generator's point of view, it is very important to estimate what the SMP will be on the next day. The determination of the price of electricity in the electricity market is based on the principle of demand and supply as in the case of general commercial products. The highest cost among the costs of active electricity sources to fulfill hourly power demand becomes the SMP. KPX forecasts the power demand of the trading day and calls for one-day-ahead bids. Therefore, the proper prediction of the SMP can help a power generator to submit its generating unit prices as close to SMP as possible and, therefore, to increase the company's income on the next scheduling day.

A good price forecasting tool in deregulated markets should be able to capture the uncertainty associated with those prices. Factors that influence the electricity price include time (hour of the day, day of the week, month, year, and special days), operating reserves, historical prices and demand, bidding strategies, temperature effects, predicted power shortfall and generation outages. In general, the most significant element driving the electricity price is power demand. The commonly used series models are linear predictors, while electricity prices are in general a nonlinear function of its input features [1,2]. So, the behavior of price signal may not be completely captured by the time series techniques [3]. To resolve this problem, some black-box forecasting models such as neural networks (NN) and fuzzy neural networks (FNN) have been presented. Mandal et al. [4] NNs and FNNs have the capability of modeling the nonlinear input/output mapping functions. However, electricity price is a time variant signal and its functional relationships rapidly vary with time [5,6]. Thus, the derived information or extracted feature of the NN or FNN rapidly loses its value. While it seems that the NN or FNN learns well the training data, they may encounter large prediction errors in the test phase. Pedregal and Trapero propose the use of the dynamic harmonic regression for forecasting the hourly price time series of electricity markets [7]. Their model belongs to the class of the unobserved components models set up as a state space system. A certain regularity of the data is an important precondition for the successful application of neural networks [8]. When using classical statistical techniques, a stationary process is assumed for the data. But, in most cases, the assumption of stationarity has to be discarded. Besides, one has to bear in mind that different kinds of nonstationarities may exist [8]. Neural networks have already been used to solve problems such as load forecasting [9], component and system fault diagnosis, security assessment and unit commitment [10]. The wavelet transform converts a price series in a set of constitutive series. These series present a better behavior than the original prices series, and therefore, they can be predicted more accurately [11]. The reason for the better behavior of the constitutive series is the filtering effect of the wavelet transform.

Most of the forecasting methods presented so far have some disadvantages and are inconvenient to use in that they require large amount of past data without regarding to their accuracy. Forecasting by a compact model which is convenient to use would be most

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

attractive for electricity market. The fundamental and novel contribution of the paper is the introduction of a new compact model for forecasting SMP based on power supply and demand data as well as past cost data. The proposed model is very simple to use and requires the least amount of data without sacrificing prediction accuracy. The proposed forecasting equation exhibits tracking performance that is good enough to use in the electricity market.

SMP FORECASTING MODEL

Accurate forecasting means small differences between predicted values by the proposed model and actual price data. Analysis of prediction error for a certain period suggests whether the forecasting algorithm is appropriate to the revealed pattern or not. If the probability distribution is used for a given period, the magnitude of errors has to be determined for accurate forecasting, and criteria for representing uncertainty are required. In this work, we employ RMSE (root mean square error) and MAPE (mean absolute percentage error) to evaluate accuracy of forecasting algorithms. RMSE and MAPE are defined as the following:

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N_t}(\mathrm{P}_{i,actual}-\mathrm{P}_{i,estimated})^2} \quad (1)$$

$$\mathrm{MAPE}=\frac{\sum_{i=1}^{N_t}\left(\frac{|\mathrm{P}_{i,actual}-\mathrm{P}_{i,estimated}|}{\mathrm{P}_{i,actual}}\right)\times 100}{\mathrm{N}_t} \quad (2)$$

where N_t is the number of observations in test data set and $P_{i,actual}$ and $P_{i,estimated}$ are the actual and estimated values, respectively.

