KEPCO Journal on Electric Power and Energy, Vol. 2, No. 4, December 2016

Study on the Prediction of Wind Power Outputs using Curvilinear Regression

곡선회귀분석을 이용한 풍력발전 출력 예측에 관한 연구

Youngdo Choy*, Solyoung Jung**, Beomjun Park**, Jin Hur**†, Sang ho Park*, Gi gab Yoon*
최영도*, 정솔영**, 박범준**, 허진**†, 박상호*, 윤기갑*

* *KEPCO Research Institute, Korea Electric Power Corporation, 105 Munji-ro Yuseong-gu, Daejeon 34056, Korea*
** *Department of Energygrid, Sangmyung University, 20, Hongjimun 2-gil, Jongno-gu, Seoul, 110-743, Korea*
† *jinhur@smu.ac.kr*

Abstract

Recently, the size of wind farms is becoming larger, and the integration of high wind generation resources into power gird is becoming more important. Due to intermittency of wind generating resources, it is an essential to predict power outputs. In this paper, we introduce the basic concept of curvilinear regression, which is one of the method of wind power prediction. The empirical data, wind farm power output in Jeju Island, is considered to verify the proposed prediction model.

Keywords: Wind Power Outputs, Wind Power Predict, Linear Regression, Curvilinear Regression

I. INTRODUCTION

최근 전력생산의 많은 부분을 차지하고 있는 원자력발전의 위험성이 대두되고 있음은 물론이며, 기후협약에 따른 이산화탄소 감축 정책으로 인하여 화석에너지를 이용한 발전에 많은 제약이 발생하고 있다 [8]. 이러한 이유로 원자력발전과 화력발전을 대체하기 위하여 오염물질을 배출하지 않는 신재생에너지에 대한 관심이 급증하고 있다. 특히, 풍력발전의 경우 다른 신재생에너지를 이용한 발전원보다 경제성이 높다는 평가를 받고 있어 전 세계적으로 많은 관심이 집중되고 있다. 국내에서도 세계적인 추세에 따라 풍력발전의 시장이 확대되고 있으며, 풍력발전의 효율적인 계통 연계에 대한 관심 역시 증가되고 있다.

풍력발전의 효율적인 계통 연계를 위해서는 풍속 등 외부적인 요인에 따라 변화하는 풍력발전 출력 예측이 필수적이다. 이에 따라 최근 다양한 풍력발전의 출력 예측 기법이 제시되고 있다. 풍력발전 출력 예측 기법에는 대표적으로 ARMA (Auto Regression Moving Average)나 ARIMA (Auto Regression Integrated Moving Average)를 사용하여 출력을 예측하는 시계열 기반의 예측 [3][4], Kriging을 이용하는 공간적인 예측 [5][6] 등이 있다. 그 중 회귀분석(Regression Analysis)은 가장 기본적인 풍력발전 예측 기법 중 하나로, 종속변수(Dependant variable)와 독립변수(Independent variable) 간의 관계를 이용하여 쉽고 빠르게 출력예측을 수행할 수 있다는 장점이 있다.

본 논문에서는 회귀분석 기법을 이용하여 독립변수인 풍속과 종속변수인 풍력발전 출력의 상관관계를 분석하고, 이를 통해 풍력발전 출력 예측을 실시하였다. 이를 위해 제주지역에 위치한 풍력발전단지의 실제 풍속과 출력을 사용하였다.

II. 회귀 분석

회귀분석이란 각 변수 값들이 주어질 때 원하는 한 변수 값을 예측하기 위해 구체적인 관계식을 도출하는 통계적 기법이다 [1][2]. 즉, 객관적으로 나타난 자료를 기반으로 종속변수와 독립변수 간의 상관관계를 분석하는 기법으로 독립변수는 우리가 관심을 가지고 예측하고자 하는 변수에 영향을 미치는 인자로 설명변수라고도 한다. 또한, 영향을 받는 변수를 독립변수라 하며, 이를 반응변수라 한다. 종속변수를 y, 종속변수를 x라고 할 때 두 변수의 관계는 다음과 같이 표현될 수 있다.

$$Y = f(x) + \varepsilon \tag{1}$$

이때, ε는 확률오차를 의미하며, $Y = f(x)$ 로는 y의 변화를 충분히 설명하지 못하기 때문에 확률적인 부분을 의미하는 오차 항이 포함된다. 오차 항은 서로 독립적이며, $E(\varepsilon)$, $Var(\varepsilon)=\sigma^2$로 가정한다.

