2005년도 대한전기학회 전력기술부문회 추계학술대회 논문집(2005.11.18~19)

ARIMA 모형을 이용한 계통한계가격 예측 방법론 개발

김대용† · 이찬주† · 박종배† · 신중린† · 전영환§
†건국대학교 전기공학과 · §홍익대학교 전기공학과

Development of SMP Forecasting Method Using ARIMA Model

Dae-Yong Kim† · Chan-Joo Lee† · Jong-Bae Park† · Joong-Rin Shin† · Yeong-Han Chun§
†Konkuk University · §Hongik University

Abstract - Since the SMP(System Marginal Price) is a vital factor to the market participants who intend to maximize the their profit and to the ISO(Independent System Operator) who wish to operate the electricity market in a stable sense, the short-term marginal price forecasting should be performed correctly. This paper presents a methodology of a day-ahead SMP forecasting using ARIMA(Autoregressive Integrated Moving Average) based on the Time Series. And also we suggested a correction algorithm to minimize the forecasting error in order to improve efficiency and accuracy of the SMP forecasting. To show the efficiency and effectiveness of the proposed method, the numerical studies have been performed using historical data of SMP in 2004 published by KPX(Korea Power Exchange).

1. 서 론

오늘날 국내 전력산업은 급속한 속도로 변하여 전력에너지도 일반 상품과 동일하게 경쟁시장에서 거래될 수 있는 경쟁적 전력시장이 형성되어 발전사업자가 생산한 전력을 한국전력거래소(Korea Power Exchange ; KPX)에서 판매하는 경쟁체제이다. 따라서 경쟁적 전력시장의 발전회사들은 자신의 수익을 최대화 할 수 있는 입찰전략을 필요로 하기 때문에 이를 해결하기 위한 많은 노력들이 시도되고 있다. 단기 전력가격 예측은 시장참여자들의 이익 극대화를 위한 중요한 요소가 된다. 따라서 정확한 계통한계가격 예측은 시장참여자들에게 있어서 이익을 극대화 할 수 있는 전략수립의 바탕이 된다. 기존에 사용된 예측방법론은 시계열자료 예측기법인 자기회귀누적이동평균(AutoRegressive Integrated Moving Average ; ARIMA)모형[1], 신경회로망(Neural Network)[2], 혼합 시스템(Hybrid System)[3] 등이 있다. ARIMA 모형은 단기예측의 정확성이 매우 높지만 새로운 자료가 투입될때 모형의 모수를 쉽게 업데이트할 수 있는 방법이 없다는 단점이 있다. 그리고 신경회로망은 복잡하고 다양한 문제에 대한 적용이 가능하나 느린 학습속도와 학습과정에서 비선형성과 오차 역전파 학습(back-propagation learning) 알고리즘을 사용하기 때문에 오차 표면상에서 바람직하지 못한 지역 최소치(local minimum)에 수렴하는 단점이 있다. 또한 혼합 시스템은 예측기법에 다양한 수학적인 도구를 적용하거나 예측방법론들을 결합하여 서로의 단점을 보완한다. 이와 같은 예측방법론 이외에도 다양한 예측기법이 시도되고 있다. 하지만 시장가격에는 많은 가변요소가 내포되어 있기 때문에 정확한 예측이 어렵다. 또한 대부분의 예측알고리즘이 과거 자료의 일정패턴에 의존하기 때문에 시변의 가격변동에 대한 불규칙적 동특성을 반영하기 어렵고, 특정일 또는 특정 시간대에 심각한 예측오차가 발생되는 것을 볼 수 있다. 따라서 원활한 시장형성을 위한 보다 정확한 시장가격 예측알고리즘이 필요하다.

