시계열 자료 분석기법에 의한 풍속 예측 연구

김건훈, 정영석, 주영철

한국에너지기술연구원 풍력발전연구단
(kkh4417@kier.re.kr / jung96@kier.re.kr / ycju@kier.re.kr))

Estimation Model of Wind speed Based on Time series Analysis

Kim, Keon-Hoon, Jung, Youngseok, Ju, Young-Chul

Korea Institute of Energy Research
(kkh4417@kier.re.kr / jung96@kier.re.kr / ycju@kier.re.kr)

Abstract

A predictive model of wind speed in the wind farm has very important meanings. This paper presents an estimation model of wind speed based on time series analysis using the observed wind data at Hangyeong Wind Farm in Jeju island, and verification of the predictive model. In case of Hangyeong Wind Farm and Haengwon Wind Farm, The ARIMA(Autoregressive Integrated Moving Average) predictive model was appropriate, and the wind speed estimation model was developed by means of parametric estimation using Maximum likelihood Estimation.

Keywords : 풍속예측모델(predictive model of wind speed), 최대우도추정법(Maximum likelihood Estimation.), ARIMA(Autoregressive Integrated Moving Average)

1. 서 론

최근 몇 년 사이의 풍력단지의 출력전력(power output)을 예측하기 위해 수많은 시스템이 개발되어 왔다. 이러한 예측에 적용되는 시간범위는 일반적으로 약 하루 또는 이틀 전의 전력시장에 대한 입찰상황(bidding condition)과 종래 발전소의 스케줄링 기법에 의해 정해진다. 따라서 예측시스템은 가급적이면 시간마다의 분석을 요구하지만 6시간에서 적어도 48시간까지 예상되는 출력전력을 제공하는 것이 보통이다. 여기에서 고려되어지는 몇 일의 시간 척도에 의한 단기예측과 0~3시간 범위의 매우 짧은 단기예측 사이에는, 기본적으로 모델링 관점으로부터 요구되는 시간범위가 매우 중요하다. 대기의 역학을 명백히 모델화하는 수치 예보시스템은 더 긴 시간에서 잘 예측되는 것에 반해, 매우 짧

은 범위는 일반적으로 이론적인 통계학상의 접근에서의 더 나은 예측결과를 나타내고 있으며, 현재 대부분의 전력 예측시스템은 수치 예보시스템을 기반으로 한다.

본 연구에서는 우선 시계열분석에 의한 풍속의 예측모형을 구현 하고자, 다양한 시계열모델 중에서도 ARIMA(Autoregr- essive Integrated Moving Average, 누적자기회귀이동평균 모형)라고 불리는 모형을 적용하여 제주 한경 풍력단지의 풍속의 예측모형을 만들고자 하였다.

2. 풍속 예측을 위한 시계열 모형

시간영역에서의 시계열자료의 분석은 Yule (1926, 1927)이 시계열을 과거 관측값들의 함수형태로 표현하고자 시도한 데서 시작하였다.[1)2)]그 후에 Walker(1931)가 자기회귀(autoregressive)의 개념을 제안[3)]하였고, Slutzky(1927)는 이동평균(moving average)의 개념을 제안하였으며, 이를 기초로 Walker (1962)가 자기회귀이동평균(Autoreg- ressive moving average; ARMA)모형을 개발하였다.[4)] 모형의 추정은 Mann과 Wald(1943)가 자기회귀(autoregressive model; AR)모형에서의 최대우도추정법(maximum likelihood estimation)을 제안하였으며[5)], Durbin(1959, 1960)이 자기회귀모형과 이동평균모형(moving average model; MA model)에서의 모수추정법을 제안하였다.[6)] 그러나 ARIMA 모형은 모수추정 계산의 복잡성 문제로 인하여 많이 사용되지 못하고 있다가 Jenkins와 Watts (1968), 그리고 Box와 Jenkins(1970)에 의해 효율적인 계산방법 및 삼 단계 모형적합법이 제안되면서부터 본격적으로 활용되기 시작하였다. 이러한 이유로 ARIMA 모형은 Box-Jenkins의 ARIMA 모형이라고도 불리며, 이들에 의하여 제안된 삼단계 모형적합법은 관찰된 시계열 자료를 하나의 시계열 모집단으로부터 추출된 표본으로 간주하여 이들이 어떤 확률적 성질을 만족하는가를 조사하고 통계적 추정 및 검정을 통하여 적절한 시계열모형을 수립하는 것이다. Box와 Jenkins에 의해 제안된 ARIMA 모형에 의한 분석방법은 모형의 식별(identification), 추정(estimation), 그리고 진단(diagnostic checking)을 한 주기로 하여 시계열자료에 가장 적합한 모형을 수립하게 된다.