1. Estimation of Power Demand

Prediction of power demand and supply is required to apply the principle of supply and demand in the forecasting of the SMP. To accomplish the prediction we have to choose appropriate variables to use in the prediction. These variables can be determined by correlation analysis. To represent the extent of correlation, we use a correlation coefficient, denoted by "r", which represents numerically the correlation characteristics between two variables. The coefficient r takes a value between −1 and 1. The correlation coefficient between two variables can be calculated by the following sequence:

$$\mathrm{S(xx)}=\sum_{i=1}^{n}(\mathrm{x}_i-\bar{\mathrm{x}})^2=\sum_{i=1}^{n}\mathrm{x}_i^2-\left(\sum_{i=1}^{n}\mathrm{x}_i\right)^2/\mathrm{n}$$

$$\mathrm{S(yy)}=\sum_{i=1}^{n}(\mathrm{y}_i-\bar{\mathrm{y}})^2=\sum_{i=1}^{n}\mathrm{y}_i^2-\left(\sum_{i=1}^{n}\mathrm{y}_i\right)^2/\mathrm{n}$$

$$\mathrm{S(xy)}=\sum_{i=1}^{n}(\mathrm{x}_i-\bar{\mathrm{x}})(\mathrm{y}_i-\bar{\mathrm{y}})=\sum_{i=1}^{n}\mathrm{x}_i\mathrm{y}_i-\left(\sum_{i=1}^{n}\mathrm{x}_i\right)\left(\sum_{i=1}^{n}\mathrm{y}_i\right)/\mathrm{n}$$

$$\mathrm{r}=\frac{\mathrm{S(xy)}}{\sqrt{\mathrm{S(xx)S(yy)}}} \quad (3)$$

where x_i and y_i are i_{th} variables, and $\bar{x}$ and $\bar{y}$ are average of x and y. n is the number of variables. A high value of the correlation coefficient (r) over 0.8 indicates high correlation. If the value of r stays within the range from 0.6 to 0.8, we can say that there exists modest correlation. We can say there exists no correlation among variables if the value of r is lower than 0.4. The correlation coefficient is calculated for the purpose of the analysis of relations among variables. The correlation coefficient between current power demand and power demand 1 week ago was found to be 0.7649 and that between current power demand and power demand 1 day ago was 0.6348.

Based on this observation, we can construct the load estimation model as follows:

$$\mathrm{P}_d=\alpha\times\mathrm{P}_{dw}+\beta\times\mathrm{P}_{dd} \quad (4)$$

Where P_d is power demand, and P_{dd} and P_{dw} are power demand 1 day ago and 1 week ago, respectively. α and β are coefficients.

The amount of power demand of one week ago and the previous day shows significant monthly variations. To compensate for these variations and to reduce estimation errors, we can use α and β which take different value for each month. Calculated values of α and β are shown in Table 1. To assess predicting performance of the model (4), estimation was performed for one week (June 22, 2009-June 28, 2009) and compared with actual load data. Results are shown in Fig. 1. As can be seen, nice tracking performance is achieved by using the simple model (4). Estimation errors are MAPE =3.67 and RMSE=6.67.

2. Estimation of Power Supply

As in the estimation of power demand, variables are chosen for prediction of power supply through correlation coefficients. It was found that the power supply of the present day is highly correlated with the amount of supply of previous day as well as the maximum power demand. The correlation coefficient between the present power supply and the supply of previous day is 0.97 and that between the present power supply and the maximum demand is 0.67.

Table 1. Values of α and β for each month

Month	α	β	Month	α	β
March	0.4	0.6	August	0.2	0.8
April	0.4	0.6	September	0.05	0.95
May	0.4	0.6	October	0.55	0.45
June	0.05	0.95	November	0.25	0.75
July	0.25	0.75	December	0.27	0.73

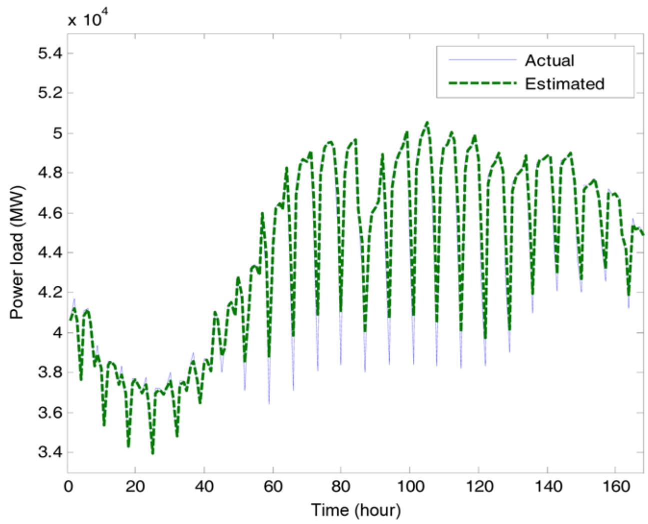

Fig. 1. Results of prediction on power load (June 22, 2009-June 28, 2009).

