

ARIMA 모형과 Piecewise Cubic Interpolation을 이용한 누락된 수요실적자료의 보정기법

*이재용 · *이찬주 · *박종배 · *신중린 · **김성수
 *건국대학교 전기공학과 · **한국전력거래소

Correction Technique of Missing Load Data Using ARIMA Model and Piecewise Cubic Interpolation

*J. Y. Lee · *C. J. Lee · *J. B. Park · *J. R. Shin · **S. S. Kim
 *Dept. of Electrical Eng. Konkuk University · **Korea Power Exchange

Abstract - This paper presents a correction technique of missing load data. In this paper, the ARIMA(Autoregressive Integrated Moving Average) model and Piecewise Cubic Interpolation are applied to seek the missing parameters. The new model has been tested under a variety of conditions and it is shown in this paper to produce excellent results. It is helpful for operators to designed the load duration curve.

상확률과정인 ARMA(Autoregressive Moving Average) 과정을 따르게 된다. 그리고 ARMA 모형에 차분연산기법을 적용한 모형은 ARIMA(p, d, q)모형으로 변환되고, 주기성을 가지지 않는 자료의 예측 모형은 아래의 식(2)와 같다.

$$\phi_p(B)(1-B)^d y_t = \theta_q(B) a_t \quad (2)$$

단, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$
 $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

1. 서 론

여기서,
 p : AR (Autoregressive) 모형의 차수.
 q : MA (Moving Average) 모형의 차수.
 a_t : 평균이 0이고 분산이 σ_a^2 인 오차.

전력산업 구조개편에 따라 우리나라 전력산업은 과거의 수직 통합적 독점구조에서 발전, 송전, 배전부문의 기능분할과 함께 전력시장을 통한 부문별 경쟁을 단계적으로 추진중이다 [1,2]. 경쟁적 전력시장의 형성은 전반적인 전력산업구조의 변화를 가져다 줄 뿐만 아니라 최종 소비자의 에너지 소비형태에도 많은 변화가 있을 것이다. 특히, 다양한 소비자에 대한 정확한 에너지 소비형태의 분석은 불확실한 전력가격에 대한 에너지 소비 합리화의 기초 자료로 활용될 수 있으며, 나아가 안정적인 전력시장 운영에 기여할 수 있다. 하지만, 현재 우리나라의 수요 점검자료에는 통신상 장애등의 원인으로 인한 많은 불확실성을 내포하고 있기 때문에 정확한 전력에너지 사용형태에 대한 분석을 수행하지 못하고 있다. 따라서 효율적인 전력수요 관리를 위한 신뢰성 있는 과거 실적 부하자료의 생성이 무엇보다도 중요하다.

만약 시계열자료(y_t)가 주기성분(s)을 포함하고 있다면, 시계열자료는 계절 ARIMA(P, D, Q), 모형으로 변환되고, 이때 주기성을 가지는 자료의 예측모형은 식(3)과 같다.

$$\Phi_p(B^s)(1-B^s)^D y_t = \Theta_q(B^s) a_t \quad (3)$$

단, $\Phi_p(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps}$
 $\Theta_q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_q B^{qs}$

본 논문은 과거 실적 부하자료에 대해서 통계학적 자료 예측 기법인 ARIMA(Autoregressive Integrated Moving Average) 모형과 수치해석적 접근방법인 3차 구간보간법(piecewise cubic interpolation)을 이용하여 수요 점검자료에서 누락된 자료를 예측 및 보정하는 기법을 제시하였다. 본 논문에서 제시하는 방법은 부하별로 다양하고 신뢰성 있는 과거 실적부하곡선에 대한 분석에 활용될 수 있으며, 안정적인 전력시장을 위한 단기간 수요예측에 적용될 수 있다.

여기서, 주기성과 비정상성을 가진 시계열 함수(y_t)를 나타내면, 상기 식(3)은 ARIMA(p, d, q)(P, D, Q) 모형으로 확장되고, 아래의 식(4)와 같이 표현된다.

$$\Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_q(B^s)a_t \quad (4)$$

그리고 ARMA(p, q) 모형에서 ($t+l$)시점의 예측값은 아래 식(5)와 같이 표현된다.