2.1 ARIMA모형

시계열자료가 시간의 진행에 따라 추세를 가져서 평균이 다르거나 분산이 다른 경우에는 정상성을 가지지 않는다. 시계열자료가 시간의 진행에 따라 분산이 다른 경우에는 로그변환이나 제곱근변환 등의 분산안정화 변환(variance stabilizing transformation)을 실시한다. 또한 시계열자료가 시간의 진행에 추세를 갖는 경우에는 차분을 실시해서 추세에 따른 평균의 차이를 제거한다. 추세를 가지는 시계열자료 {yt}의 d차 차분시계열자료 {wt = ▽dyt}가 ARMA(p, q)과정이면 시계열 자료 {Yt}가 차수 p, d, q인 ARIMA 모형 (autoregressive integrated moving average model, ARIMA model, 누적자기회귀이동평균모형)을 갖는다고 하고, 이러한 시계열자료를 ARIMA(p, d, q)과정이라고 한다. 실제

1) GU Yule, Why do we sometimes get nonsense correlations between time-series.?, Journ. Roy. Stat. Soc, 1926.

2) Udny Yule, G., On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, Volume 226, pp. 267-298.

3) CAMPBELL, MJ, AND WALKER, AM: 'A survey of statistical work on the McKenzie River series of annual Canadian lynx trappings for the years 1821-1934, and a new analysis', J. Roy. Stat. Soc. A 140 (1977), 411-431.

4) Walker, A. M. (1962). Large-sample estimation of parametes for autoregressive pro-. cesses with moving-average residuals, Biometrika, 49, 117- 131

5) Mann, H.B. and Wald, A., 1943, On the statistical treatment of linear stochastic difference equations. Econometrica, 11,173-220.

6) J. Durbin, "Efficient Estimation of Parameters in Moving Average Models,". Biometrika,. vol. 46,. pp. 306-316, 1959.

시계열자료의 d는 0, 1, 2 중의 하나가 되며 일반적으로 차분의 차수는 3 이하로 가정된다.

ARIMA(p, d, q)는 다음과 같이 표현된다.

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-p} \quad (1)$$

이 식은 후향연산자를 이용하여 다음과 같이 간편하게 표현될 수 있다.

$$\phi(B) w_t = \theta(B) \epsilon_t \quad (2)$$

여기서,

$w_t = \triangle^d y_t = (1-B)^d y_t$

$\{\epsilon_t\}$는 평균이 0이고 분산이 σε2인 백색잡음과정

$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$: AR부분

$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$: MA부분

d는 차분 차수

또 위의 식은 다음과 같이 쓸 수 있는데, 이 ARIMA(p, d, q)모형은 d = 0이면 ARMA(p, q)모형, p = 0이면 IMA(d, q)모형, q = 0이면 ARI(p, d)모형이라고도 부른다.

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1 - B)^d y_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\epsilon_t \quad (3)$$

ARIMA 모형에 의한 시계열자료 분석에 있어서 가장 중요한 단계는 추세요인이나 계절요인과 같은 결정적 부분을 제거하거나 변수변환이나 차분 방법을 이용해서 비정상 시계열자료를 정상시계열자료로 변환하는 것이다. 그 다음 단계로는 이와 같이 변환된 정상 시계열자료에 적합한 모형을 선택하게 된다. 시계열자료에 자료에 모형을 적합시키는 Box와 Jenkins의 방법은

①모형의 식별(model identification)
②모형의 추정(model estimation)
③모형의 진단(model diagnostic checking)

의 세 단계를 거쳐서 진행된다. 모형적합을 위한 이들 3단계는 ARIMA 모형의 경우만이 아니라 다른 모형을 사용한 적합에도 동일하게 사용된다. 단지 적합되는 모형의 종류에 따라서 각 단계에서 사용되는 방법에는 약간의 차이가 있다.

2.2 모형의 식별

ARIMA 모형에서 모형의 식별(model identification)은 모형의 차수 p와 q를 정하는 것이다. 모형의 식별단계에서 가장 우선적으로 시도되어야 하는 작업은 시계열도표(time series plot), 표본상관도표(sample correlogram), 표본자기상관함수(sample autocorrelation function), 표본부분자기상관함수(sample partial autocorrelation function) 등의 그림을 그려서 시계열자료의 형태에 대한 대체적인 이해를 얻는 것이다.