A. 회귀 분석의 분류

회귀 분석은 종속변수 예측을 위해 사용되는 독립변수의 개수에 따라 단순회귀분석과 다중회귀분석으로 나뉘게 된다 [1][7][9]. 종속변수의 개수가 1개이면 단순회귀분석, 2개 이상이면 다중회귀분석으로 분류된다. 또한 독립변수와 종속변수 간의 선형성여부에 따라 선형과 비선형모형으로 나뉜다.

B. 선형회귀분석(Linear Regression Analysis)

선형회귀분석은 종속변수 예측을 위해 사용되는 독립변수의 개수에 따라 단순선형회귀모형(Simple Linear

Manuscript received July 11, 2016, revised October 26, 2016, accepted November 3, 2016
ISSN 2465-8111(Print), 2466-0124(Online), DOI http://dx.doi.org/10.18770/KEPCO.2016.02.04.627

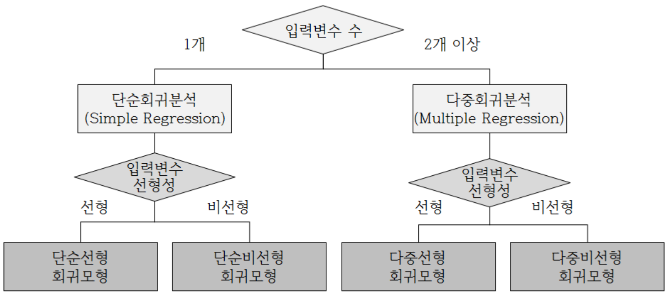


Fig. 1. 회귀분석의 종류.

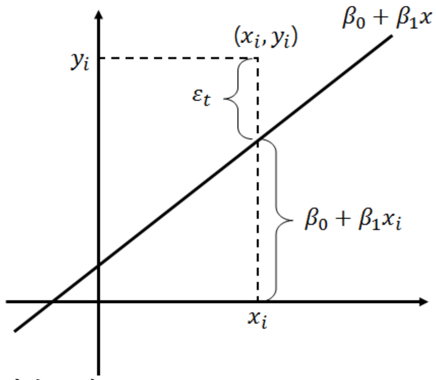


Fig. 2. 선형회귀 그래프.

Regression Model)과 다중선형회귀모형(Multiple Linear Regression Model)으로 나뉘게 된다. 단순선형회귀모형은 독립변수(설명변수) x가 하나인 모형이며, 다음과 같은 수식으로 표현할 수 있다. 이때, β_0, β_1는 회귀계수를 의미한다.

$$Y=\beta_0+\beta_1 x+\varepsilon \tag{2}$$

다중선형회귀모형은 설명변수의 개수가 2개 이상인 모형을 의미하며, 독립변수가 k개 일 때 다중선형회귀모형은 Eq. 3으로 표현된다. 이때, $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$는 회귀계수 또는 매개변수를 의미한다.

$$Y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k+\varepsilon \tag{3}$$

C. 회귀계수 추정기법

선형회귀모형의 회귀계수를 추정하는 방법 중 가장 기본적인 방법은 최소자승법(Least Square Method, LSM)이다. 최소자승법은 실제 값과 추정치간의 차이가 최소가 되도록 하는 방식으로 회귀오차 자승의 합이 최소가 되는 회귀식을 찾는 방식이다.

단순선형회귀모형을 기준으로 독립변수 x가 n개의 값을 가진다고 하고, 그에 따른 종속변수 y의 관측 값이 n개 있다고 가정하면, 단순회귀모형은 다음과 같이 표현될 수 있으며, Fig. 2와 같이 나타난다.

$$Y=\beta_0+\beta_1 x_i+\varepsilon_i, \quad i=1,2,\cdots,n \tag{4}$$

이때, ε_i는 서로 독립적이며, $E(\varepsilon)=0$, $Var(\varepsilon)=\sigma^2$로 가정한다.