2. 본 론

$$\hat{y}_t(l) = -\sum_{i=1}^p \phi_i E(y_{t+i-l}) + \sum_{i=1}^q \theta_i E(a_{t+i-l}) + E(a_{t+l}) \quad (5)$$

2.1 ARIMA 모형

전력계통에서 측정되는 수요데이터는 평균과 분산 및 자기공분산이 시간의 변화에 영향을 받는 비정상시계열 자료이다. 비정상시계열 자료를 연속적인 차분을 통해 정상 시계열로 변환은 Box-Jenkins에 의해서 제시되었으며[3-5], 제시한 모형은 전력계통의 수요예측 연구분야에 보고된 사례가 있다[6,7]. 이러한 비정상적인 시계열 자료(y_t)에 ARIMA 모형을 적용하면 시계열 함수(y_t)는 아래의 식(1)과 같다.

$$\nabla^d y_t = (1-B)^d y_t \quad (1)$$

단, B : 시간지연의 후향 연산자(backshift operator)
 상기 식(1)은 추세요인이 제거된 정상시계열이며 이것은 정

본 장에서는 비정상시계열 특성을 지닌 전력계통 수요자료에 대해서 ARIMA 모형의 정식화를 통해 나타내었다. 적용된 모형은 일정 기간동안 누락된 실적 수요자료의 수가 많은 경우에 적용하였으며, ARMA(p, q) 모형을 이용하여 누락된 실적값을 보정하였다.

2.2 Piecewise Cubic Interpolation을 이용한 보정

본 논문에서 적용한 ARIMA 모형은 누락된 실적 수요자료에 대해서 그 값을 예측하여 보정하는데 사용되지만, 모형 분석에 필요한 시계열 예측값의 개수가 적어도 50개 이상되어야 그 값을 신뢰할 수 있는 단점을 지니고 있다. 전력계통에서 측정되는 수요자료는 유·무선을 이용한 자료송신방식을 취하고

있으며, 이것은 통신장애등의 원인으로 인하여 취득한 수요자료의 연속성을 보장하기가 어렵다. 따라서 본 장에서는 누락된 자료의 연속성을 보장함과 동시에 예측값을 보정하기 위해서 수치해석적 접근방법론의 3차 구간보간법(piecewise cubic interpolation)을 적용하였으며, 아래의 식(6)과 같이 나타내었다.

$$P(x) = \frac{3h_k s^2 - 2s^3}{h_k^3} y_{k+1} + \frac{h_k^3 - 3h_k s^2 + 2s^3}{h_k^3} y_k + \frac{s^2(s-h_k)}{h_k^2} d_{k+1} + \frac{s(s-h_k)^2}{h_k^2} d_k \quad (6)$$

$$\text{단, } h_k = x_{k+1} - x_k, \quad \delta_k = \frac{y_{k+1} - y_k}{h_k}$$

$$s = x - x_k, \quad d_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k}$$

$$d_{k+1} = \frac{y_{k+2} - y_{k+1}}{x_{k+2} - x_{k+1}}$$

여기서, h_k 는 k-번째 부분구간의 길이를 나타내며, δ_k 는 k-번째 부분구간의 평균 기울기를 나타낸다. 그리고 d_k 는 x_k 에서 내삽의 기울기를 나타내며, $\delta_k \neq d_k$ 이다.

상기 식(6)은 $x_k \leq x \leq x_{k+1}$ 에서 아래와 같은 4가지 조건을 만족하는 Hermit Interpolants이다.

- $P(x_k) = y_k$
- $P(x_{k+1}) = y_{k+1}$
- $P'(x_k) = d_k$
- $P'(x_{k+1}) = d_{k+1}$

PCHIP(Piecewise Cubic Hermite Interpolating Polynomial)을 적용한 3가지 경우의 d_k 에 관한 정의는 아래와 같다.

- δ_{k-1} 과 δ_k 의 부호가 다른 경우
- δ_{k-1} 과 δ_k 중 하나가 0인 경우
- δ_{k-1} 과 δ_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(7)과 같다.

$$\frac{1}{d_k} = \frac{1}{2} \left(\frac{1}{\delta_{k-1}} + \frac{1}{\delta_k} \right) \quad (7)$$

만약 두 구간의 길이가 다르다면 상기 식(7)은 아래의 식(8)과 같이 변형된다.

$$\frac{w_1 + w_2}{d_k} = \frac{1}{2} \left(\frac{w_1}{\delta_{k-1}} + \frac{w_2}{\delta_k} \right) \quad (8)$$

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

본 논문에서는 PCHIP과 ARIMA 모형을 적용하여 누락된 수요자료를 보정 및 예측하였다. 전력수요자료는 각각의 부하 특성, 계절특성등에 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 PCIP을 이용하여 보정하였다.