모형을 식별할 때에 유의할 점은 p와 q의 값이 커지면 추정해야 할 모수의 수가 증가할 뿐만 아니라 모수 추정의 효율성도 떨어지고 예측모형의 해석에도 어려움이 있다는 것이다. 따라서 모형의 식별단계에서는 가능한 한 간단한 모형이 선호되며 이를 간결의 법칙(principle of parsimony, 모수절약의 원칙)이라고 부른다. 모형의 차수에 따른 모수의 수는 p+q+2가 되는데 AR(p)모형의 p값은 보통 2이하가 되며, 만약 p의 값이 3보다 커지게 되면 전체 모수의 수가 작은 ARMA(p, q)모형을 선택하는 것이 좋다.

모형의 식별 단계에서는

① 시계열도표(time series plot)에 의한 방법

② 표본 자기상관함수(sample autocorrelation function; SACF)와 표본 부분자기상관함수(sample partial autocorrelation function; SPACF)에 의한 방법

③ 통계량에 의한 방법

등의 세 가지 방법이 주로 사용된다.

2.3 모형의 추정

모형의 식별과정을 통하여 최적모형이 선택되면 선택된 모형에 포함된 모수를 가장 효율적인 추정방법으로 추정하여야 한다. 일반적으로 사용되는 모수추정 방법으로는 조건부 최소자승법, 비조건부 최소자승법, 최대우도법, 비선형추정법 등이 있다. 어느 방법이든지 안정적인 ARMA 모형의 적합을 위해서는 자료수가 최소한 30 이상이거나 또는 일반적으로 50 이상인 것이 바람직하다. 자료수가 이보다 작을 경우에는 가급적 AR모형의 적합을 시도하는 것이 좋다. 왜냐하면 대개의 경우 MA모형의 적합이 AR모형의 적합보다 힘들기 때문이다. ARIMA(p, d, q) 모형으로부터 추정해야 할 모수의 개수는 시계열의 평균, 오차의 분산을 포함하여 p+q+2 개이다.

일반적으로 많이 쓰이는 모형의 모수 추정 방법(model estimation)으로는 조건부 최소자승추정법(conditional least squares estimation method), 비조건부 최소자승추정법(unconditional least squares estimation method), 최대우도추정법(maximum likelihood estimation method), 비선형최소자승추정법(nonlinear least squares estimation method), 적률추정법(method of moment estimation), 베이즈추정법(Bayesian estimation method) 등이 있다.

2.3 모형의 진단

모형이 식별되고 모수가 추정된 후에는 추정된 모수의 유의성 검정이 실시된다. 추정된 모수들이 모두 유의하면 다음단계에서는 잔차분석을 통해서 잠정적으로 선택된 모형이 주어진 시계열자료에 잘 적합하는가를 확인하는 과정을 거치게 되는데 이를 모형의 진단(model diagnostic checking)단계라고 한다. 추정된 모수 중에 유의하지 않은 모수들이 있으면 이들을 모형에서 제거하고 모수추정 과정을 다시 거친 후에 모형의 진단을 실시한다.

추정된 모형이 관측된 시계열에 적합하면 추정 후의 잔차가 순수 오차인 백색잡음과정이 되어야 한다. 모형의 진단 단계에서는 주로 잔차분석(residual analysis)과 과다적합진단(overfitting diagnostics) 등을 수행한다. 잔차분석에는 잔차의 시계열도표, 잔차의 자기상관함수와 부분자기상관함수 등에 대한 검토가 이루어지며, 포트맨토(portmanteau) 통계량이나 포트맨토 통계량을 수정한 Box-Pierce 통계량이 사용된다. 잠정모형이 적합한 것으로 판정되면 이 모형은 예측모형으로 사용되지만 잠정모형이 적합하지 않은 것으로 판정되면 다시 모형의 식별단계로 돌아가서 모형의 진단단계에서 얻어진 정보를 활용하여 새로운 잠정모형을 탐색하게 된다.

3. 제주도내 풍력발전 단지 풍속 예측

3.1 제주 한경 풍력발전 단지

제주도내 풍력자원 실측치 중에서 제주 한경 풍력발전 단지의 건설을 위해 실측되었던 풍속측정자료를 이용하여, 한경 풍력발전 단지의 풍속을 예측코자 한다. 우선, 실측치는 한경 단지내에서 실측된 1999년 6월 1일 0시부터 2000년 5월 31일 23시까지의 60분 평균자료를 이용하였다.

(1) 한경 풍력단지 시계열도표

다음의 그림 1에는 한경 풍력단지 내에서 실측된 풍속자료의 시계열도표를 보이고 있다. 전체적으로 변화가 매우 심하며 어떤 추세는 보이지 않고 있으나, 평균값의 변화는 계절적으로 변화하는 모습을 보이고 있다. 따라서, 시계열도표의 정상성을 확보하기 위하여, 1차 차분이 필요할 것으로 판단된다.