Appropriate variables to be used in the prediction are selected from the analysis of correlation. Based on these variables, we can develop the supply estimation model given by

$$P_s = P_{sd} \times \left(\frac{P_{max}}{P_{max\_d}}\right)^{0.06} \tag{5}$$

where P_s and P_{sd} are current power supply and power supply one day ago, respectively, and P_{max} and P_{max_d} are maximum power demand at present and one day ago, respectively.

The maximum demand obtained from the previous model (4) is used as the maximum supply in Eq. (5). To evaluate the predicting performance of the model (5), estimation was performed for March 1, 2009-December 31, 2009 and compared with actual supply data. Results are shown in Fig. 2. As can be seen, excellent prediction performance is achieved by using the prediction model (5). In this work, prediction with RMSE less than 20 is regarded as "permissible." Estimation errors are MAPE=0.67 and RMSE=0.92.

3. Prediction of SMP

In general, factors that impact the SMP include power demand, ambient temperature and variations in fuel prices. The main variable that derives the SMP is the power demand. But, in addition to seasonal variations, sometimes the relationship between the SMP and the power demand display distorted pattern as shown in Fig. 3. To improve forecasting performance, we can employ the supply and demand principle. If the suppliers try to set higher prices on specific items, production of these items increases. If the consumers believe that they can acquire those items at lower prices, production of those items decreases while demand increases. If demand on some items is lower than expected, the suppliers lower the prices and vice versa. Fig. 4 shows the well-known interrelationship between demand and supply. As the amount Q increases, the demand curve (D) and the supply curve (S) exhibit an inverse proportional relation to each other. Based on this observation, we can derive the following equation:

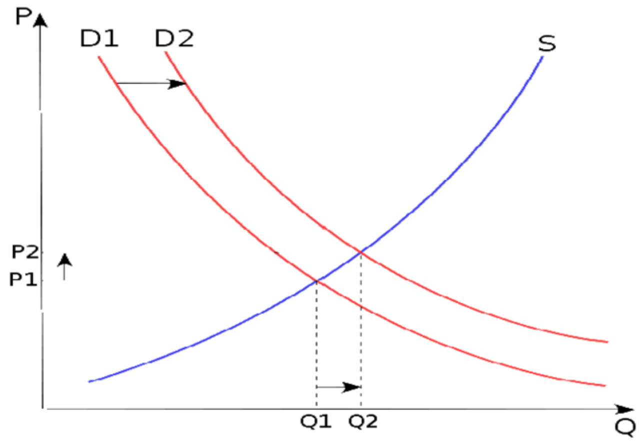

Fig. 4. The supply and demand curve.

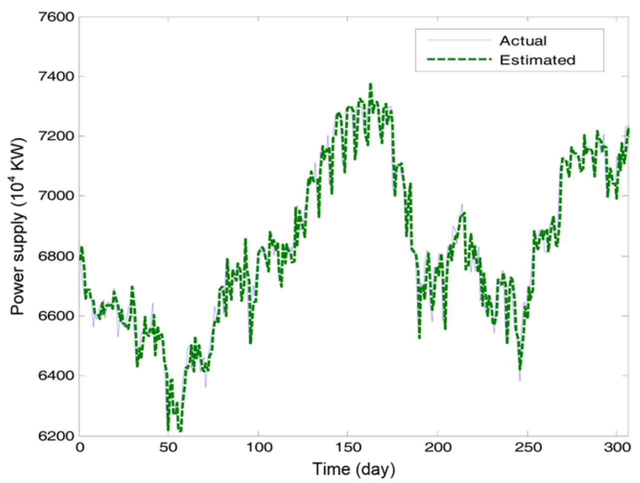

Fig. 2. Results of prediction on power supply (March 1, 2009-December 31, 2009).

