

## A Correction Technique of Missing Load Data Based on ARIMA Model

朴宗培\* · 李瓚柱\*\* · 李才勇\*\*\* · 愼重麟§ · 李昌浩§§

(Jong-Bae Park · Chan-Joo Lee · Jae-Yong Lee · Joong-Rin Shin · Chang-Ho Lee)

**Abstract** - Traditionally, electrical power systems had the vertically-integrated industry structures based on the economics of scale. However power systems have been recently reformed to increase the energy efficiency of the power system. According to these trends, Korean power industry has been partially restructured, and the competitive generation market was opened in 2001. In competitive electric markets, correct demand data are one of the most important issue to maintain the flexible electric markets as well as the reliable power systems. However, the measuring load data can have the uncertainty because of mechanical trouble, communication jamming, and other things. To obtain the reliable load data, an efficient evaluation technique to adjust the missing load data is needed.

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

**Key Words** : Time Series Analysis, ARIMA(Autoregressive Integrated Moving Average), Load Forecast, PCHIP (Piecewise Cubic Interporation), Branch and Bound.

### 1. 서 론

전력산업 구조개편이 시행됨에 따라 수직통합 체제를 이루고 있던 전력산업은 에너지효율성 재고적 측면에서 급격한 속도로 발전·송전·배전·판매 등 기능별로 수직 분할되고 있으며, 전력거래를 위한 도소매시장을 형성하고 있다[1,2]. 경쟁적 전력시장의 형성은 전반적인 전력산업구조의 변화를 가져다 줄 뿐만 아니라 최종 소비자의 에너지 소비형태에도 많은 변화를 줄 것이다. 이러한 전력산업의 구조개편과 전력시장의 형성으로 인하여 최종소비자들은 전력소비형태에 대한 다양한 정보와 관련 분석자료를 필요로 하지만 현재로서는 이러한 자료가 제공되지 못하고 있는 실정이다. 또한 개별 수용가에 대한 에너지사용 데이터 취득에 있어서 통신상의 장애등의 원인으로 인한 불확실성의 내포는 수요관리(DSM: Demand Side Managment)의 신뢰성을 저하시키는 원인으로 작용하고 있다. 따라서 전력에너지의 합리적인 사용을 유도하기 위하여 소비자의 전력사용형태에 대한 세밀한 분석이 요구되어지고 각종 연구의 기초자료 및 정책 자료의

활용에 있어 소비형태에 대한 체계적이고 합리적인 분석이 필요하다. 하지만, 현재 우리나라의 수요 검침자료에는 통신상 장애등의 원인으로 인한 불확실성을 내포하고 있기 때문에 정확한 전력에너지 사용형태에 대한 분석을 부분적으로 수행하지 못하고 있으며 이를 위한 효율적인 전력수요관리를 위한 신뢰성 있는 과거 실적 부하자료의 생성이 무엇보다도 중요하다.

본 논문에서는 산업종별·계약종별·계약전력별에 대한 수요실적자료를 바탕으로 다양하고 신뢰성있는 과거 부하실적곡선에 대한 분석을 수행하였고, 검침자료에서 누락된 데이터에 대한 보정 알고리즘 및 과거자료를 기반으로 하여 단기 간 에너지 사용패턴을 추정하는데 필요한 적합한 알고리즘을 제시하였다. 본 논문의 알고리즘은 개별 수용가의 과거 실적 부하자료에 대해서 수치해석적 기법인 3차구간보간법과 통계학적 자료 예측기법인 자기회귀누적이동평균모형(ARIMA: Autoregressive Integrated Moving Average) 모형을 확장하여 예측하였으며, 이를 통해서 얻은 예측된 개별 수용가의 에너지사용량에 대해서 최적화 기법인 분기한정법을 적용하여 전체 에너지사용량을 보정하는 기법을 제시하였다. 또한 실제 누락된 수요실적자료를 가지고 개별 수용가에 대한 사례 연구를 수행함으로써 제시한 방법론의 타당성을 입증하였다. 본 논문에서 제시하는 방법론은 부하별로 다양하고 신뢰성 있는 과거 실적부하곡선에 대한 분석에 활용될 수 있으며, 안정적인 전력시장의 운용 및 수요관리 정책에 적용될 수 있을 것으로 판단된다.

\* 正會員 : 建國大學 電氣工學科 助敎授 · 工博

\*\* 正會員 : 建國大學 電氣工學科 博士課程

\*\*\* 正會員 : 建國大學 電氣工學科 碩士

§ 正會員 : 建國大學 電氣工學科 敎授 · 工博

§§ 正會員 : 韓國電氣研究員 電力産業研究室 室長

接受日字 : 2004年 3月 26日

最終完了 : 2004年 4月 29日

2. 수요실적자료 보정

2.1 일반적인 ARIMA 모형

시계열분석은 시간에 종속적인 시계열의 측정값 간의 상호 관계를 밝히고 이를 바탕으로 미래 시계열을 합리적으로 예측하는데 그 목적이 있으며, 전력계통분야에서 시계열분석을 이용한 장기간, 단기간 수요예측, 시장 가격 예측 등 다수의 연구사례가 이미 보고된 바 있다[3,4,5,6,7]. 전력계통에서 측정되는 수요데이터는 평균, 분산 및 자기공분산이 시간의 변화에 영향을 받는 비정상시계열 자료이다. Box-Jenkins는 비정상시계열 자료를 연속적인 차분과 확률과정을 통해 정상시계열 자료로 변환하는 과정을 일반화 하였고, 이에 대한 모형의 개념은 아래의 그림과 같은 절차를 따른다[8,9].