본 논문에서는 계통한계가격 예측을 위해서 시계열 자료 예측기법 가운데 하나인 ARIMA 모형 기반의 계통한계가격 예측보정 알고리즘을 제시한다. 시계열 자료 예측에 있어서 가장 큰 영향을 미치는 것은 바로 직전의 계통한계가격이다. 이것을 입력 자료로 사용하는 시간대별과 주간대별은 월별을 입력요소로 사용함으로서 돌출적인 예측값을 보정한다. 본 논문에서 제시한 방법론의 타당성을 검증하기 위해서 2004년 8월(첨두기간) 한 달을 예측하였다. 그 방법론의 타당성 평가를 위한 기준으로 효율 및 정확성 측정을 위해서는 평균절대비율오차(Mean Absolute Percentage Error ; MAPE)를 사용하여 타당성을 검증하였다.

2. 본 론

2.1 일반적인 ARIMA 모형

전력계통에서 측정되는 계통한계가격 자료는 평균, 분산 및 자기공분산이 시간의 변화에 영향을 받는 비정상 시계열 자료이다. Box-Jenkins는 이 같은 비정상시계열 자료를 연속적인 차분(differencing)과 확률과정을 통해 정상시계열로 변환을 일반화하였다. ARMA 모형은 시계열 y_t가 과거의 시계열자료 값들로 표현될 수 있다는 것과 과거의 오차들의 선형결합으로 표현될 수 있다는 것을 나타내고 있다. 차수가 p와 q인 ARMA(p, q)모형은 다음과 같다.

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

여기서, ϕ_i : 자기회귀계수
p : 자기회귀모형의 차수
ε_t : 평균이 0이고 분산이 일정한 백색잡음
θ_i : 이동평균계수
q : 이동평균모형의 차수

ARMA 모형의 차수(p, q)가 높아지면 그들 모형에 포함된 계수(ϕ_i, θ_i)가 많아져서 모형이 복잡하게 된다. 따라서 이를 일반적으로 적용할 수 있는 단순화된 형태의 조건으로 표현하는 것이 필요하다. 이를 위하여 후향전위연산자(backward shift operator)를 이용하기로 한다. 후향전위연산자는 B^m으로 정의하며 연산의 지수 m은 차수를 나타내고 다음과 같이 표현한다[4].

$$(B^m) y_t = y_{t-m} \qquad (2)$$

또한 정식 (1)은 추세요인이 없는 정상시계열 자료에만 적용될 수 있는 모형이다. 그러나 현실적으로 존재하는 전력계통의 계통한계가격은 시간의 변화에 영향을 받는 비정상시계열이다. 이를 한 두 번의 차분과정을 거치면 대부분 추세요인이 제거된 정상시계열로 전환될 수 있다. 따라서 차분연산자인 $\nabla^d y_t = (1-B)^d y_t$를 사용하여 모형 식으로 표현하면 다음과 같다.

$$(1-B)^d(1-\phi_1 B^1 - \phi_2 B^2 - \ldots\ldots - \phi_p B^p)y_t$$

$$= (1-\theta_1 B^1 - \theta_2 B^2 - \ldots\ldots - \theta_q B^q)\varepsilon_t$$

$$(1-B)^d \phi_p(B) y_t = \theta_q(B)\varepsilon_t$$

$$y_t = \left(\frac{1}{1-B}\right)^d \frac{\theta_q(B)}{\phi_p(B)}\varepsilon_t \tag{3}$$

여기서, d : 차분차수

$\phi_p(B)$: $1-\phi_1 B^1 - \phi_2 B^2 - \ldots\ldots - \phi_p B^p$

$\theta_q(B)$: $1-\theta_1 B^1 - \theta_2 B^2 - \ldots\ldots - \theta_q B^q$

이처럼, ARMA모형과 차분연산기법을 합한 정식 (3)와 같은 모형을 자기회귀누적이동평균모형 또는 ARIMA (p, d, q)모형이라 한다. 여기서, 시계열 y_t는 과거의 시계열 값들과 오차값들에 의해 영향을 받는다는 것을 의미한다.