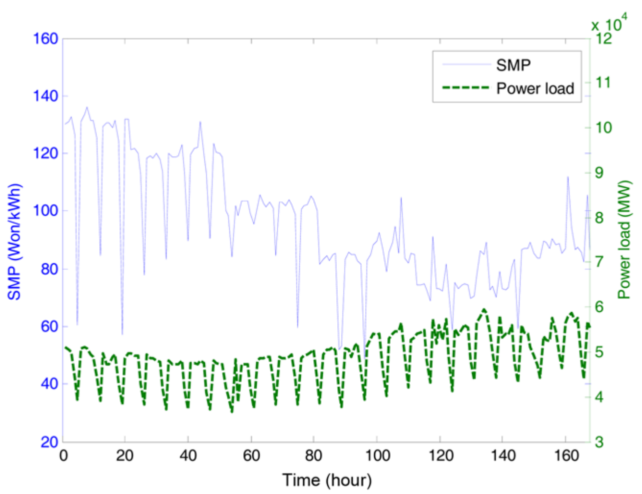

Fig. 3. Chronological demand and SMP curves (August 10, 2009-August 16, 2009).

$$SMP \times \frac{P_s}{P_d} = SMP_w \times \frac{P_{sw}}{P_{dw}} \tag{6}$$

where P_s and P_d are power supply and demand respectively, and P_{sw} and P_{dw} are power supply and demand on 1 week ago, respectively. SMP_w is the SMP of 1 week ago. Eq. (6) can be rearranged as follows:

$$SMP = SMP_w \times \frac{P_d}{P_{dw}} \times \frac{P_{sw}}{P_s} \tag{7}$$

The pattern of SMP data may be regular or irregular. So if we use the data as it is, it is impossible to find a similar data pattern. The pattern shown by new forecasting Eq. (7), being based on data of past seven days (one week), depends upon the pattern of past seven days. Forecasting may be affected significantly by the variations in seasonal fuel prices. For this reason, regular and irregular patterns must be classified according to the season. Thus, regular and irregular patterns for summer are different from those for winter or spring. We may anticipate similar patterns during spring and fall. Data of the past seven days showing irregular pattern are replaced by data of 14 days ago to reduce forecasting errors to give Eq. (8). Variations in fuel prices matter usually in the first week of a month. To incorporate the effect of variations in fuel prices and to decrease forecasting error, a term denoting variations in fuel prices is added to Eq. (7) to give Eq. (10). It should be noted that Eq. (9) is used only in the forecasting of the first week of each month.

$$\mathrm{SMP} = \mathrm{SMP}_{w2} \times \frac{\mathrm{P}_d}{\mathrm{P}_{dw2}} \times \frac{\mathrm{P}_{sw2}}{\mathrm{P}_s} \tag{8}$$

$$\mathrm{SMP} = \mathrm{SMP}_{w2} \times \frac{\mathrm{P}_d}{\mathrm{P}_{dw}} \times \frac{\mathrm{P}_{sw}}{\mathrm{P}_s} \times \frac{\mathrm{C}_{LNG}}{\mathrm{C}_{LNGw}} \tag{9}$$

where SMP_w and SMP_{w2} are SMPs of one week and two weeks ago, respectively, and P_{dw2} and P_{sw2} are power demand and supply on two weeks ago, respectively. C_{LNG} and C_{LNGw} denote cost of LNG at present and one week ago, respectively.

RESULTS AND DISCUSSION

The past SMP data used in the proposed model were collected from Korea Electricity Trading Office. The information available includes hourly price historical data of the one and two weeks previous to the day of the week whose prices are to be predicted during the period from March 31, 2009 to December 31, 2009.

Because the SMP exhibits different patterns according to the season, we select a typical week for each season. The spring week selected is the third week of March (March 16, 2009-March 22, 2009), the summer week is the fourth week of June (June 22, 2009-June 28, 2009), the autumn week is the fourth week of October (October 19, 2009-October 25, 2009), and the winter week is the second week of December (December 7, 2009-December 13, 2009).