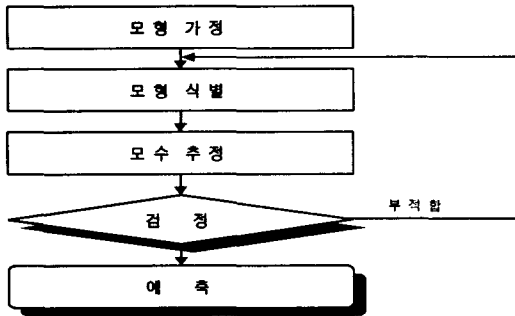


그림 1 박스-젠킨스의 시계열분석 절차  
Fig. 1 Time series analysis procedure by Box-Jenkins

그림 1에서 모형의 식별은 자기회귀모형(AR: Auto Regressive Model)과 이동평균모형(MA: Moving Average Model) 부분의 차수를 결정하는 것이고, 모수추정은 자기회귀부분과 이동평균부분이 가지는 계수들의 최우수 추정량을 구하는 것이며 모형검정은 식별과 추정에 의해 선택된 모형이 주어진 자료에 대해 적합성을 판단하는 것이다. 이러한 시계열분석 절차에 따라 과거 실적 부하자료를 이용하여 개별 수용가에 대한 수요함수의 모형을 설정할 수 있고, 설정된 모형을 이용하여 누락된 수요실적자료의 보정이 가능하다. 자기회귀누적 이동평균 모형은 자기회귀모형과 이동평균모형의 결합으로 비정상성을 제거하기 위해 적절한 차분을 적용한 모형을 말한다. 박스-젠킨스의 모형은 자기회귀모형, 이동평균모형, 자기회귀이동평균모형으로 나누어지고 시계열자료( $y_t$ )에 대한 일반적인 자기회귀 모형은 아래 식 (1)과 같다[9].

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

여기서,  $\phi_i$  : 자기회귀계수.  
 $p$  : 자기회귀모형의 차수.  
 $\epsilon_t$  : 평균이 0이고 분산이 일정한 백색오차.

또한 이동평균모형이란 현재 시계열자료( $y_t$ )가 연속적인 백색오차항들의 영향을 받는다는 것으로서, 백색잡음오차의 누적평균으로 표현될 수 있다. 일반적인 이동평균 모형은 아

래 식 (2)와 같다.

$$y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (2)$$

여기서,  $\theta_i$  : 이동평균계수.  
 $q$  : 이동평균모형의 차수.

위의 두 가지 모형은 시계열자료( $y_t$ )가 과거의 시계열자료 값들로 표현될 수 있다는 것과 과거의 오차값들로 표현될 수 있다는 것을 나타내고 있다. 하지만 경우에 따라서는 시계열 자료( $y_t$ )는 과거의 시계열자료 값들과 과거의 오차값들 모두의 영향을 받는다고 할 수 있다. 이러한 모형은 자기회귀모형과 이동평균 모형의 혼합된 모형인 자기회귀이동평균 모형으로 나타낼 수 있고 자기회귀이동평균 모형은 아래 식 (3)과 같이 된다.

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

박스-젠킨스는 이론적으로 도출된 자기회귀모형, 이동평균모형 및 자기회귀이동평균모형의 자기상관계수와 부분 자기상관계수가 차수  $p$  또는  $q$ 의 값에 따라 독특한 형태를 가짐을 이용하여 시계열자료의 적합한 모형과 차수를 선정하였다. 그리고 자기회귀이동평균 모형은 상기 식 (3)에 후향연산자 함수를 이용하여 아래의 식 (4)와 같이 정식화 할 수 있다.

$$\begin{aligned} (1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p) y_t &= (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q) \epsilon_t \\ \rightarrow (1 - \sum_{i=1}^p \phi_i B^i) y_t &= (1 - \sum_{j=1}^q \theta_j B^j) \epsilon_t \\ \rightarrow \phi_p(B) y_t &= \theta_q(B) \epsilon_t \end{aligned} \quad (4)$$

여기서,  $y_t$  : 시간  $t$  에서 개별수용가의 수요실적자료.  
 $B$  : 후향 연산자(backward-shift operator);  
 $B^m y_t = y_{t-m}$

식 (4)는 모형차수  $p$ 를 갖는 자기회귀모형과 모형차수  $q$ 를 갖는 이동평균모형을 혼합한 자기회귀이동평균 모형(ARMA(p,q))을 후향연산자(Back-shift operator)를 사용하여 간략화를 나타내는 것이다. 이 모형은 시계열( $y_t$ )이 과거의 시계열 관측값 및 확률오차들에 의해 영향을 받는다는 것을 의미한다.

2-2. 확장된 ARIMA 모형

일반적으로 개별 수용가의 수요는 시간대별, 요일별, 계절별에 따라 수요함수( $y_t$ )에 대해 각각 서로 다른 영향을 받는다. 때문에 수요실적함수는 각각의 인자에 따라서 서로 의존적인 관계로 일반화가 될 수 있으며, 이는 예측모형의 차수 설정에 중요한 요소로 적용된다. 따라서 본 논문에서는 각각의 인자에 대한 수요실적 예측모형을 식 (4)를 이용하여 아래의

식 (5)와 같이 확장하였다.

$$\phi_p^h(B) \phi_p^d(B) \phi_p^w(B) y_t = \theta_q^h(B) \theta_q^d(B) \theta_q^w(B) \varepsilon_t \quad (5)$$

여기서,  $\phi_p^h(B)$  : 차수 p를 갖는 시간대별 자기회귀모형 (Hourly autoregressive model)

$\phi_p^d(B)$  : 차수 p를 갖는 요일별 자기회귀모형 (Daily autoregressive model)

$\phi_p^w(B)$  : 차수 p를 갖는 주간별 자기회귀모형 (Weekly autoregressive model)

$\theta_q^h(B)$  : 차수 q를 갖는 시간대별 이동평균모형 (Hourly moving average model)

$\theta_q^d(B)$  : 차수 q를 갖는 요일별 이동평균모형 (Daily moving average model)

$\theta_q^w(B)$  : 차수 q를 갖는 주간별 이동평균모형 (Weekly moving average model)

상기 식 (5)에서 수요함수( $y_t$ )는 시간대별 자기회귀모형과 요일별 자기회귀모형, 그리고 주간별 자기회귀모형의 곱의 형태를 이루고, 이것은 이동평균모형에 대해서도 동일하게 적용하여 확장하였다. 하지만 대부분의 수요실적자료는 시계열 자료의 변화 패턴이 일정한 평균값을 중심으로 일정한 변동폭을 갖지 않는 비정상성을 가지기 때문에 적절한 차분을 통해 정상성 시계열로 변환해야 한다. 따라서 상기 식(5)에 연속적인 차분을 적용하여 추세요인을 제거하면 정상 시계열로 분석이 가능하고, 이것은 아래의 식 (6)과 같다.

