A Hybrid Correction Technique of Missing Load Data Based on Time Series Analysis

Chan-Joo Lee*, Jong-Bae Park†, Jae-Yong Lee*, Joong-Rin Shin* and Chang-Ho Lee**

Abstract - Traditionally, electrical power systems had formed the vertically integrated industry structures based on the economics of scale. However, power systems have been recently reformed to increase their energy efficiency. According to these trends, the Korean power industry underwent partial reorganization and competition in the generation market was initiated in 2001. In competitive electric markets, accurate load data is one of the most important issues to maintaining flexibility in the electric markets as well as reliability in the power systems. In practice, the measuring load data can be uncertain because of mechanical trouble, communication jamming, and other issues. To obtain reliable load data, an efficient evaluation technique to adjust the missing load data is required. This paper analyzes the load pattern of historical real data and then the tuned ARIMA (Autoregressive Integrated Moving Average), PCHIP (Piecewise Cubic Interpolation) and Branch & Bound method are applied to seek the missing parameters. The proposed method is tested under a variety of conditions and also tested against historical measured data from the Korea Energy Management Corporation (KEMCO).

Keywords: ARIMA (Autoregressive Integrated Moving Average), Load Forecast, PCHIP (Piecewise Cubic Interpolation), Branch and Bound Algorithm, Time Series Analysis.

1. Introduction

Recently, electrical power systems around the world have been reformed in efforts to increase the energy efficiency of power systems in general. According to these trends, the Korean power industry, vertically integrated and centrally planned and operated, has been partially restructured and the competitive generation market was opened in the country in 2001[1].

In the competitive electricity market, both electric utilities and energy consumers will be variously changed. In particular, energy consumers may require a variety of information to analyze their load patterns and electric prices. Among information, the accurate load data is one of the most important factors used to maintain the flexible electricity market as well as reliable power systems.

In practice, the measured load data may contain uncertainty, due to mechanical trouble, communications jamming, and other issues. To obtain reliable load data, an efficient evaluation technique to adjust the missing load data is vital.

This paper analyzes the load pattern of the historical real data and then the extended ARIMA (Autoregressive Integrated Moving Average) model, PCHIP (Piecewise Cubic Interpolation), and Branch & Bound method are applied to seek the missing parameters. In addition, the proposed method is tested under a variety of conditions and also tested using historical measured data from the Korea Energy Management Corporation (KEMCO).

2. A correction technique of missing load data

2.1 ARIMA model

The analysis of time series has a main goal to predict future values of the time series variable. Several models using time series analysis have previously been proposed with load forecasting, such as short-term load forecasting, mid-term load forecasting, and long-term load forecasting, resulting in the presentation of numerous studies [2-6].

The measured load data in power systems have a non-stationary behavior since the electrical demand does not have a fixed mean level. To analyze time series with non-stationary behavior, the stationary processing procedure that has a constant mean, variance, and autocorrelation through time is considered necessary. The ARIMA methodology to analyze time series with non-stationary behavior is developed by Box and Jenkins [7, 8]. ARIMA processes are a class of stochastic processes used to analyze time series and the general scheme is as follows:

The Transactions of the KIEE, Vol.53A, No.7, JULY. 2004, pp.405-413 : A paper recommended and approved by the Editorial Board of the KIEE Power Engineering Society for translation for the KIEE International Transaction on PE.

† Corresponding Author: Dept. of Electrical Engineering, Konkuk University, Korea. (jbaepark@konkuk.ac.kr)

* Dept. of Electrical Engineering, Konkuk University, Korea. (jubily@konkuk.ac.kr, jrshin@kkucc.konkuk.ac.kr)

** Korea Electrotechnology Research Institute, Korea. (chrhee@keri.re.kr)

Received March 26, 2004 ; Accepted April 29, 2004

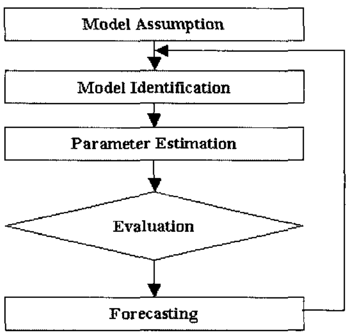


Fig. 1 Time series analysis procedure by Box-Jenkins

In the step of model assumption, a class of models is formulated assuming certain hypotheses. In the step of model identification, a model is identified for the observed data. To identify the specific number and type of ARIMA parameters to be estimated, the major tools used in the identification step are auto correlation (ACF) and partial autocorrelation (PACF). At the parameter estimation step, the model parameters are estimated. The autoregressive parameters, the number of differencing passes, and moving average parameters are decided in this stage. If the hypotheses of the model are validated, proceed to the forecasting step, and otherwise proceed to the identification step to refine the model. Finally, the model is ready for forecasting. Therefore, the load forecasting for each customer can be obtained by these procedures.