Table 2 summarizes the forecasting errors for one week of each season. We can see that good forecasting is achieved regardless of seasons, which means that we do not need to take into account seasonal characteristics to use the proposed forecasting model. Incorporation of price data of the past one or two weeks means a relatively recent trend is taken into account in the forecasting to give more accurate prediction. The power demand depends upon the ambient temperature, and changes in ambient temperature are significant factors causing fluctuations in power demand. For this reason, it is necessary to include the ambient temperature as a key variable in the forecasting model. But the ambient temperature of a day does not show much difference from the day of one week ago. Thus, we do not need to include the temperature as a variable in the forecasting model if the model is based on data of one week ago. Moreover, temperature variations below 5 °C do not affect the forecasting significantly. To illustrate this observation, forecasting errors were determined and compared for the forecasting models with and without temperature term. In this case the rate of temperature change is represented in Kelvin (K). The difference in forecasting error between the model with temperature term and that without temperature term was found to be 0.01% MAPE, which can be ignored.

Fig. 5 provides the forecasting data for the third week of March. The prediction behavior of the proposed model for the spring week-

Table 2. The SMP forecasting errors for 1 week of each season

Test period	MAPE	RMSE
March 16, 2009-March 22, 2009	3.27	6.13
June 22, 2009-June 28, 2009	5.01	7.21
October 19, 2009-October 25, 2009	6.08	8.81
December 7, 2009-December 13, 2009	3.59	5.99

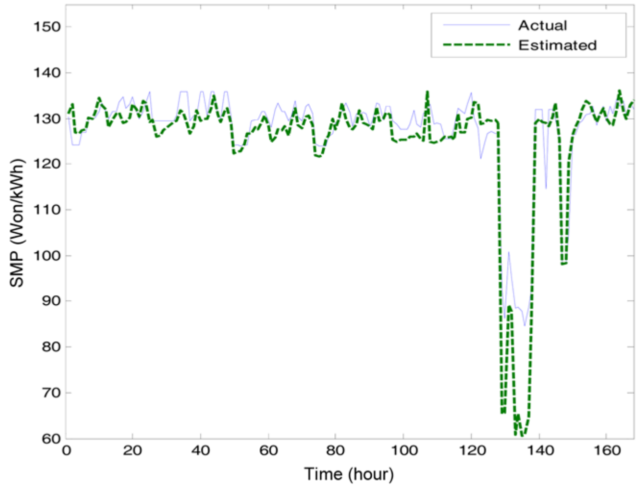

Fig. 5. Prediction of SMP (March 16, 2009-March 22, 2009).

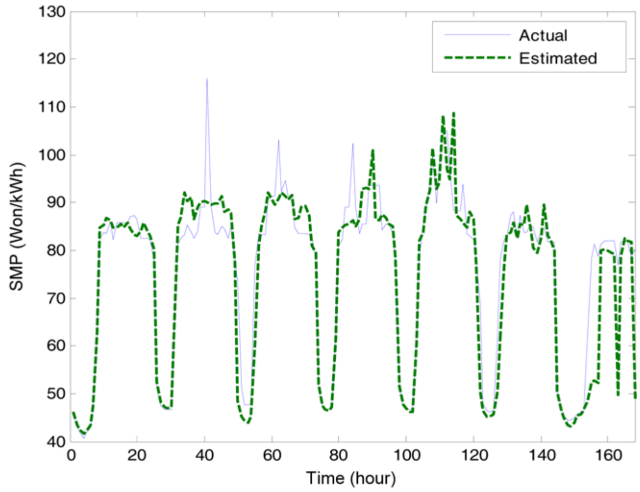

Fig. 6. Prediction of SMP (June 22, 2009-June 28, 2009).

days is very appropriate with a weekly error of MAPE=1.94 and RMSE=3.07 (time span: 1-144 hours). But, results of forecasting for the spring weekend exhibit high prediction error (MAPE=11.28, RMSE=14.39, time span: 144-168 hours). This trend can be seen in other seasons. Forecasting data for a summer week is shown in Fig. 6. For the fourth week of June, the prediction errors for weekdays are MAPE=4.23 and RMSE=5.64 (time span: 1-144 hours) and those for the weekend are MAPE=9.70 and RMSE=13.19 (time span: 144-168 hours). Fig. 7 shows the forecasting data for the fourth week of October. In this case, the prediction errors for weekdays are MAPE=4.67 and RMSE=6.54 (time span: 1-144 hours) and those for the weekend are MAPE=14.57 and RMSE=15.34 (time span: 144-168 hours). Even with the relatively high RMSEs for weekends, all RMSEs are less than the criterion of 20, which means the predictions are acceptable. As for the winter week, the performance of the proposed model is good and accuracy is reasonable enough both for weekdays and weekend. The prediction errors for the second Sunday of December, 2009, are MAPE=2.48 and RMSE=2.77. In summary, the daily average RMSE is 11.65 with the lowest value of 2.38×10^{-6}. As described before, irregular life patterns in week-