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

여기서,  $d$  : 차분차수

식 (6)은 비정상성 추세요인을 제거한 수요실적자료의 자기회귀누적이동평균모형(ARIMA)을 나타낸 것이다. 하지만 박스-젠킨스가 제시한 모형식별과정의 자기회귀누적이동평균모형의 차수(ARIMA(p,d,q)) 결정[8,9]은 일반적으로 모형식별자의 경험적 판단에 의존하기 때문에 본 논문에서는 누락된 수요실적자료를 보정하기 위한 수요실적예측 모형의 차수를 실험적인 방법으로 도출하여 식 (7)과 같이 나타내었다.

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

식 (7)에서  $(1 - \phi_1^h B^1 - \phi_2^h B^2 - \phi_3^h B^3 - \phi_4^h B^4 - \phi_5^h B^5)$ 는 시간대별 자기회귀모형으로서  $t$ 시간부터  $t-5$ 시간까지의 시계열자료를 나타낸 것이고,  $(1 - \phi_{24}^d B^{24} - \phi_{48}^d B^{48} - \phi_{72}^d B^{72} - \phi_{96}^d B^{96} - \phi_{120}^d B^{120})$ 는  $t$ 시간에 해당하는 5일전 자기회귀모형차수를 나타낸 것이다. 그리고  $(1 - \phi_{168}^w B^{168} - \phi_{336}^w B^{336} - \phi_{504}^w B^{504} - \phi_{672}^w B^{672})$ 는  $t$ 시간에 해당하는 4주전 자기회귀모형차수를 나타낸 것으로

예측에 필요한 자기회귀모형의 최소 차수를 나타내었다. 그리고 수요자료의 비정상성은 차분차수( $d$ )는 정밀성을 위해 1차항으로 설정하여 정상성 시계열자료로 변환하였고, 이동평균모형의 차수는 각각의 항에서  $t-1$ ,  $t-2$ 까지 결측된 수요자료 보정에 필요한 최소 자료구간임을 나타내었다.

### 2-3. 3차 구간 보간법(Piecewise Cubic Interpolation)

ARIMA 모형을 이용하여 시계열을 분석하기 위해서는 적절한 개수의 연속적인 시계열 자료의 확보가 무엇보다도 중요하며, 확보된 시계열자료를 바탕으로 그 시계열의 주기성 판단 및 모형식별을 할 수 있다[8,9]. 하지만 전력계통의 수요자료는 여러 가지 요인으로 인해 결측구간이 발생하고, 이로 인해 취득한 수요자료의 연속성을 보장하기가 어렵다. 이것은 결측구간이 있는 수요자료의 보정에 대한 ARIMA 모형을 적용한 예측한 결과의 신뢰성을 떨어뜨리는 요인이 된다. 따라서 본 논문에서는 이러한 ARIMA 모형의 단점을 보완하기 위해서 수치해석적 접근 방법인 3차구간보간법을 이용하였고, 아래의 식 (8)과 같이 나타내었다.

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

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

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

여기서,  $h_k$  : k-번째 부분구간의 길이,

$d_k$  : k-번째 부분구간에서의 기울기, ( $\delta_k \neq d_k$ )

$\delta_k$  : k-번째 부분구간의 평균 기울기.

상기 식 (8)은  $x_k \leq x \leq x_{k+1}$ 에서 아래와 같은 4가지 조건을 만족한다.

$$\begin{aligned} \text{i) } P(x_k) &= y_k & \text{ii) } P(x_{k+1}) &= y_{k+1} \\ \text{iii) } P'(x_k) &= d_k & \text{iv) } P'(x_{k+1}) &= d_{k+1} \end{aligned}$$

3차 구간보간법을 적용한 3가지 경우의  $d_k$ 에 관한 정의는 아래와 같다.

- $\delta_{k-1}$  과  $\delta_k$ 의 부호가 다른 경우
- $\delta_{k-1}$  과  $\delta_k$  중 하나가 0인 경우
- $\delta_{k-1}$  과  $\delta_k$ 의 부호가 같은 경우

첫 번째 경우와 두 번째 경우  $d_k$ 의 값은 0이고, 세 번째 경우  $d_k$ 의 값은 아래 식 (9)와 같다.

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

만약 두 구간의 길이가 다르다면 상기 식 (9)는 아래의 식 (10)과 같이 변형된다.

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

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

본 논문에서는 3차 구간보간법과 확장된 ARIMA 모형을 이용하여 수요검침자료에서 누락된 수요자료를 보정하는 알고리즘을 제시하였다. 전력 수요자료는 각각의 부하특성, 계절특성 등에 따라 약간의 차이는 있지만 개별 부하의 일부하 수요자료를 시간에 따라 나열하면 일정한 주기특성을 갖는 것을 알 수 있다. 따라서 단기간 시간대별 수요예측을 수행하기 위해서는 주기적 특성이 반영된 ARIMA 모형이 적용되어야 하며, 본 논문에서는 누락된 자료로 인해서 주기성이 결여된 수요자료에 대해서 수치해석적 기법인 3차 구간반복법을 이용하여 보정하였다.

2-4. 분기 한정법 (Branch & Bound Algorithm)

개별 수용가에 측정된 수요의 결측자료는 수요자료만을 가지고 그 원인을 분석하기가 어렵다. 즉, 누락된 수요자료가 통신상의 장애로부터 발생했는지, 기기의 오동작으로부터 발생했는지, 또는 부하의 유지보수로 인한 실제 에너지사용량의 값이 없는지 등등에 대해서 원인판단을 해석하기가 어렵다. 또한 보정된 수요자료의 수용가의 전체 에너지 사용량은 실제 혹은 추정된 에너지 사용량 보다 항상 크거나 같을 것이다. 만약 에너지사용량이 같은 경우에는 결측자료에 대한 보정이 적합하다고 판단할 수 있지만, 서로 다른 경우에는 보정된 수요자료의 신뢰성이 떨어진다. 때문에 실제 혹은 추정된 수용가의 전체 에너지 사용량에 대해서 보정된 전체 에너지 사용량의 해석이 필요하며, 본 논문에서는 이를 위해 최적화 기법인 분기 한정법을 누락구간의 적합성 판단결정문제에 적용하였다.