Box-Jenkins는 이론적으로 도출된 ARIMA(p, d, q)모형의 자기상관함수와 부분자기상관함수가 차수 p 또는 q의 값에 따라 독특한 형태를 가짐을 이용하여 시계열 자료의 적합한 모형과 차수를 선정하였다. 또한 모형의 정상성 여부를 점검하는데 유용하게 사용하고 과소차분(under-differencing)과 과대차분(over-differencing)을 점검하는 데에도 사용한다[5-6].

2.2 SMP예측 ARIMA 모형 유도

일반적으로 계통한계가격은 시간대별(hourly), 주간대별(weekly), 월별(monthly)에 따라 계통한계가격함수(y_t)에 대해 각각 서로 다른 영향을 받는다. 때문에 예측 계통한계가격함수는 각각의 인자에 따라서 일반화가 될 수 있으며, 이는 예측모형의 차수설정에 중요한 요소로 적용된다. 따라서 각각의 인자에 대한 t시간대 계통한계가격 예측모형은 정식 (3)을 이용하여 다음과 같이 나타낼 수 있다.

- 시간대별 모형 $$y_t^h = \left(\frac{1}{1-B}\right)^d \frac{\theta_q^h(B)}{\phi_p^h(B)}\varepsilon_t \tag{4}$$

- 주간대별 모형 $$y_t^w = \left(\frac{1}{1-B}\right)^d \frac{\theta_q^w(B)}{\phi_p^w(B)}\varepsilon_t \tag{5}$$

- 월별 모형 $$y_t^m = \left(\frac{1}{1-B}\right)^d \frac{\theta_q^m(B)}{\phi_p^m(B)}\varepsilon_t \tag{6}$$

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

$\phi_p^w(B)$: 차수 p를 갖는 주간대별 자기회귀모형

$\phi_p^m(B)$: 차수 p를 갖는 월별 자기회귀모형

$\theta_q^h(B)$: 차수 q를 갖는 시간대별 이동평균모형

$\theta_q^w(B)$: 차수 q를 갖는 주간대별 이동평균모형

$\theta_q^m(B)$: 차수 q를 갖는 월별 이동평균모형

2.3 보정 알고리즘

예측모형의 최종목표는 정확한 계통한계가격 예측에 있으며 확장된 ARIMA 모형을 이용하여 시간대별(y_t^h), 주간대별(y_t^w), 월별(y_t^m)에 대한 각각의 예측값을 다음과 같이 선정한다. 여기서 y_t^{max}는 t시간대 예측값(y_t^h, y_t^w, y_t^m)들 가운데 가장 큰 예측값을 나타내며 $\max[y_t^h, y_t^w, y_t^m]$으로 정의하고, y_t^{mid}는 t시간대 예측값들 가운데 중간 크기 값을 나타내며 $\mathrm{mid}[y_t^h, y_t^w, y_t^m]$으로 정의하고, y_t^{min}는 t시간대 예측값들 가운데 가장 작은 크기 값을 나타내며 $\min[y_t^h, y_t^w, y_t^m]$으로 정의한다. 이는 다음과 같은 식과 같다.

$$y_t^{max} = \max[y_t^h, y_t^w, y_t^m] \tag{7-a}$$

$$y_t^{mid} = \mathrm{mid}[y_t^h, y_t^w, y_t^m] \tag{7-b}$$

$$y_t^{min} = \min[y_t^h, y_t^w, y_t^m] \tag{7-c}$$

이 경우 모형에 의한 최선의 예측은 선정한 예측값들을 이용하여 예측오차를 최소화하는 것이다. 본 논문에서는 예측오차를 다음과 같은 식으로 최소화한다.

$$Y_t = \begin{cases} \dfrac{y_t^{max} + y_t^{mid}}{2} & (y_t^{max} - y_t^{mid} < y_t^{mid} - y_t^{min}) \\ \dfrac{y_t^{mid} + y_t^{min}}{2} & (y_t^{max} - y_t^{mid} > y_t^{mid} - y_t^{min}) \\ y_t^{mid} & (y_t^{max} - y_t^{mid} = y_t^{mid} - y_t^{min}) \\ y_t^{mid} & (y_t^{max} = y_t^{mid} = y_t^{min}) \end{cases} \tag{8}$$