The ARIMA model consists of autoregressive (AR) model, moving average (MA) model, and autoregressive moving average (ARMA) model. The general formulation of the AR model is as follows [8]:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ...\phi_p y_{t-p} + \varepsilon_t \quad (1)$$

where, ϕ_i : autoregressive parameters,

p : autoregressive order,

ε_t : white noise.

Each element in the time series can also be affected by past errors that cannot be accounted for in the autoregressive component. The general formulation of the MA model is the following:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} \quad (2)$$

where, ϕ_i : moving average parameters,

p : moving average order.

The ARMA model forecasts a current value by means of a linear combination of previous values, previous noises and current noise. This model consists of the time series model of AR and MA. The general formulation of the ARMA model is the following:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ...\phi_p y_{t-p} + \varepsilon_t$$
$$- \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} \quad (3)$$

The input data of the ARIMA model must be stationary. If the input data have non-stationary behaviors, these data are needed to be a differencing until they have stationary behaviors. From equation (3), the AR model using the backward-shift operator is as follows:

$$(1 - \phi_1 B^1 - \phi_2 B^2 - ... - \phi_p B^p) y_t$$
$$= (1 - \theta_1 B^1 - \theta_2 B^2 - ... - \theta_q B^q) \varepsilon_t$$
$$\rightarrow (1 - \sum_{i=1}^{p} \phi_i B^i) y_t = (1 - \sum_{j=1}^{q} \theta_j B^j) \varepsilon_t$$
$$\rightarrow \phi_p(B) y_t = \theta_q(B) \varepsilon_t \quad (4)$$

where, y_t : output,

B : backward-shift operator; $B^m y_t = y_{t-m}$.

The purpose of the ARIMA model is forecasting the value of (y_t). This model means that the current value is influenced by the previous values and noises.

2.2 The extended ARIMA model

The electrical demand has different values of level according to the hourly, daily, weekly, and seasonal demand. These features can be used as important factors to model demand function. If y_t denotes the electrical demand at time t, the proposed general ARMA formulation is the following:

$$\frac{w_1 + w_2}{d_k} = \frac{1}{2}\left(\frac{w_1}{\delta_{k-1}} + \frac{w_2}{\delta_k}\right) \quad (5)$$

where, $\phi_p^h(B)$: hourly autoregressive model,

δ_{k-1} : daily autoregressive model,

$\phi_p^w(B)$: weekly autoregressive model.

$X^* = [x_1, x_2, ...x_n]$: hourly moving average model,

$\theta_q^d(B)$: daily moving average model,

$\theta_q^w(B)$: weekly moving average model.

As can be seen in equation (5), demand function at time t can be obtained as the product of the hourly, daily, and weekly model. However, the model in equation (5) is not sufficiently general to include the main features of the load data since it has non-stationary behavior. To reflect these features of load data, a suitable differencing to be stationary is needed. The extended ARIMA model at any specific time t can be formulated as follows:

$$\phi_p^h(B)\phi_p^d(B)\phi_p^w(B)(1-B)^d y_t = \theta_q^h(B)\theta_q^d(B)\theta_q^w(B)\varepsilon_t \quad (6)$$

where, d : difference operator.

In equation (6), the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q), i.e. $ARIMA(p,d,q)$, are determined by heuristic method. The selected $ARIMA(p,d,q)$, is as follows:

$$\begin{aligned}
&(1-\phi_1^h B^1-\phi_2^h B^2-\phi_3^h B^3-\phi_4^h B^4-\phi_5^h B^5)\\
&\times(1-\phi_{24}^d B^{24}-\phi_{48}^d B^{48}-\phi_{72}^d B^{72}-\phi_{96}^d B^{96}-\phi_{120}^d B^{120})\\
&\times(1-\phi_{168}^w B^{168}-\phi_{336}^w B^{336}-\phi_{504}^w B^{504}-\phi_{672}^w B^{672})\\
&\times(1-B^1)y_t\\
&=(1-\theta_1^h B^1-\theta_2^h B^2)\times(1-\theta_{24}^d B^{24}-\theta_{48}^d B^{48})\\
&\times(1-\theta_{168}^w B^{168}-\theta_{336}^w B^{336})\varepsilon_t \quad (7)
\end{aligned}$$