3. 사례 연구

본 논문에서 제시한 방법론은 에너지 관리공단의 10년간 실측 전력수요데이터에 적용하였고, 이 가운데 1999년 하절기 주중 첨두 기간(7.20-7.24)을 이용하여 그 타당성을 검증하였다. 그리고 본 논문에서 제시한 방법론의 ARIMA 모형 차수 설정 및 예측치는 통계학적 범용 소프트웨어인 SAS Ver. 8.1을 사용하여 모의시험 하였다. 본 논문에서 적용한 하절기 시간대별 실측 수요데이터는 아래의 그림 1과 같고 각각의 누락구간에 대한 설명은 표 1과 같다.

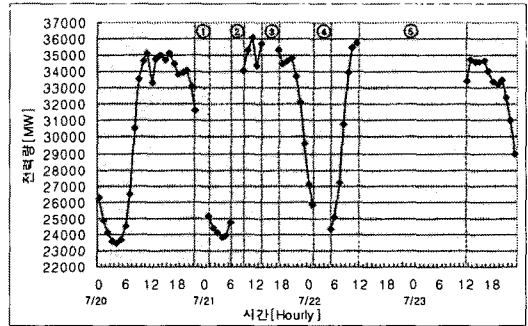


그림 1 시간대별 실측 수요자료

표 1 누락된 수요실적자료 구간

번호	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	$x_1^5, x_2^5, \dots, x_{23}^5, x_{24}^5$

단, x 는 해당 누락구간에서의 시간대별 누락된 수요실적자료

표 1의 누락된 수요실적자료 구간 ①은 δ 의 부호가 모두 (-)인 경우이며, 구간 ②는 δ 의 부호가 모두 (+)인 경우이고, 구간 ③은 δ 의 부호가 서로 다른 경우이다. 그리고 구간 ④는 δ 가 0인 경우이며, 주기성이 결여된 누락구간은 ⑤에 나타내었다. 이것은 시간대별 수요실적자료에서 네 구간(①,②,③,④)의 수요실적자료가 누락되어 주기적 특성을 갖는 ARIMA 모형을 적용은 부적합하다. 따라서 수요실적자료가 누락된 네 구간에 대해서 PCHIP을 이용하여 시간대별 수요실적값을 보정하였고, 적용된 보정값들은 아래의 그림 2와 같다.

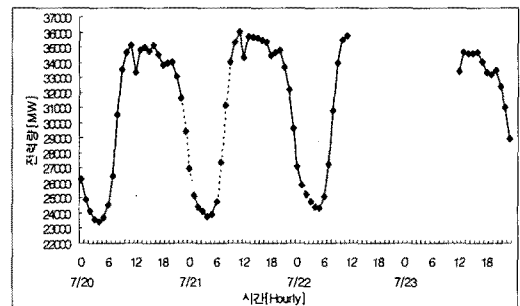


그림 2 PCHIP을 이용하여 보정된 수요자료

상기 그림2에서 실선은 주어진 수요실적자료이고, 점선은 네 개의 누락구간의 시간대별 누락된 수요실적자료를 PCHIP을

이용하여 구한 보정값을 나타낸다. PCHIP을 이용하여 수요자료의 보정값을 구하고나서 주기적 특성이 결여된 구간 (5(총 24시간)의 누락된 24개의 수요실적자료에 대해서 본 논문에서 제시하는 ARIMA 모형을 적용하여 보정하였으며, 아래의 그림 3과 같다.

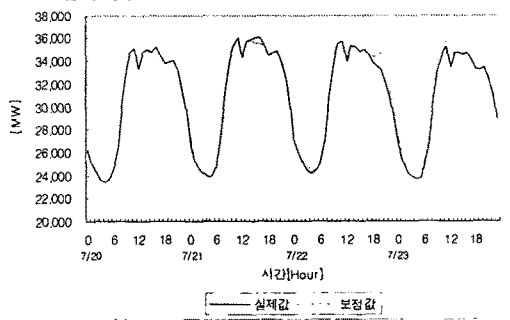


그림 4 시간대별 실적 수요값과 예측 수요값

그림 4를 통해서 누락된 네 개의 구간에 대해서 실제 수요곡선과 보정된 수요곡선의 추이와 편차가 작은 것을 알 수 있으며, 각각의 누락 구간에 대해서 실제 수요전력[MW]과 보정된 수요전력[MW]은 표 2에 나타내었다.