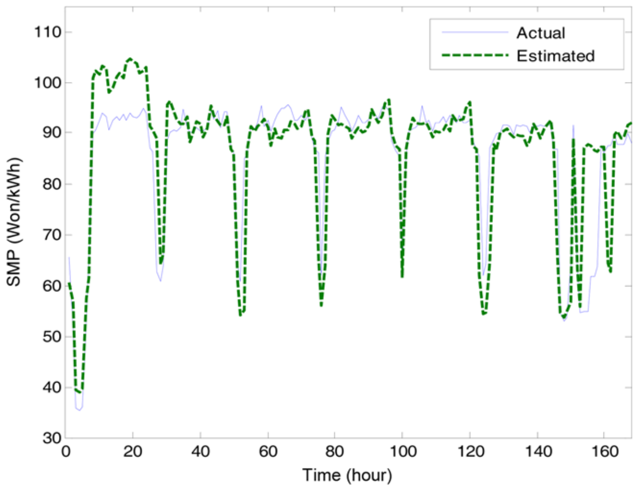

Fig. 7. Prediction of SMP (October 19, 2009-October 25, 2009).

Table 3. The SMP forecasting errors for weekdays and weekends

Test period	MAPE	RMSE
March 16, 2009-March 21, 2009 (weekdays)	1.94	3.06
March 22, 2009 (weekend)	11.27	14.39
June 22, 2009-June 27, 2009 (weekdays)	4.23	5.63
June 28, 2009(weekend)	9.69	13.18
October 19, 2009-October 24, 2009 (weekdays)	4.67	6.54
October 25, 2009 (weekend)	14.57	15.34
December 7, 2009-December 12, 2009 (weekdays)	3.78	6.37
December 13, 2009 (weekend)	2.48	2.76

ends contribute to the relatively large errors. Table 3 summarizes above results.

In the proposed model, the prediction for a specific day of the week is performed based on data of the same day of one and two weeks ago. This scheme is adequate especially for weekdays when life patterns of consumers are steady (i.e., do not show unusual variations). Relatively large errors in the weekend indicate life style patterns of consumers are irregular. We can assume that the life pattern of a weekend depends upon the weather conditions which influence outdoor activities. From the statistical analysis of weather data, it is found that the relatively high prediction errors for weekends are correlated with average ambient temperatures. For example, the average temperature on March 22, 2009, is 4.9 °C higher than that on one week ago with cloudy on both days. Due to the increase of the average temperature, outdoor activities may be increased to give unusual high prediction errors. The difference of the daily temperature range between June 28, 2009, and June 21, 2009, is only 0.9 °C but there was heavy rainfall on June 21 while it was cloudy on June 28. It is obvious that weather conditions affect highly the weekend outdoor activities again to give high prediction errors. Winter exhibits somewhat different patterns. Cold weather prohibits outdoor activities, which may be the major reason that the performance of the proposed model is good for winter and accuracy is reasonable enough both for weekdays and weekend. In short, we have to take into account weather conditions for the weekend forecasting. Inclusion of information on the weather into the SMP forecasting

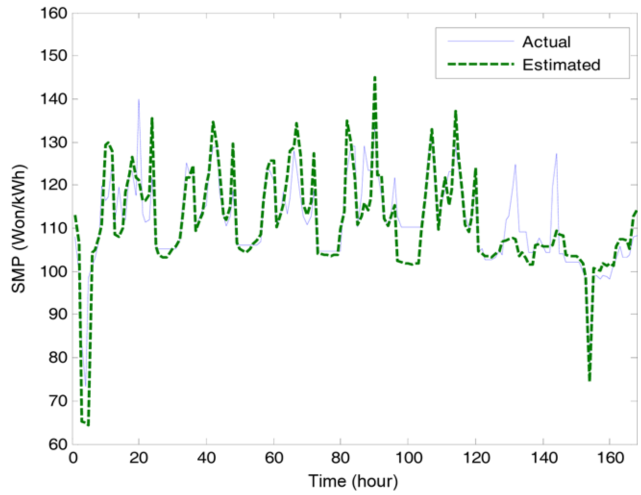

Fig. 8. Prediction of SMP (December 7, 2009-December 13, 2009).