The term of $(1-\phi_1^h B^1-\phi_2^h B^2-\phi_3^h B^3-\phi_4^h B^4-\phi_5^h B^5)$ in equation (7) denotes the time series of the previous 5 hours to predict the next hour as the hourly autoregressive model. The term of $(1-\phi_{24}^d B^{24}-\phi_{48}^d B^{48}-\phi_{72}^d B^{72}-\phi_{96}^d B^{96}-\phi_{120}^d B^{120})$ denotes the time series of the previous 5 days and the term of $(1-\phi_{168}^w B^{168}-\phi_{336}^w B^{336}-\phi_{504}^w B^{504}-\phi_{672}^w B^{672})$ denotes the previous 4 weeks. The term of $(1-B^1)$ related to the stationary property of the series denotes the first order differentiation. The moving average parameters use the previous 2 hours.

2.3 Piecewise Cubic Interpolation

To predict future values using the ARIMA model, the guarantee of continuous data through any time is needed. Without ensuring continuous data, the periodicity and identification of model cannot be considerable [7-10]. In practice, the measured load data at the customers' sites contain missing blocks due to mechanical difficulty, communication jamming, and other aspects. These data with the missing blocks may break the reliability of the outcome from almost all the prediction procedures. To obtain the reliable load data, an efficient and reliable technique to adjust the missing load data is essential.

In this paper, the PCHIP (Piecewise cubic interpolation) algorithm, which is one of the numerical algorithms, is used to correct the missing load data. The general function can be expressed as follows:

$$\begin{aligned}
P(x)&=\frac{3h_k s^2-2s^3}{h_k^3}y_{k+1}+\frac{h_k^3-3h_{ks}^2+2s^3}{h_k^3}y_k\\
&+\frac{s^2(s-h_k)}{h_k^2}d_{k+1}+\frac{s(s-h_k)^2}{h_k^2}d_k \quad (8)
\end{aligned}$$

let, $h_k = x_{k+1}-x_k$, $s = x-x_k$, $d_k=\delta_{k-1}$,

$$d_{k+1}=\frac{y_{k+2}-y_{k+1}}{x_{k+2}-x_{k+1}},\ \delta_k=\frac{y_{k+1}-y_k}{h_k}$$

where, h_k : the length of the k-th subinterval,

d_k : the slope of the k-th subinterval, $(\delta_k \neq d_k)$

δ_k : the divided difference.

This function on the interval $x_k \le x \le x_{k+1}$ satisfies four interpolation conditions as follows:

i) $P(x_k)=y_k$ ii) $P(x_{k+1})=y_{k+1}$

iii) $P'(x_k)=d_k$ iv) $P'(x_{k+1})=d_{k+1}$

The slope d_k can be defined as follows:

- δ_{k-1} and δ_k have opposite signs,
- either δ_{k-1} or δ_k is zero,
- δ_{k-1} and δ_k have identical signs.

If δ_{k-1} and δ_k have contrasting signs, or if either of them is zero, the value of d_k is zero. If δ_{k-1} and B have identical signs and the two intervals have the same length, the slope d_k is as follows:

$$\frac{1}{d_k}=\frac{1}{2}\left(\frac{1}{\delta_{k-1}}+\frac{1}{\delta_k}\right) \quad (9)$$

If δ_{k-1} and δ_k have equivalent signs, the slope $P(x_k)=y_k$ is as follows:

$$\frac{w_1+w_2}{d_k}=\frac{1}{2}\left(\frac{w_1}{\delta_{k-1}}+\frac{w_2}{\delta_k}\right) \quad (10)$$

where, $w_1=2h_k+h_{k-1}$, $w_2=h_k+2h_{k-1}$.

In this section, the PCHIP algorithm, which is one of the numerical algorithms, is presented. This algorithm is selected for use in this paper in order to obtain reliability of the outcome by ARIMA technique.

2.4 Branch and Bound

The optimization problem is largely classified into a continuous problem and discrete problem. Generally, optimization of the discrete problem is very difficult since the discrete problem requires computationally more time than the continuous problem. The methods used to solve discrete optimization problems are Exhaustive Enumeration, Branch and Bound, and Dynamic Programming [11]. The branch and bound algorithm is a general method applied to various combinatorial optimization problems. This algorithm is one of the search-based enumeration techniques that enumerate the entire solution space implicitly.