최적화 문제를 다루는 방법은 크게 이산적인 문제를 해석하는 방법과 연속적인 문제를 해석하는 방법이 있다. 이산적인 최적화 문제를 해석하는 방법은 연속적인 최적화 문제를 해석하는 방법보다 계산하는데 시간이 많이 소요될 뿐 만 아니라 접근 방법에 있어서 어려운 점이 많다[9,10]. 그래서 대부분의 이산적인 최적화 문제는 수학적으로 연속적인 모델링에 의해서 접근된다. 이산적인 최적화 기법에는 이산변수의 모든 조합을 통해서 최적화 해법을 찾는 기법과 지역 탐색을 이용하여 해법을 얻는 분기 한정법, 그리고 동적 프로그래밍 기법 등이 있다. 본 논문에서는 누락된 수요실적자료를 보정하기 위해 지역 탐색을 통해 해의 도출이 보장되는 분기 한정법을 적용하였다. 본 논문에서 적용한 분기 한정법은 누락된 수요실적자료에 대해 ARIMA 모형을 이용하여 개별 수용가의 수요실적을 보정한 후, 실제 측정된 혹은 추정된 개별 수용가의 연간 수요실적과 보정된 수요실적을 최적화하는데 사용된다. 본 논문에서 적용한 분기 한정법의 순서도는 아래의 그림 2와 같다.

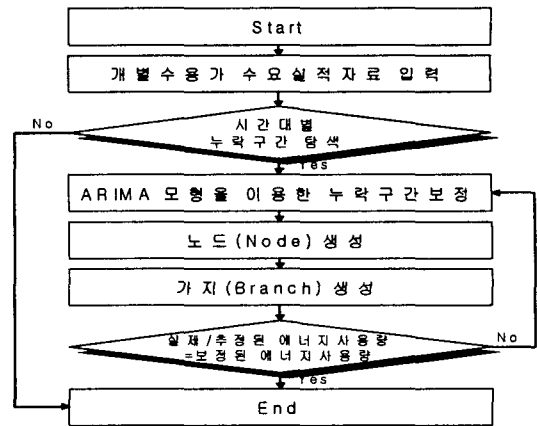


그림 2 분기 한정법을 이용한 수요실적자료 보정 알고리즘 순서도  
Fig. 2 Flowchart of a corrective technique for missing load data using BB method

그림 2에서 개별 수용가의 수요실적 자료가 입력되면 시간대별로 누락구간의 존재의 유무를 판단해야 한다. 만약 누락구간이 존재한다면 ARIMA 모형을 이용하여 누락된 구간의 수요자료를 보정한다. 여기에서 보정된 수용가의 에너지사용량은 모든 누락구간에 대해서 수요자료를 보정하기 때문에 실제 혹은 추정된 개별수용가의 에너지사용량보다 항상 크거나 같다. 따라서 ARIMA 모형을 이용하여 보정된 수용가의 에너지사용량은 실제 혹은 추정된 수용가의 에너지사용량과 전체 에너지사용량을 비교하여 누락구간의 적합성을 판단해야 한다. 이러한 적합성을 판단하기 위해서 적용된 분기 한정법은 부분해를 탐색하기 위해 기본적으로 트리(tree)구조를 가지며, 이러한 트리는 노드와 가지로 구성된다. 이때  $n$ 개의 누락구간별로 보정된 에너지사용량은 각각의  $n$ -노드로 생성이 되며, 각각의 노드에서 생성된 부분해의 집합인 스테이지(stage)는 스테이지 내에 최적 부분해를 기준으로 노드를 생성하며, 각 스테이지에서 생성된 노드와 노드 간에는 가지로 연결된다. 이때 노드간의 연결되는 가지는 각 누락구간의 보정된 에너지사용량으로 결정되고, 본 논문에서 제시하는 식 (11)의 목적함수에 따라서 개별 수용가의 전체 에너지사용량이 보정된다. 본 논문에서 제시한 분기 한정법의 목적함수는 아래와 같다.

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

여기서,  $ME_i$  : 개별 수용가의 측정 또는 추정된 전력 사용량[MWh].

$FCE_i$  : 개별 수용가의 예측된 전력사용량 [MWh].

$I$  : 기간.

만약, 개별 수용가의 누락구간이  $n$  개이고, 각 구간에서 예측된 에너지사용량을  $X^* = [x_1, x_2, \dots, x_n]$ 이라고 정의한다면, 각 노드와 가지에 대한 분기 한정법은 아래 그림 3과 같이 적용된다.

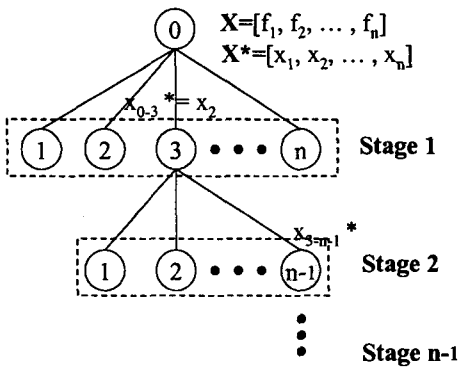


그림 3 분기한정법 알고리즘  
Fig. 3 Branch and bound algorithm

그림 3에서 알 수 있듯이 분기 한정법은 각 단계(stage)와 노드(node), 그리고 가지(branch)로 이루어진다. 여기서,  $X$ 는 자유변수이며,  $X^*$ 는 각 누락구간에서의 보정된 에너지사용량을 나타낸다. 만약 Stage 1에서 모든 노드에 대해서 목적함수에 적합한 에너지사용량을 탐색하여 얻은 결과가 3-번 노드라면, 최적노드인 3-번 노드에서 다시 가지를 펼친다. 이렇게 전체  $n-1$ 번의 단계를 반복하면 이산적인 변수에 대해서 최적값을 얻을 수 있다.

본 논문에서 적용한 분기 한정법은 보정된 개별 수용가의 에너지사용량에 대한 판별에 적용되며, 이때 부하의 유지보수로 인한 수요자료의 누락구간은 복수개로 가정하고 모의 시험하였다. 이는 부하의 유지보수로 인한 누락구간이 없거나 하나인 경우 분기 한정법의 적용이 불필요하기 때문이다.

### 3. 사례 연구

#### 3-1 사례 1: 일반용 수용가

본 논문에서는 에너지관리공단에 의해서 검침된 실측 전력수용가데이터 중 특정 일반용 수용가의 하계 첨두부하 실측자료(2002년7월29일~2002년8월02일)를 이용하여 제시한 알고리즘의 타당성을 검증하였다. 그리고 본 논문에서 제시한 방법론의 ARIMA 모형 차수 설정 및 예측치는 통계학적 범용 소프트웨어인 SAS Ver. 8.1을 사용하여 모의시험 하였다 [11,12]. 본 논문에서 제시한 알고리즘의 신뢰성을 검증하기 위한 임의의 자료 결측구간과 설비유지보수 기간은 아래와 같이 가정하였다.