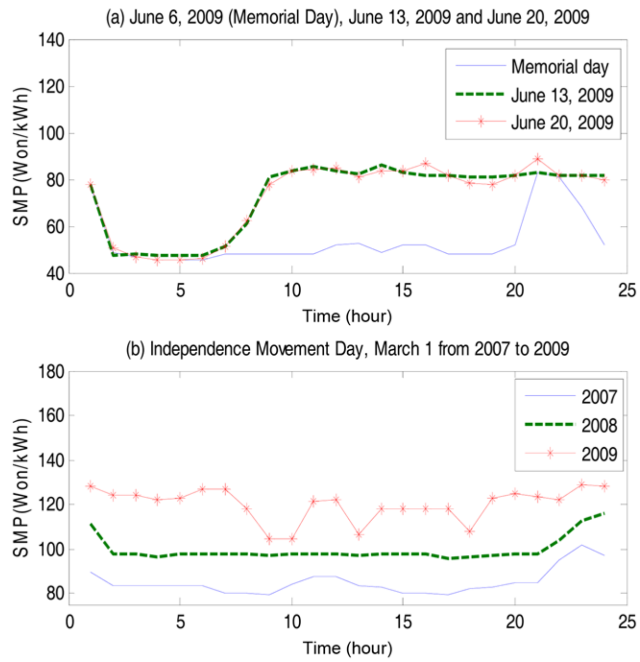

Fig. 9. (a) Comparison of SMP forecasting between holiday and normal days (b) Comparison of SMP forecasting on a holiday for 3 years (2007-2009).

model is very complicated and requires detailed analysis on weather data for long period. For simplicity, we do not include weather data in the proposed model.

Forecasting results for holidays or festive days show a similar pattern to weekends. For accurate prediction, we need information on the day of week for a given holiday. For example, we should not employ the forecasting scheme for weekdays if, for instance, Tuesday is the Memorial day. Differences in the forecasting patterns between a holiday and an ordinary day can be seen in Fig. 9. In Fig. 9(a), the solid line denotes SMP forecasting results for the Memorial day, the dotted line indicates those for one week after the Memorial day, and the starred (*) line represents those for two weeks after the Memorial day. Fig. 9(b) shows SMP forecasting

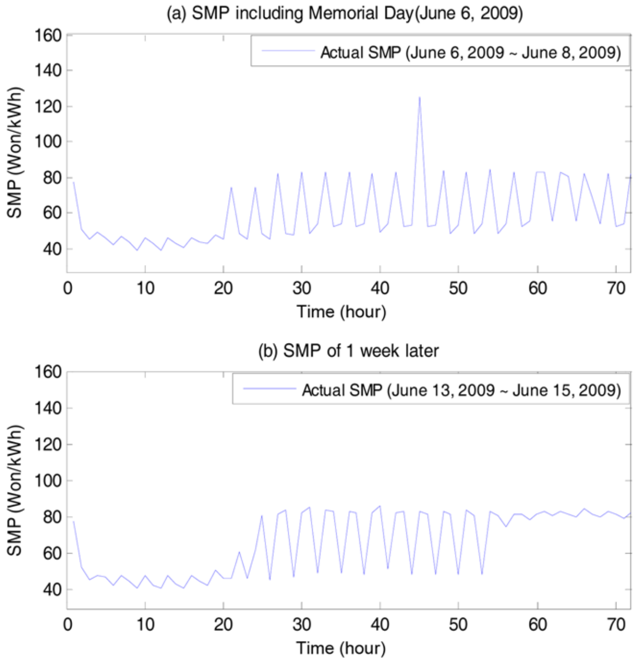

Fig. 10. Effects of a holiday on the SMP forecasting: (a) the Memorial day (June 6, 2009) and 1 and 2 days after the holiday; (b) normal weekdays (June 13, 2009-June 15, 2009).

results for the anniversary of the Independence Movement of March 1st, 1919. We can see quite a different prediction pattern for each year, which may be caused by the fact that March 1st of each year comes on a different day of the week. We can also see somewhat different patterns from those of other holidays and weekends.