The measured load data at customers' sites may have a missing block in practice. Since there are several factors, such as mechanical difficulty, communications jamming, and maintenance schedule, it is difficult to analyze the reason for the missing block in the measured load data. In this paper, the branch and bound algorithm is use to search the missing blocks in the total demand. The flowchart of the correction technique for missing load data is as follows:

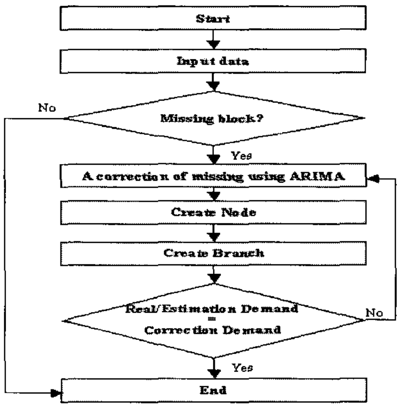


Fig. 2 Flowchart of a correction technique

In the flowchart, if input data contain missing blocks, the missing load data are corrected using ARIMA technique. Since the corrected demand from the ARIMA model may be larger than or equal to the measured demand at customers' sites, the evaluation of the corrected demand is needed. The branch and bound algorithm in this paper is used as the evaluation technique of the corrected demand. The objective function is as follows:

$$Min \sum_{i=1}^{I} |(ME_i - FCE_i)| \quad (11)$$

where, ME_i: the measured or estimated demand[MWh],

FCE_i: the corrected demand[MWh],

I: a time domain.

To evaluate the corrected demand, the relation between the measured demand and the corrected demand is defined as a minimization problem. The calculation procedure of the branch and bound algorithm is as follows:

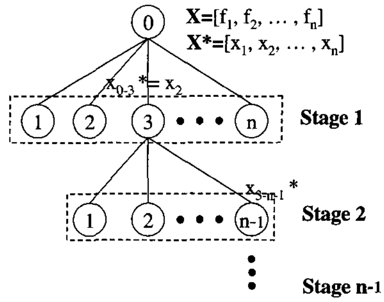


Fig. 3 Branch and bound algorithm

As can be seen in Fig. 3, the branch and bound algorithm consists of the stage, node, edge, and branch. Where, X denotes the free variables, $X^* = [x_1, x_2, \ldots, x_n]$ is a set of the corrected demand in each of the missing blocks, and the number of missing blocks is n.

In this paper, the branch and bound algorithm is only used to seek and evaluate the missing parameter of the corrected demand.

3. Case Studies

3.1 Case 1: Residential

The historical measured load data from the Korea Energy Management Corporation (KEMCO) are utilized to test the proposed correction technique. Numerical results with the ARIMA model are presented by using SAS, which is one of the common statistics programming tools. To test the validity of the proposed correction technique, the assumption for missing blocks is shown in Table 1.

Table 1 Assumption of missing load data

No.	Missing Period	Real value [Wh]	Assumption
1	07:31-09:15, July 29th	8318.15	Missing
2	17:01-18:45, July 29th	8234.10	Missing
3	21:31-24:15, July 31st	13020.00	Missing
4	08:16-10:00, July 31st	9350.25	Missing
5	16:01-17:45, July 31st	12019.30	Missing
6	03:16-05:00, October 1st	3938.55	Missing
7	12:46-14:30, October 1st	13485.10	Maintenance

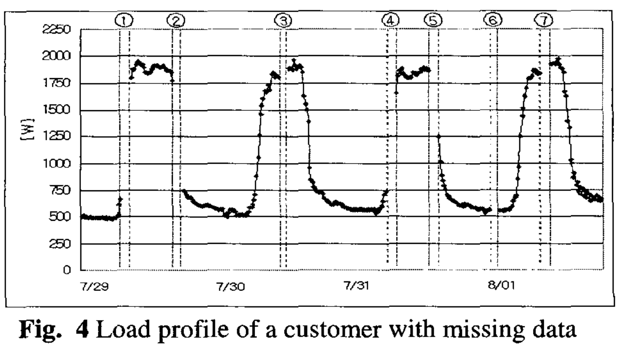


Fig. 4 Load profile of a customer with missing data

Fig. 4 indicates the load profile of a customer with missing data.

Fig. 5 presents the results of the corrected load data using the proposed algorithms, the extended ARIMA model and PCHIP.