다음의 그림 2에는 그림 1의 실측 풍속자료에 대한 1차 차분의 시계열도표를 보이고 있다. 차분하기 이전의 시계열 도표에 비해 훨씬 안정적임을 보이고 있고, 정상성을 가지고 있음을 보이고 있다. 즉, ARIMA의 모수 추정에서 d는 1로서 설정하여 분석하였다.

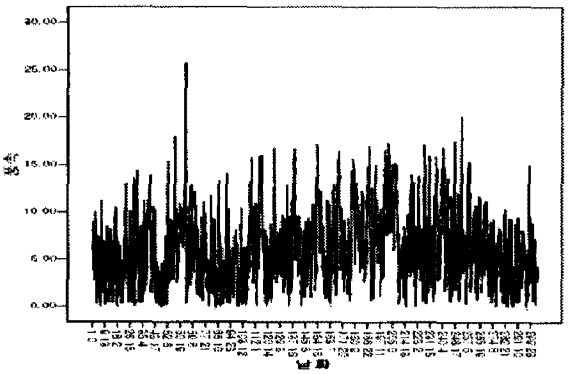


그림 1. 실측 풍속자료의 시계열도표(한경 풍력단지)

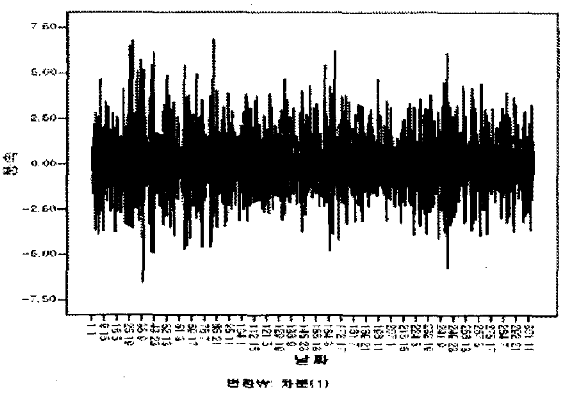
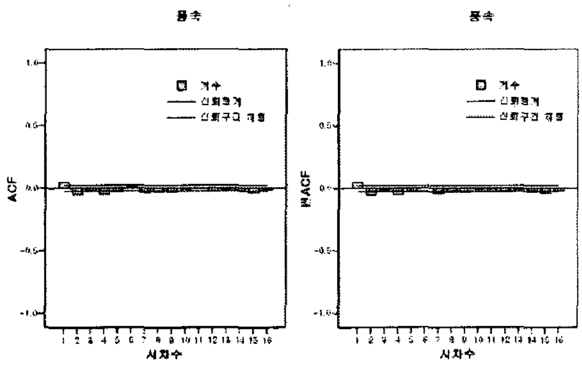


그림 2. 실측 풍속자료 1차 차분의 시계열도표 (한경 풍력단지)

(2) 풍속예측 모형의 식별

예측모형의 식별을 위해서 표본 자기상관함수(SACF)와 표본 부분자기상관함수(SPACF)에 의한 방법을 위해 풍속자료의 1차 차분에 대한 표본 자기상관함수와 표본 부분자기상관함수를 도출하였다. 다음의 그림 3 (a)에서는 1차 차분의 표본 자기상관함수를 보이고 있고, 그림 3 (b)에서는 표본 부분자기상관함수를 보이고 있다.



(a)표준 자기 상관함수 (b)표준 부분자기상관함수

그림 3. 한경 풍력단지 1차 차분의 SACF와 SPACF

AR모형의 차수 p를 주로 결정하는 표준자기상관함수나 MA모형의 차수 q를 주로 결정하는 표준 부분자기상관함수 모두 차수에 따라 급격하게 줄어들며, 특별히 95% 신뢰구간을 넘어가는 차수는 보이지 않는다. 두 함수 모두 1시차수 이후에는 급격히 줄어드는 형상으로서, 고차의 AR 차수 p나 고차의 MA 차수 q를 필요로 하지는 않고 있다. 따라서, 간결의 법칙에 의해 AR모형의 차수도 1로, MA 모형의 차수도 1로 추정하고, 예측모형을 형성코자 한다. 즉, ARIMA 모형의 차수를 각각, 1,1,1로 설정하여 ARIMA (1,1,1)을 풍속예측의 모형으로 추정하였다.

(2) 풍속예측 모형의 추정 및 예측

앞서 식별된 풍속 예측 모형 ARIMA (1,1,1)에 대해 모형을 추정하였다. 추정방법은 앞서 언급한 바와 같이 최대우도추정법을 사용하였고, 추정된 결과를 다음의 표 1에서 보이고 있다.