표 1 수요자료 누락구간

Table 1 Assumption for missing load data

구간	구간	구간	구간
1	7월29일07시31분-7월29일09시15분	8318.15	결측구간
2	7월29일17시01분-7월29일18시45분	8234.10	결측구간
3	7월30일21시31분-7월30일24시15분	13020.0	결측구간
4	7월31일08시16분-7월31일10시00분	9350.25	결측구간
5	7월31일16시01분-7월31일17시45분	12019.3	결측구간
6	8월01일03시16분-8월01일05시00분	3938.55	결측구간
7	8월01일12시46분-8월01일14시30분	13485.1	유지보수기간

본 사례연구에서 적용된 일반용 수용가의 하계첨두기간의 일주간 수요 가운데 누락구간은 아래의 그림 4와 같다.

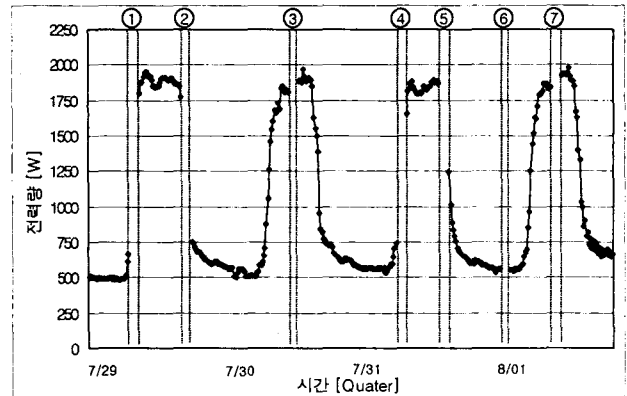


그림 4 누락된 수요데이터 프로파일  
Fig. 4 Load profile of a customer with missing data

상기 자료를 바탕으로 아래 그림 5는 본 논문에서 제시한 ARIMA 모형과 3차구간보간법을 이용하여 얻은 보정값을 나타낸 것이다.

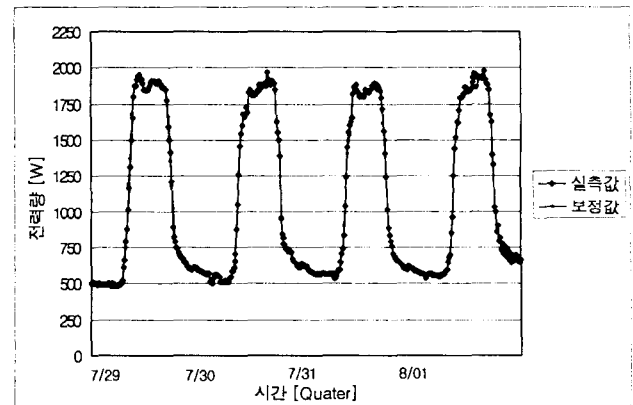


그림 5 실측 수요 데이터 값과 보정 데이터 값  
Fig. 5 A result data using the collection technique for missing load data

상기 그림에서 실측값은 일반용 수용가에서 실측된 자료를 나타낸 것이고, 보정값은 본 논문에서 제시한 알고리즘의 타당성을 검증하기 위해 각 구간별로 임의로 결측시킨 구간에 대한 보정 결과값을 나타낸 것이다. 그림 4에서 구간 1의 경우 시계열 분석에 필요한 수요자료가 부족하기 때문에 3차구간보간법을 사용하여 결측자료를 보정하였고, 나머지 구간의 경우 확보된 시계열 자료를 바탕으로 확장된 ARIMA 모형을 이용하여 보정값을 산출하였다. 개별 구간에 대한 실측값과 보정값은 아래 표 2와 같고 결과에 대한 요약은 표 3과 같다.

표 2 각 구간별 실측값, 보정값 및 오차율

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

791.7	882	1011.7	1170.6	1310.7	1499.8	1661.65
748.7	874.1	1026.9	1195.6	1368.5	1533.9	1680.1
5.43	0.89	1.50	2.13	4.40	2.27	1.72
1589.7	1503.6	1413.3	1207.5	898.8	829.5	791.7
1660.6	1515.2	1351.3	1183	1024.4	889.6	792.6
4.45	0.77	4.38	2.20	13.97	7.24	0.11
1815.4	1829.1	1850.1	1885.8	1883.7	1874.2	1881.6
1808	1813.6	1822	1832.6	1844.5	1857.2	1869.8
0.41	0.84	1.51	2.82	2.08	0.90	0.62
837.9	1041.6	1241.1	1452.1	1552.9	1598.1	1626.4
904.01	1050.1	1244.4	1407.7	1487.3	1535.2	1607.2
7.89	0.82	0.27	3.06	4.22	3.93	1.18
1850.1	1855.3	1838.5	1792.3	1709.4	1563.4	1410.1
1873.3	1873.4	1859.6	1832.3	1618.7	1560.3	1501.4
1.25	0.97	1.14	2.22	5.30	0.19	6.47
570.15	569.1	567	555.45	564.9	554.4	557.55
565.48	564.77	557.78	560.80	553.88	557.80	560.37
0.81	0.75	1.62	0.96	1.95	0.61	0.50
1855.3	1897.3	1938.3	1957.2	1937.2	1955.1	1944.6
1842.7	1876.2	1867.8	1863.4	1857.7	1862.0	1872.5
0.67	1.11	3.63	4.79	4.10	4.75	3.70

표 3 각 구간별 결과 요약(사례연구 1)

Table 3 Summary of result data for case 1

1구간	8,318.1	8,427.8	2.624	5.431	0.895
2구간	8,234.1	8,416.7	4.711	13.974	0.113
3구간	13,019.9	12,847.7	1.315	2.821	0.407
4구간	9,350.1	9,235.9	3.052	7.889	0.265
5구간	12,019.1	12,119.0	2.512	6.474	0.198
6구간	3,938.5	3,920.8	1.034	1.950	0.505
7구간	13,485.0	13,042.3	3.256	4.792	0.679