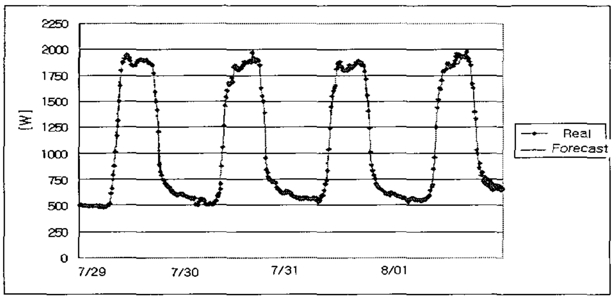


Fig. 5 Results from the correction technique

In Fig. 5, the real and forecast values denote the measured and corrected load data, respectively. Since the ARIMA model needs data from the previous 5 hours, the result of the first missing block is obtained by using PCHIP and all other results are obtained by using the extended ARIMA model. Tables 2 and 3 show the comparison and summary of result data for case 1 respectively.

Table 2 Comparison of measured data, estimated data, and error

No.	Time	Real	Forecasting	Error
1 (July, 29th)	07:31-07:45	791.7	748.7	5.43%
	07:46-08:00	882	874.1	0.90%
	08:01-08:15	1011.7	1026.9	1.50%
	08:16-08:30	1170.6	1195.6	2.14%
	08:31-08:45	1310.7	1368.5	4.41%
	08:46-09:00	1499.8	1533.9	2.27%
	09:01-09:15	1651.65	1680.1	1.72%
2 (July, 29th)	17:01-17:15	1589.7	1660.6	4.46%
	17:16-17:30	1503.6	1515.2	0.77%
	17:31-17:45	1413.3	1351.3	4.39%
	17:46-18:00	1207.5	1183	2.03%
	18:01-18:15	898.8	1024.4	13.97%
	18:16-18:30	829.5	889.6	7.25%
	18:30-18:45	791.7	792.6	0.11%
3 (July, 30th)	12:46-13:00	1815.4	1808	0.41%
	13:01-13:15	1829.1	1813.6	0.85%
	13:16-13:30	1850.1	1822	1.52%
	13:31-13:45	1885.8	1832.6	2.82%
	13:46-14:00	1883.7	1844.5	2.08%
	14:01-14:15	1874.2	1857.2	0.91%
	14:16-14:30	1881.6	1869.8	0.63%
4 (July, 31st)	07:31-07:45	837.9	904.01	7.89%
	07:46-08:00	1041.6	1050.1	0.82%
	08:01-08:15	1241.1	1244.4	0.27%
	08:16-08:30	1452.1	1407.7	3.06%
	08:31-08:45	1552.9	1487.3	4.22%
	08:46-09:00	1598.1	1535.2	3.94%
	09:01-09:15	1626.4	1607.2	1.18%
5 (July, 31st)	17:01-17:15	1850.1	1873.3	1.25%
	17:16-17:30	1855.3	1873.4	0.98%
	17:31-17:45	1838.5	1859.6	1.15%
	17:46-18:00	1792.3	1832.3	2.23%
	18:01-18:15	1709.4	1618.7	5.31%
	18:16-18:30	1563.4	1560.3	0.20%
	18:30-18:45	1410.1	1501.4	6.47%
6 (Oct., 1st)	12:46-13:00	570.15	565.48	0.82%
	13:01-13:15	569.1	564.77	0.76%
	13:16-13:30	567	557.78	1.63%
	13:31-13:45	555.45	560.8	0.96%
	13:46-14:00	564.9	553.88	1.95%
	14:01-14:15	554.4	557.8	0.61%
	14:16-14:30	557.55	560.37	0.51%
7 (Oct., 1st)	07:31-07:45	1855.3	1842.7	0.68%
	07:46-08:00	1897.3	1876.2	1.11%
	08:01-08:15	1938.3	1867.8	3.64%
	08:16-08:30	1957.2	1863.4	4.79%
	08:31-08:45	1937.2	1857.7	4.10%
	08:46-09:00	1955.1	1862	4.76%
	09:01-09:15	1944.6	1872.5	3.71%

Table 3 Summary of result data for case 1

No.	Real [Wh]	Forecasting [Wh]	Average Error	Max Error	Min Error
1	8318.15	8427.8	2.62%	5.43%	0.90%
2	8234.1	8416.7	4.71%	13.97%	0.11%
3	13019.9	12847.7	1.32%	2.82%	0.41%
4	9350.1	9235.91	3.05%	7.89%	0.27%
5	12019.1	12119	2.51%	6.47%	0.20%
6	3938.55	3920.88	1.03%	1.95%	0.51%
7	13485	13042.3	3.26%	4.79%	0.68%
Total	68364.9	68010.29	2.64%	13.97%	0.11%

As can be seen in Table 3, the average mean errors of missing blocks are around 2%. The maximum and minimum mean errors are approximately 13% and 0.1% in the second missing block, respectively.