이상에서 추정된 모수와 식별된 모형에 의해서 한경 풍력단지내의 풍속에 대해 실측치와 예측치의 비교를 다음의 그림 4에서 보이고 있다. 그림에서 풍속의 비교는 2000년 5월 30일 0시부터 2000년 6월 2일 23시까지의 풍속의 실측치와 예측치를 비교 하였다.

표 1. 한경 풍력단지 풍속예측을 위해 추정된 모수

Number of residuals	7280
Standard error	1.1693729
Log likelihood	-11467.459
AIC	22940.919
SBC	22961.597
AR1	-.48130550
MA1	-.53937127
CONSTANT	-.00049024

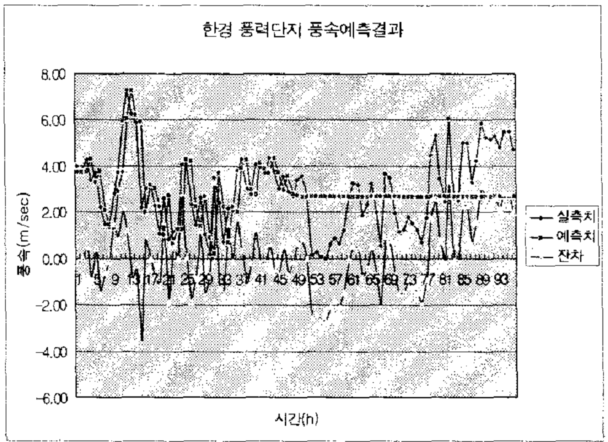
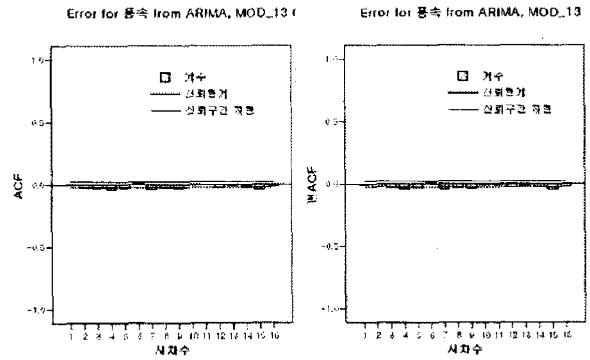


그림 4. 풍속 실측치와 예측치의 비교(한경 풍력단지)

풍속자료 예측을 위해서 실측된 자료가 사용되는 2000년 5월 31일까지의 예측치는 실측치에 대해 잘 예측하고 있음을 보이고 있다. 그러나 실측자료가 사용되지 않는 2000년 6월 1일 부터의 예측치는 일치하고 있지 못하고 있음을 보이고 있으며, 다만 풍속 변화의 평균적인 수치를 예측하고 있는 것으로 사료된다.

(3) 풍속예측 모형의 진단

예측된 모형의 진단을 위하여 잔차에 대하여 잔차의 자기상관함수와 부분자기상관함수를 도출 하였다. 그림 5에서와 같이 자기상관함수나 부분 자기상관함수 모두 모든 시차수에 대해서 신뢰한계 내에 존재하므로 풍속예측 모형은 적절한 것으로 볼 수 있다.



(a) 자기상관함수 (b) 부분 자기상관함수

그림 5 한경 풍력발전 단지의 풍속 잔차의 자기상관함수와 부분 자기상관함수

5. 결론

본 연구는 제주도내 실측자료를 통해 대표적으로 제주 한경 풍력단지의 풍속을 예측할 수 있는 모형을 개발하고 이의 모수를 추정한 후 예측모형의 검증을 수행 하였다. 제주 한경의 경우에 ARIMA(1,1,1)의 예측 모형이 적절한 것으로 식별되었고, 이에 최대우도추정법에 의한 모수의 추정이 이루어져 예측모형이 완성되었다. 그러나 과거의 실측치가 적용되는 예측기간에는 풍속 예측치와 실측치가 유사하게 일치하고 있으나, 과거의 실측치가 적용되지 않는 예측기간에는 빠르게 변화하는 풍속을 잘 추종하지 못하는 모형의 한계도 도출 되었다. 따라서 예측모형에 대해서는 수치기상모델을 포함하는 제반 모델의 검토가 필요할 것으로 보인다.

참 고 문 헌

1. 시계열자료의 분석과 예측, 2005, 김현철, 교육과학사
2. 시계열 분석과 응용, 2007, 이종협, 자유아카데미