여기서, Y_t : t시간대 최종 계통한계가격 예측값

3. 사례연구

본 논문에서는 KPX에서 제공하는 시간대별 계통한계가격 자료를 이용하여 2004년 8월 한 달간을 예측하고 제시한 알고리즘의 타당성을 검증한다. 그리고 제시한 방법론의 ARIMA 모형의 모수추정 및 예측치은 통계학적 범용 소프트웨어인 SAS Ver. 8.1을 사용하여 모의 시험한다. SAS의 ARIMA 프로시저에서는 ESTIMATE 문에서 CLS(Conditional Least Squares), ULS(Unconditional Least Squares), ML(Maximum Likelihood) 중 하나를 지정할 수 있으며 디폴트는 CLS 이다[4]. 이때, 시계열을 분석하기 위해서 일반적으로 필요한 시계열 자료 수는 최소 30개 이상이어야 정확한 예측을 할 수 있다.

그림1-4는 예측에 대한 모든 입력요소를 고려하여 2004년 8월 한 달간을 예측한 계통한계가격과 KPX에서 제공하는 실제 계통한계가격의 오차율을 나타낸다. 본 논문에서 제시한 ARIMA 모형별 예측값은 평균 3.27[%]의 비교적 작은 오차율을 가지고 있지만 특정 시간대에 큰 오차율을 가지는 것을 확인할 수 있다. 따라서 본 논문에서 제시한 보정 알고리즘을 통해 보정된 최종 계통한계가격의 오차율은 그림 4와 같다.

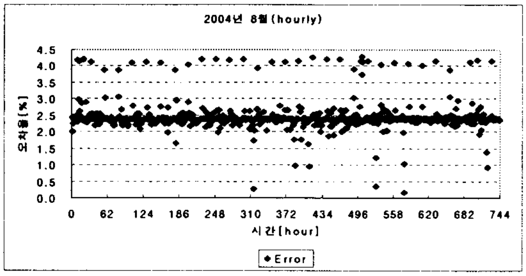

그림 1 시간대별에 대한 예측값과 실제값의 오차율

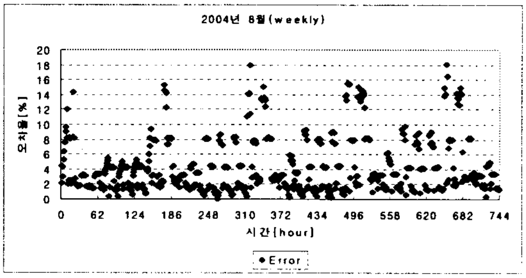

그림 2 주간대별에 대한 예측값과 실제값의 오차율

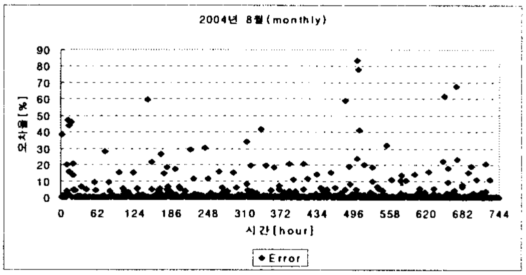

그림 3 월별에 대한 예측값과 실제값의 오차율

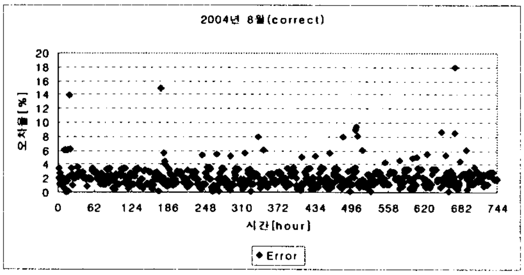

그림 4 보정알고리즘을 통한 예측값과 실제값의 오차율

예측보정 알고리즘의 타당성을 검증한 결과 요약은 표 1과 같다.