표 2 실제 수요전력값과 보정된 수요전력값

	7/20 23시	7/21 00시	7/21 07시	7/21 08시	7/21 14시	7/21 15시	7/21 16시
실제값	29536	26488	26866	30919	35894	36027	36028
예측값	29410	26925	27341	31168	35677	35621	35620
편차	126	-437	-475	-249	217	406	508
	7/22 02시	7/22 03시	7/22 04시	7/22 12시	7/22 13시	7/22 14시	7/22 15시
실제값	24977	24405	24152	33987	35279	35158	34738
예측값	25266	24771	24442	34127	35427	35709	35602
편차	-299	-366	-288	-140	-148	-551	-864
	7/22 16시	7/22 17시	7/22 18시	7/22 19시	7/22 20시	7/22 21시	7/22 22시
실제값	34958	34512	33831	33474	33219	32076	30875
예측값	35493	35173	34339	34461	34710	33548	33084
편차	-535	-661	-509	-987	-1492	-1472	-1189
	7/22 23시	7/23 00시	7/23 01시	7/23 02시	7/23 03시	7/23 04시	7/23 05시
실제값	28769	26534	25179	24306	23838	23743	23780
예측값	29520	27036	25776	25192	24680	24379	24245
편차	-751	-502	-598	-886	-742	-636	-465
	7/23 06시	7/23 07시	7/23 08시	7/23 09시	7/23 10시	7/23 11시	
실제값	24844	26628	30420	33219	34600	35178	
예측값	25051	27151	30745	33922	35418	35701	
편차	-207	-523	-325	-703	-818	-523	

그리고 각각의 누락구간에 대해서 실제 수요전수요전력값의 오차율은 아래의 그림 5와 같다.

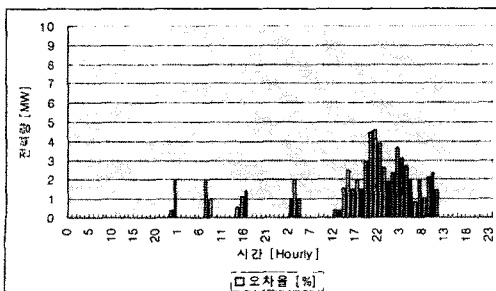


그림 5 시간대별 실적 수요값과 보정 수요값의 오차율

상기 그림 5를 통해서 실적 수요값과 보정 수요값의 오차율이 5% 미만으로 비교적 적은 오차율을 보여주고 있으며, 이를 통해서 본 논문에서 제시한 방법론의 타당성을 검증 및 수행하였다.

4. 결 론

본 논문에서는 ARIMA 모형과 PCHIP을 이용하여 전력계통에서 발생할 수 있는 수요실적자료에 대한 보정기법을 제시하였다. 또한 제시한 보정기법은 하절기 실적 수요실적자료에 적용하여 그 타당성을 검증 및 수행하였다. 제시한 기법은 오차율이 약 5% 미만의 수요예측 결과값을 나타내었으며, 이를 통해서 제시한 기법을 이용한 누락된 수요실적자료의 보정값이 신뢰할 수 있는 수요 예측값으로 판단된다. 본 논문에서 제시한 기법은 안정적인 전력시장을 위한 단기간 수요예측에 활용될 수 있으며, 부하별로 다양하고 신뢰성있는 수요예측에도 적용이 가능하다.

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 2001.
- [2] 박종근, 김발호, 박종배, 정도영, "전력산업구조개편 개론", 기초전력공학공동연구소, 1999. 8.
- [3] 김연형, "시계열 분석", 자유아카데미, 1994.
- [4] 조신섭/손영숙 "SAS/ETS를 이용한 시계열 분석", 율곡출판사, 2002.
- [5] SAS 교육팀, "기초통계분석", SAS Korea.
- [6] Amjady, N, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability", IEEE Trans. on Power systems, Vol. 16, pp. 498-505, Aug. 2001.
- [7] 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.