상기 표에서 개별 구간에 대한 실측값과 보정값의 오차율은 평균 2.643%로 비교적 적은 오차율을 나타내고 있으며,

구간 2에서 13.974[%]의 최대 오차율과 0.113[%]의 최소오차율을 나타내었다. 이처럼 본 논문에서 제시하는 3차 구간보간법과 ARIMA 모델을 이용하여 결측값 보정의 타당성을 검증하였다. 하지만 누락된 구간의 모든 수요자료가 결측되었다고 볼 수 없기 때문에 수용가의 전체 에너지사용량에 대해서 검토해 볼 필요가 있다. 그래서 본 논문에서는 구간 7을 유지보수기간이라 가정하였고, 이때 각 수용가는 경제적 이윤을 극대화하기 위해 유지보수는 가장 적은 횟수를 실시한다고 가정하여 분기한정법을 이용하여 유지보수기간을 탐색하였다. 아래 표 4는 본 논문에서 적용한 분기한정법을 이용하여 산출된 결과 값을 나타내고 있다.

표 4 분기 한정법 적용결과

Table 4 Result from branch and bound algorithm

구간	1구간	2구간	3구간	4구간	5구간	6구간	7구간
누락된 구간별 실제전력량 [Wh]	7539.0	8234.1	13020.0	9350.25	12019.35	3938.55	0
분기별 선택된 전력량 [Wh]	8427.8	8416.7	12847.7	9236.17	12119.19	3920.88	0

상기 표에서 보면 알 수 있듯이 본 논문에서 제시한 알고리즘의 신뢰성을 검증하기 위해서 가정한 유지보수기간인 구간 7을 탐색하는 것을 볼 수 있다.

### 3-2 사례 2: 산업용 수용가

산업용 수용가의 경우 생산방식, 작업스케줄등에 따라 각각의 산업용 수용가의 부하패턴이 매우 상이하다. 더욱이 수용가의 시계열자료가 자기회귀모형과 이동평균모형을 갖지 않기 때문에 시계열분석에서 선행되는 자료의 모형식별이 주관적 판단에 의존할 뿐 만 아니라 식별과 추정에 의해 선택된 모형에 대해 적합성을 판단하는 어렵다. 따라서 여기에서는 시계열자료의 개수를 확장된 실제 수요검침자료가 누락된 특정 산업용 수용가의 2사분기(2002년5월 - 2002년 8월) 수요자료에 대해서 사례연구를 수행하여 본 논문에서 제시한 알고리즘의 타당성을 검증하였다. 아래 표 5는 산업용 전력을 사용하는 수용가의 2사분기 동안의 실제 누락구간의 정보와 본 논문에서 제시한 알고리즘을 이용하여 보정된 전력량을 나타내고 있다.

표 5 각 구간별 누락된 수요데이터 자료

Table 5 Missing data for each period

1	7월10일22시31분-7월10일24시00분	$x_{0710}^1, x_{0710}^2, \dots, x_{0710}^6$	2661.730
2	7월14일18시31분-7월14일24시00분	$x_{0714}^1, x_{0714}^2, \dots, x_{0714}^{22}$	9408.368
3	7월23일00시01분-7월23일24시00분	$x_{0723}^1, x_{0723}^2, \dots, x_{0723}^{96}$	31284.08
4	7월27일23시01분-7월27일24시00분	$x_{0727}^1, x_{0727}^2, \dots, x_{0727}^4$	1803.473
5	7월28일16시01분-7월28일24시00분	$x_{0728}^1, x_{0728}^2, \dots, x_{0728}^{32}$	11865.60
6	7월31일19시31분-7월31일24시00분	$x_{0806}^1, x_{0806}^2, \dots, x_{0806}^{18}$	7694.640

7	8월07일18시31분-8월07일24시00분	$x_{0807}^1, x_{0807}^2, \dots, x_{0807}^{22}$	9171.314
8	8월08일17시31분-8월08일24시00분	$x_{0808}^1, x_{0808}^2, \dots, x_{0808}^{26}$	10031.55
9	8월14일22시46분-8월14일24시00분	$x_{0814}^1, x_{0814}^2, \dots, x_{0814}^5$	2234.426
10	8월15일21시46분-8월15일24시00분	$x_{0815}^1, x_{0815}^2, \dots, x_{0815}^9$	3984.114
11	8월16일09시46분-8월16일24시00분	$x_{0816}^1, x_{0816}^2, \dots, x_{0816}^{57}$	15590.11
12	8월20일17시01분-8월20일24시00분	$x_{0820}^1, x_{0820}^2, \dots, x_{0820}^{28}$	11140.62
13	8월21일22시01분-8월21일24시00분	$x_{0821}^1, x_{0821}^2, \dots, x_{0821}^8$	3594.102
14	8월22일21시01분-8월22일24시00분	$x_{0822}^1, x_{0822}^2, \dots, x_{0822}^{12}$	5109.030
15	8월23일14시01분-8월23일24시00분	$x_{0823}^1, x_{0823}^2, \dots, x_{0823}^{40}$	12401.53
16	8월24일19시16분-8월24일24시00분	$x_{0824}^1, x_{0824}^2, \dots, x_{0824}^{19}$	7869.459
17	8월26일23시16분-8월26일24시00분	$x_{0826}^1, x_{0826}^2, x_{0826}^3$	1378.855
18	8월27일22시16분-8월27일24시00분	$x_{0827}^1, x_{0827}^2, \dots, x_{0827}^7$	3184.550
19	8월28일21시15분-8월29일08시15분	$x_{0828}^1, x_{0828}^2, \dots, x_{0829}^{44}$	4836.726
20	8월29일20시16분-8월30일07시00분	$x_{0829}^1, x_{0829}^2, \dots, x_{0830}^{43}$	6391.680

여기서,  $x$ 는 해당 누락구간에서의 시간대별 누락된 수요실적자료

여기서 주어진 수용가의 수요자료를 가지고 누락구간에 대한 분석이 어렵다. 따라서 본 논문에서는 구간 3을 유지보수 기간이라 가정하였고, 본 논문에서 제시한 분기한정법을 이용하여 탐색된 결과를 아래의 표 6과 같이 나타내었다.