Since it cannot determine whether all load data of the missing blocks are missed, evaluation of the corrected load data is required. Table 4 presents results from the branch and bound algorithm. It assumes that the 7th missing block is the maintenance schedule.

Table 4 Result from branch and bound

[Wh] \ No.	1	2	3	4	5	6	7
Real	8318.15	8234.1	13019.9	9350.1	12019.1	3938.55	0
Forecasting	8427.80	8416.7	12847.7	9235.91	12119.0	3920.88	0

As can be seen in Table 4, demand for the 7th missing block is zero. Therefore, it can see that the searched block from the branch and bound algorithm corresponds to the assumption.

3.2 Case 2: Commercial

The hourly data used to test the validity of the proposed correction technique are from May 1st to August 31st, 2002. In the measured load data, the missing data for each period is as follows:

Table 5 Missing data for each period

No.	Real Missing Period	Num. of Missing Data	Corrected Demand [Wh]
1	22:31-24:00, July 10th	6	2661.73
2	18:31-24:00, July 14th	22	9408.36
3	00:01-24:00, July 23rd	96	31284.08
4	23:01-24:00, July 27th	4	1803.47
5	16:01-24:00, July 28th	32	11865.60
6	19:31-24:00, July 31st	18	7694.64
7	18:31-24:00, Oct. 07th	22	9171.31
8	17:31-24:00, Oct. 08th	26	10031.55
9	22:46-24:00, Oct. 14th	5	2234.42
10	21:46-24:00, Oct. 15th	9	3984.11
11	09:46-24:00, Oct. 16th	57	15590.11
12	17:01-24:00, Oct. 20th	28	11140.62
13	22:01-24:00, Oct. 21st	8	3594.10
14	21:01-24:00, Oct. 22nd	12	5109.03
15	14:01-24:00, Oct. 23rd	40	12041.53
16	19:16-24:00, Oct. 24th	19	7869.45
17	23:16-24:00, Oct. 26th	3	1378.85
18	22:16-24:00, Oct. 27th	7	3184.55
19	21:15-08:15, Oct. 28-29th	44	4836.72
20	20:16-07:00, Oct. 29-30th	43	6391.68

Table 6 presents the results from the branch and bound algorithm. It assumes that the 3rd missing block is the maintenance schedule.

Table 6 Result from branch and bound

[Wh] \ No.	1	2	3	4
Forecasting	2661.73	9408.36	0	1803.47
[Wh] \ No.	5	6	7	8
Forecasting	11865.60	7694.64	9171.31	10031.55
[Wh] \ No.	9	10	11	12
Forecasting	2234.42	3984.11	15590.11	11140.62
[Wh] \ No.	13	14	15	16
Forecasting	3594.10	5109.03	12041.53	7869.45
[Wh] \ No.	17	18	19	20
Forecasting	1378.85	3184.55	4836.72	6391.68

3.3 Case 3: Peak Demand

The validity of the proposed correction technique is tested to use the hourly load data for four days in July 1999 (from July 20th to 24th), which is typically a peak demand week. Fig. 5 and Table 7 indicate the hourly measured load data and the missing load data, respectively.

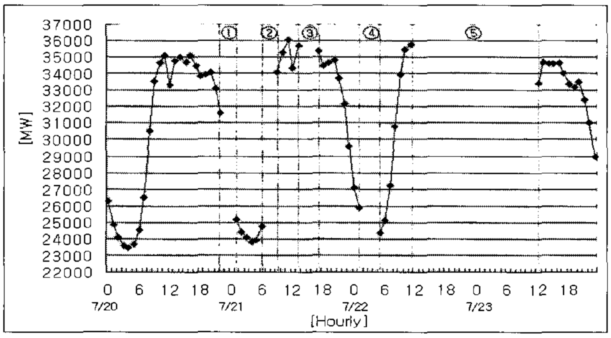


Fig. 6 Hourly load profile

Table 7 Missing load data

No.	Real Missing Period	Num. of Missing Data
1	22:00, July 20th -01:00, July 21st	2
2	06:00, July 21st -09:00, July 21st	2
3	13:00, July 21st -17:00, July 23rd	3
4	01:00, July 22nd -05:00, July 22nd	3
5	11:00, July 22nd -12:00, July 23rd	24

In Table 7, the divided difference (δ) of each block is as follows:

- 1st block: δ_{k-1} and δ_k have negative signs.
- 2nd block: δ_{k-1} and δ_k have positive signs.
- 3rd block: δ_{k-1} and δ_k have opposite signs.
- 4th block: either δ_{k-1} or δ_k is zero.
- 5th block: the missing block without the periodicity (24 hours).