표 1 각 입력 자료별 결과 요약

방법론 / 오차율[%]	시간대별	주간대별	월별	산술평균	보정 알고리즘
8월 평균	2.45	4.48	2.89	3.27	**2.15**
주중 평균	2.43	3.30	1.67	2.47	**2.04**
주말 평균	2.50	7.38	6.00	5.29	**2.40**

표 1에서 보정 알고리즘 구간에 대한 실제값과 예측값의 오차율은 8월 한 달간 평균 2.15[%]로 매우 작은 오차율을 나타내고 있으며, 2.04[%]의 주중 평균 오차율과 2.40[%]의 주말 평균 오차율을 나타내었다. 그림 4에서 볼 수 있듯이 예측보정 알고리즘을 통해 예측된 예측값의 오차율은 매우 작은 값을 갖는 것을 알 수 있으며, 이를 통해 예측 결과값이 매우 정확하다는 것을 확인 할 수 있다.

4. 결 론

본 논문에서는 정확한 계통한계가격 예측을 위해서 시계열자료 예측기법의 하나인 ARIMA 모형을 확장하여 사용하였다. 또한 예측보정 알고리즘을 통해 각 입력요소별로 예측된 계통한계가격의 예측오차를 최소화하였다. KPX에서 제공하는 과거의 계통한계가격을 입력 자료로 사용하여 2004년 8월 한 달간을 예측하였으며, 이를 실제 계통한계가격과 비교 분석하여 타당성을 검증 및 수행하였다. 제시한 방법론은 평균오차율이 약 2.15[%]의 예측 결과값을 나타내었으며, 하루 평균 최대 오차일은 1일(일요일)이고 2.95[%]의 오차율을 갖고 최소 오차일은 11일(수요일)이며 1.78[%]의 매우 정확한 예측을 보였다. 이를 통해서 제시한 기법을 이용한 계통한계가격 예측값은 신뢰할 수 있으며 여러 시장참여들에게 그들의 목적에 부합할 수 있는 유용한 정보의 제공이 가능하다. 본 논문에서 제시한 예측보정 알고리즘은 안정적인 전력시장을 위한 단기간 수요예측에 활용될 수 있으며, 부하별로 다양하고 신뢰성 있는 수요예측에도 적용이 가능할 것이다. 또한 제시한 기법을 통해 각 발전회사는 경쟁적 전력시장에서 자신의 이익을 최대화하기 위한 최적의 입찰전략을 수립하는데 이용할 수 있으며, 시장운영자는 전력시장을 예측하고 분석할 수 있을 것으로 판단된다.

향후 연구에서는 주말 또는 특정일의 예측 정확도가 주중에 비해 상대적으로 떨어지는 단점을 보완하고 가끔 돌출적인 예측값이 발생하는 문제를 해결하기 위한 새로운 예측기법이 모색되어야 할 것으로 생각된다.

감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력연구원(R-2005-B-112) 주관으로 수행된 과제임.

[참 고 문 헌]

[1] J. Contreras, R. Espinola, F.J. Nogales, and A.J. Conejo, "ARIMA Models to Predict Next-Day Electricity Prices", *IEEE Transactions on Power Systems*, Vol. 18, No. 3, pp. 1014-1020, Aug. 2003.

[2] B.R. Szkuta, L.A. Sanabria, and T.S. Dillon, "Electricity Price Short-Term Forecasting Using Artificial Neural Network", *IEEE Transactions on Power Systems*, Vol. 14, No.3, pp. 851-857, Aug. 1999.

[3] A.A. El Desouky and M.M. ElKateb, "Hybrid Adaptive Techniques for Electric-Load Forecast Using ANN and ARIMA" *IEE Proc.-Gener. Transm. Distrib.*, Vol. 147, No. 4, pp. 213-217, July 2000.

[4] 박유성, 김기환, "SAS/ETS를 이용한 시계열자료분석I", 자유아카데미, 2002.

[5] 김연형, "시계열 분석", 자유아카데미, 1994.

[6] Wei and W.S. William, "Time Series Analysis : Univariate And Multivariate Methods", Addison-Wesley, 2005.