When a holiday comes in between weekends (for example, Wednesday is the Memorial day), the SMP prediction for one day and two days after the holiday is influenced by that for the holiday. Fig. 10 illustrates SMP forecasting results for a period including the Memorial day (Fig. 10(a)) and those after one week (Fig. 10(b)). As shown in Fig. 10, we can see quite a different behavior between the week containing a holiday and the normal week without holidays. The effect of a holiday on the following few days in the SMP forecasting is clear. The problem is that we do not have enough data on holidays (i.e., information on past SMPs and weather conditions for holidays) because a holiday comes one time in a year. Even if we collected data for 10 years, we would have only 10 data for holidays. Thus, it is inevitable to suffer some SMP forecasting errors for holidays.

CONCLUSION

We propose a new forecasting scheme of short-term system marginal price (SMP) using the law of supply and demand on power. Prediction of SMP is based upon past data on power supply and demand as well as cost data of one and two weeks ago. The performance of the proposed forecasting model is good and the accuracy is reasonable enough both for weekdays and weekend. The average RMSE is 11.65 with the minimum value of 2.38×10^{-6}. The relatively large RMSE originates from the irregular life patterns in holidays and weekends. If we perform the SMP forecasting excluding holidays and weekends, the average RMSE is reduced to 9.65. The salient feature of the proposed approach is that it exhibits excellent predicting performance in short-term forecasting.

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.

NOMENCLATURE

C_{LNG} : cost of LNG [Won/kWh]
C_{LNHw} : cost of LNG 1week ago [Won/kWh]
P_d : power demand [MW]
P_{dd} : power demand 1 day ago [MW]
P_{dw} : power demand 1 week ago [MW]
P_{dw2} : power demand 2 weeks ago [MW]
P_s : power supply [10^4 KW]
P_{sd} : power supply 1 day ago [10^4 KW]
P_{sw} : power supply 1 week ago [10^4 KW]
P_{sw2} : power supply 2 weeks ago [10^4 KW]
P_{max} : max power demand during whole day [10^4 KW]
P_{max_d} : max power demand 1 day ago [10^4 KW]
T : ambient temperature [ºC]
T_w : ambient temperature 1 week ago [ºC]
SMP_{w2} : SMP 1 week ago [Won/kWh]
SMP_w : SMP 2 weeks ago [Won/kWh]

REFERENCES

1. P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi and A. K. Srivastava, *IEEE Trans. Power Syst.*, **22**(4), 2058 (2007).
2. P. Mandal, K. Anurag, Srivastava, T. Senjyu and M. Negnevitsky, *Int. J. Energy Res.*, **34**(6), 507 (2009).
3. N. Amjady and M. Hemmati, *IEEE Power Energy Magazine*, **4**(2), 20 (2006).
4. N. Amjady, *IEEE Trans. Power Syst.*, **21**(2), 887 (2006).
5. A. M. Gonzalez, A. M. S. Roque and J. Garcia-Gonzalez, *IEEE Trans. Power Syst.*, **20**(1), 13 (2005).
6. D. J. Pedregal and J. R. Trapero, *Energy Convers. Manage.*, **48**(5), 1710 (2007).
7. A. J. R. Reis and A. P. A. da Silva, *IEEE Trans. Power Syst.*, **20**(1), 189 (2005).
8. K. Y. Lee, Y. T. Cha and J. H. Park, *IEEE Trans. Power Syst.*, **7**(1), 124 (1992).
9. A. J. Conejo, M. A. Plazas and R. Espinola, *IEEE Trans. Power Syst.*, **20**(2), 1035 (2005).
10. F. J. Nogales, J. Contreras, A. J. Conejo and R. Espinola, *IEEE Trans. Power Syst.*, **17**(2), 342 (2002).
11. D. S. Kirschen, *IEEE Trans. Power Syst.*, **18**(2), 520 (2003).
12. R. Baldick, R. Grant and E. Khan, *Journal of Regulatory Economics*, **25**(2), 143 (2004).