Fig. 7 shows the corrected load curve for missing blocks (from 1st to 4th) using PCHIP algorithm.

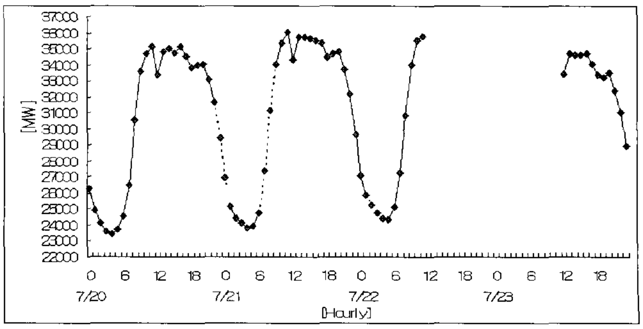


Fig. 7 Corrected load data using PCHIP algorithm

Fig. 8 presents the corrected load curve of the 5th missing block using the extended ARIMA model.

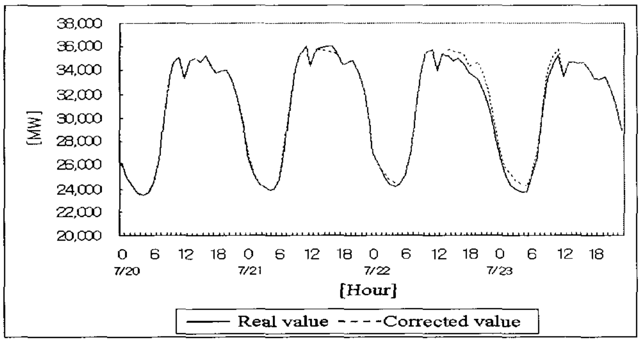


Fig 8 Hourly measuring load curve and forecasting result curve

Table 8 Hourly measured load data and corrected load data

No.	Time	Real [Mwh]	Forecating[MWh]	Error
1	23:00, July 20th	29536	29410	0.43%
	00:00, July 21st	26488	26925	1.65%
	07:00, July 21st	26866	27341	1.77%
	08:00, July 21st	30919	31168	0.81%
	14:00, July 21st	35894	35677	0.60%
	15:00, July 21st	36027	35621	1.13%
	16:00, July 21st	36028	35520	1.41%
2	02:00, July 22nd	24967	25266	1.20%
	03:00, July 22nd	24405	24771	1.50%
	04:00, July 22nd	24152	24442	1.20%
	12:00, July 22nd	33987	34127	0.41%
	13:00, July 22nd	35279	35427	0.42%
	14:00, July 22nd	35158	35709	1.57%
	15:00, July 22nd	34738	35602	2.49%
3	16:00, July 22nd	34958	35493	1.53%
	17:00, July 22nd	34512	35173	1.92%
	18:00, July 22nd	33831	34339	1.50%
	19:00, July 22nd	33474	34461	2.95%
	20:00, July 22nd	33219	34710	4.49%
	21:00, July 22nd	32076	33548	4.59%
	22:00, July 22nd	30875	32064	3.85%
4	23:00, July 22nd	28769	29520	2.61%
	00:00, July 23rd	26534	27036	1.89%
	01:00, July 23rd	25179	25776	2.37%
	02:00, July 23rd	24306	25192	3.65%
	03:00, July 23rd	23938	24680	3.10%
	04:00, July 23rd	23743	24379	2.68%
	05:00, July 23rd	23780	24245	1.96%
5	06:00, July 23rd	24844	25051	0.83%
	07:00, July 23rd	26628	27151	1.96%
	08:00, July 23rd	30420	30745	1.07%
	09:00, July 23rd	33219	33922	2.12%
	10:00, July 23rd	34600	35418	2.36%
	11:00, July 23rd	35178	35701	1.49%
	12:00, July 23rd	36028	35520	1.40%

Table 8 indicates the hourly results of the corrected load data for each block and Table 9 represents the summary of result data for case 3.