표 6 사례 2의 분기 한정법 적용결과  
Table 6 Result data using branch and bound

					5 구간
	2661.73	9408.368	0	1803.473	11865.6
					10 구간
	7694.64	9171.314	10031.55	2234.426	3984.114
					15 구간
	15590.11	11140.62	3594.102	5109.03	12401.53
					20 구간
	7869.459	1378.855	3184.55	4836.726	6391.68

3-3 사례 3: 하절기 첩두수요

본 장에서는 에너지관리공단의 10년간 실측 전력수요데이터 가운데 전체 계통수요의 하절기 주중 첩두 기간(1999년7월20일-1999년7월24일)을 이용하여 제시한 알고리즘의 타당성을 검증하였다. 본 논문에서 적용한 하절기 시간대별 실측 수요데이터는 아래의 그림 5와 같고 각각의 누락구간에 대한 설명은 표 7과 같다.

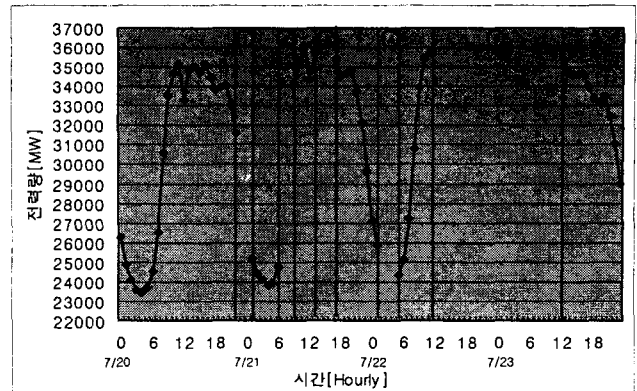


그림 5 시간대별 실측 수요자료  
Fig. 5 Hourly measuring load data

표 7 누락된 수요실적자료 구간  
Table 7 Missing load data

구간	누락구간	누락 자료수
1	7월20일22시-7월21일01시	$x_1^1, x_2^1$
2	7월21일06시-7월21일09시	$x_1^2, x_2^2$
3	7월21일13시-7월21일17시	$x_1^3, x_2^3, x_3^3$
4	7월22일01시-7월22일05시	$x_1^4, x_2^4, x_3^4$
5	7월22일11시-7월23일12시	$x_1^5, x_2^5, \dots, x_{23}^5, x_{24}^5$

여기서,  $x$ 는 해당 누락구간에서의 시간대별 누락된 수요실적자료

표 7에서 누락된 수요실적자료의 구간 1은 평균기울기( $\delta$ )의 부호가 모두 음(-)인 경우이며, 구간 2는 평균기울기( $\delta$ )의 부호가 모두 양(+ )인 경우이고, 구간 3은 평균기울기( $\delta$ )의 부호가 서로 다른 경우이다. 그리고 구간 4는 평균기울기( $\delta$ )가 0인 경우이며, 주기성이 결여된 누락구간은 구간 5에 나타내었다. 이것은 시간대별 수요실적자료에서 네 구간(1, 2, 3, 4)의 수요실적자료가 누락되어 시계열자료의 연속성이 보장되어야 하는 ARIMA 모형의 적용이 부적합하다. 따라서 수요실적자료가 누락된 네 구간에 대해서 3차구간 보간법을 이용하여 시간대별 수요실적값을 보정하였고, 적용된 보정값들은 아래의 그림 6과 같다.

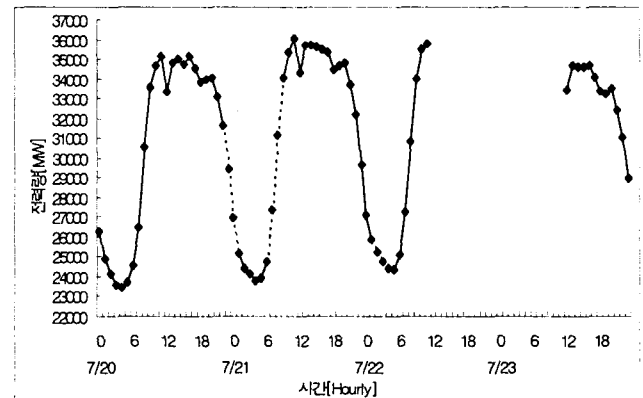


그림 6 3차구간법을 이용하여 보정된 수요자료  
Fig. 6 Corrected load data using piecewise cubic interpolation algorithm

상기 그림6에서 실선은 주어진 수요실적자료이고, 점선은 네 개의 누락구간의 시간대별 누락된 수요실적자료를 3차 구간보간법을 이용하여 구한 수요자료 보정값을 나타낸다. 3차 구간보간법을 이용하여 수요자료의 보정값을 구하고 나서 주기적 특성이 결여된 구간 5(총 24시간)의 누락된 24개의 수요실적자료에 대해서 확장된 ARIMA 모형을 순차적으로 적용하여 보정하였으며, 그 결과는 아래의 그림 7과 같다.

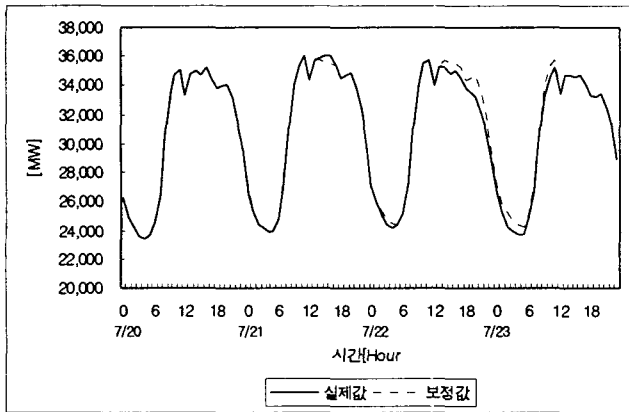


그림 7 시간대별 실측 수요값과 예측 수요값  
Fig. 7 Hourly measuring load curve and forecasting result curve

그림 7을 통해서 누락된 네 개의 구간에 대해서 실제 수요곡선과 보정된 수요곡선의 추이와 편차가 작은 것을 알 수 있으며, 각각의 누락구간에 대해서 실제 수요전력량[MWh]과 보정된 수요전력량[MWh]은 표 8과 같고, 결과에 대한 요약은 표 9와 같다.