Table 9 Summary of result data for case 3

No.	Real [Wh]	Forecasting [Wh]	Average Error	Max Error	Min Error
1	221758	221662	1.11%	1.77%	0.43%
2	212686	215344	1.25%	2.49%	0.41%
3	232945	239788	2.97%	4.59%	1.50%
4	176249	180828	2.61%	3.65%	1.89%
5	184889	187988	1.64%	2.36%	0.83%
Total	1028527	1045610	1.92%	4.59%	0.41%

In Table 9, the maximum mean error is about 13% at the missing block 3, the minimum mean error is approximately 0.1% at the missing block 2, and the average mean error is around 1.9%.

These results present the validity of the proposed correction technique for missing load data.

5. Conclusion

This paper presents the efficient correction technique of missing load data based on the ARIMA model. The extended ARIMA model, PCHIP, and B&B algorithm is used to correct the missing load data.

The proposed correction technique is tested under a variety of conditions to establish its validity. Results from the case studies show the proposed correction technique could effectively exploit load forecasting.

Acknowledgements

This work was financially supported by MOCIE through the IERC program.

Reference

[1] Ministry of Commerce, Industry and Energy Korea Electricity Commission, *Restructuring of Electric Power Industry in Korea*, Oct. 2001.

[2] Amjady, N., "Short-term hourly load forecasting using time-series modeling with peak load estimation capability"", *IEEE Tran. on Power Systems*, Vol. 16, pp. 798-805, Nov. 2001.

[3] El Desouky, A.A., Elkateb, M.M., "Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA", *IEE Proceedings*, Vol. 147, Issue 4, pp. 213-217, July 2000.

[4] Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J., "ARIMA models to predict next-day electricity prices", *IEEE Trans. on Power Systems*, Vol. 18, Issue 3, pp. 1014-1020, Aug. 2003.

[5] M.Y.Cho, J.C.Hwang, C.S.Chen, "Customer Short Term Load Forecasting By Using ARIMA Transfer Function Model", *IEEE Proceedings*, Vol. 1, pp. 21-23, Nov. 1995.

[6] Amjady, N, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability", *IEEE Trans. on Power Systems*, Vol. 16, Issue 3, pp. 498-505, Aug. 2001.

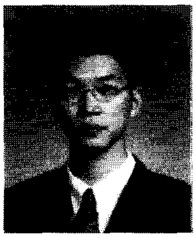
[7] G.E.P.Box, G.M.Jenkins, G.C.Reinsel, *Time Series Analysis Forecasting and Control*, Third ed.

[8] Jayu Academy, *Time Series Analysis*, 1994.

[9] Yulgok, *Time Series Analysis Using SAS/ETS*, 2002.

[10] SAS Korea, *Basic Statistics Analysis*.

[11] P.Venkataranman, *Applied Optimization with Matlab Programming*, John Wiley & Sons, pp. 318-335, 2002.



Chan-Joo Lee

He received the B.E. degree from Anyang University, South Korea, in 2000, and the M.S. degree from Konkuk University, South Korea, in 2002, respectively. He is currently working toward his Ph.D. degree in Electrical Engineering from Konkuk University.

Jong-Bae Park

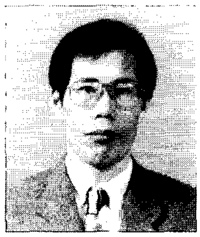
He received the B.E., M.S. and Ph.D. degrees in Electrical Engineering from Seoul National University, South Korea, in 1987, 1989, and 1998, respectively. From 1989-1988, he worked as a researcher of Korea Electric Power Corporation (KEPCO), and since 2001 he has been an assistant professor at Konkuk University, Korea. His research interests are power system planning and economic studies.

Jae-Yong Lee

He received the B.E. degree from Anyang University, South Korea, in 2002, and the M.S. degree from Konkuk University, South Korea, in 2004, respectively.

Joong-Rin Shin

He received the B.E., M.S., and Ph.D. in Electrical Engineering from Seoul National University, South Korea, in 1977, 1984, and 1989, respectively. Since 1990, he has been with the Department of Electrical Engineering, Konkuk University in Seoul, Korea, where currently is a professor. His major research field is power system analysis, power system planning and reliability evaluation.



Chang-Ho Lee

He received the B.E. from Chonnam National University, in 1981, and the M.S., Ph.D. in Economics from Pusan National University, South Korea, in 1987 and 1998, respectively. He is currently working at the Korea Electrotechnology Research Institute.