표 8 실제 수요전력량값과 보정된 수요전력량값  
Table 8 Hourly measuring load data and forecasting result data

29536	26488	26866	30919	35894	36027	36028
29410	26925	27341	31168	35677	35621	35520
0.4	1.6	1.8	0.8	0.6	1.1	1.4
24967	24405	24152	33987	35279	35158	34738
25266	24771	24442	34127	35427	35709	35602
1.2	1.5	1.2	0.4	0.4	1.6	2.5
34958	34512	33831	33474	33219	32076	30875
35493	35173	34339	34461	34710	33548	32064
1.5	1.9	1.5	2.9	4.5	4.6	3.9
28769	26534	25179	24306	23938	23743	23780
29520	27036	25776	25192	24680	24379	24245
2.5	1.9	2.4	3.6	3.1	2.7	2.0

24844	26628	30420	33219	34600	35178	36028
25051	27151	30745	33922	35418	35701	35520
0.8	2.0	1.1	2.1	2.4	1.5	1.4

표 9 각 구간별 결과요약(사례연구 3)  
Table 9 Summary of result data for case 3

1구간	221,758	221,662	1.113	1.768	0.427
2구간	212,686	215,344	1.255	2.487	0.412
3구간	232,945	239,788	2.975	4.589	1.502
4구간	176,249	180,828	2.607	3.645	1.892
5구간	184,889	187,988	1.639	2.364	0.833

상기 표에서 전체 계통수요의 개별 구간에 대한 실측값과 보정값의 오차율은 평균 1.9%로 비교적 적은 오차율을 나타내고 있으며, 구간 3에서 13.974[%]의 최대 오차율과 구간 2에서 0.113[%]의 최소오차율을 나타내었다. 이를 통해 3차 구간보간법과 ARIMA 모형을 이용한 결측값 보정의 타당성을 검증하였다.

### 5. 결 론

본 논문에서는 통신상의 에러 및 검침기기의 고장 등으로 전력계통 상에서 발생할 수 있는 수요실적자료의 누락을 해결하기 위하여 시계열자료의 예측기법의 하나인 기존의 ARIMA 모형을 확장하여 누락된 검침자료를 보정하였으며 에너지 제약조건을 고려한 분기한정법을 제시함으로써 누락된 수요데이터에 대한 효과적인 보정방법을 제시하였다. 또한 본 논문에서 제시한 보정기법은 에너지사용 패턴이 서로 상이한 개별 수용가와 연간 하절기 수요실적자료에 적용하여 그 타당성을 검증 및 수행하였다. 제시한 기법은 오차율이 약 5% 미만의 예측 결과값을 나타내었으며, 이를 통해서 제시한 기법을 이용한 누락된 수요실적자료의 보정값이 신뢰할 수 있는 수요값으로 판단된다. 본 논문에서 제시한 기법은 안정적인 전력시장을 위한 단기간 수요예측에도 활용될 수 있으며, 부하별로 다양하고 신뢰성 있는 수요예측에도 적용이 가능하다. 또한 본 논문의 결과는 크게는 전력수요관리 계획 수립에 필요한 소비자 그룹별 전력수용예측의 정확성을 향상시켜 보다 정밀한 미래의 성과계량을 수행할 수 있는 기본 자료를 제공할 것으로 생각된다.

### 감사의 글

본 논문은 에너지관리공단의 학술진흥사업으로 수행된 논문입니다.



참 고 문 헌

- (1) Ministry of Commerce, Industry and Energy Korea Electricity Commission, "Restructuring of Electric Power Industry in Korea", October 2011.
- (2) 박종근, 김발호, 박종배, 정도영, "전력산업구조개편 개론", 기초전력공학공동연구소, 1999. 8.
- (3) Amjady, N., "Short-term hourly load forecasting using time-series modeling with peak load estimation capability", IEEE Trans. on Power Systems, Vol. 16, pp. 798-805, Nov. 2011.
- (4) 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.
- (5) 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.
- (6) M.Y.Cho, J.C.Hwang, C.S.Chen, "Customer Short Term Load Forecasting By Using ARIMA Transfer Function Model", IEEE Proceedings 1995, Vol. 1, pp. 21-23, Nov. 1995.
- (7) 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.
- (8) G. E. P. Box, G. M. Jenkins, G. C. Reinsel, Time Series Analysis Forecasting and Control, Third ed.
- (9) 김연형, "시계열 분석", 자유아카데미, 1994.
- (10) P. Venkataranman, Applied Optimization with Matlab Programming, John Wiley & Sons, pp.318-335, 2002.
- (11) 조신철/손영숙 "SAS/ETS를 이용한 시계열 분석", 울곡출판사, 2002.
- (12) SAS 교육팀, "기초통계분석", SAS Korea.

저 자 소 개



**박종배(朴宗培)**

1963년 11월 24일생. 1987년 서울대 공대 전기공학과 졸업. 1989년 동 대학원 전기공학과 졸업(석사). 1998년 동 대학원 전기공학과 졸업(박사). 현재 건국대학교 공과대학 전기공학과 조교수.

Tel.: 02-450-3483, Fax: 02-447-9186  
E-mail: jbaepark@konkuk.ac.kr



**신중린(愼重麟)**

1949년 9월 22일생. 1977년 서울대 공대 전기공학과 졸업. 1984년 동 대학원 전기공학과 졸업(석사). 1989년 동 대학원 전기공학과 졸업(박사). 현재 건국대학교 공과대학 전기공학과 교수.

Tel.: 02-450-3487  
E-mail: jrshin@konkuk.ac.kr



**이찬주(李瓊柱)**

1974년 12월 15일생. 2000년 안양대 공대 전기공학과 졸업. 2002년 건국대 대학원 전기공학과 졸업(석사). 현재 동 대학원 전기공학과 박사과정.

Tel.: 02-458-4778  
E-mail: jubily@konkuk.ac.kr



**이창호(李昌浩)**

1955년 10월 5일생. 1981년 전남대 사회대 학 졸업. 1987년 부산대 대학원 경제학과 졸업(석사). 1998년 동 대학원 경제학과 졸업(박사). 현재 한국전기연구원 전력산업연구실 실장

Tel.: 031-420-6120, Fax: 031-420-6129  
E-mail: chrhee@keri.re.kr



**이재용(李才勇)**

1976년 11월 1일생. 2002년 안양대 공대 전기공학과 졸업. 2004년 건국대 대학원 전기공학과 졸업(석사).

Tel.: 02-458-4778, Fax: 02-447-9186  
E-mail: jlee4800@hotmail.com