Fig. 1에 따라 (x_i, y_i)에 대한 오차는 다음과 같이 표현되며, 오차 제곱의 합은 Eq. 6과 같이 표현된다.

$$\varepsilon_i=y_i-\beta_0-\beta_1 x_k \tag{5}$$

$$Q(\beta_0,\beta_1)=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(y_i-\beta_0+\beta_1 x_i)^2 \tag{6}$$

최소제곱법은 오차 제곱의 합 Q가 최소로 되는 회귀식을 찾는 방식이므로 오차를 최소화하기 위한 회귀계수를 구하기 위해서는 β_0, β_1에 대해 각각 편미분한 값이 0이 되어야 한다.

$$\left.\frac{\partial Q(\beta_0,\beta_1)}{\partial\beta_0}\right|_{\widehat{\beta_0},\widehat{\beta_1}}=-2\sum_{i=1}^{n}\left(y_i-\widehat{\beta_0}-\widehat{\beta_1}x_i\right)=0 \tag{7}$$

$$\left.\frac{\partial Q(\beta_0,\beta_1)}{\partial\beta_1}\right|_{\widehat{\beta_0},\widehat{\beta_1}}=-2\sum_{i=1}^{n}\left(y_i-\widehat{\beta_0}-\widehat{\beta_1}x_i\right)x_i=0 \tag{8}$$

$\widehat{\beta_0}$, $\widehat{\beta_1}$을 최소제곱추정값이라고 할 때, 위의 연립방정식을 만족하는 정규방정식(normal equations)은 다음과 같다.

$$n\widehat{\beta_0}+\widehat{\beta_1}\sum x_i=\sum y_i \tag{9}$$

$$\widehat{\beta_0}\sum x_i+\widehat{\beta_1}\sum x_i^2=\sum x_i y_i \tag{10}$$

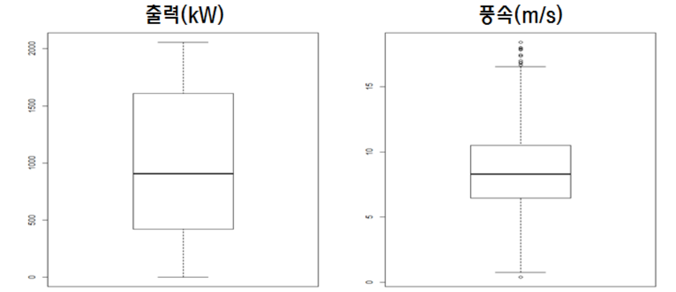
Eq. 9에서 얻은 $\widehat{\beta_0}=\bar{y}-\widehat{\beta_1}\bar{x}$를 Eq. 10에 대입하여 $\widehat{\beta_1}$을 얻을 수 있다. 즉, 최소제곱추정값 $\widehat{\beta_0}$, $\widehat{\beta_1}$는 다음과 같이 정리할 수 있다.

$$\widehat{\beta_1}=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \tag{11}$$

$$\widehat{\beta_0}=\bar{y}-\widehat{\beta_1}\bar{x} \tag{12}$$

D. 곡선회귀분석(Curvilinear Regression Analysis)

앞서 설명한 선형회귀분석의 경우 독립변수와 종속변수가 직선의 관계를 갖는 경우에는 적합한 분석 방법이지만, 독립변수와 종속변수가 곡선의 관계를 갖는 경우에는 많은 오차를 발생시킨다 [1][9]. 이와 같이 변수간의 관계가 직선이 아닌 자유 곡선의 관계를 갖는 경우 이를 근사하기 위해 곡선회귀분석을 사용할 수 있다. 곡선회귀분석은 독립변수가 1개인 회귀분석 모형으로, 독립변수 x, 종속변수 y에 대해 m차 곡선회귀분석은 Eq. 13과 같이 표현된다.



	Min	1st Qu	Median	3rd Qu	Max	Average
Output	0.00	421.10	905.10	1608.00	2057.00	973.80
Wind Speed	0.39	6.48	8.31	10.52	18.43	8.61

Fig. 3. 출력, 풍속 실측데이터 분포.

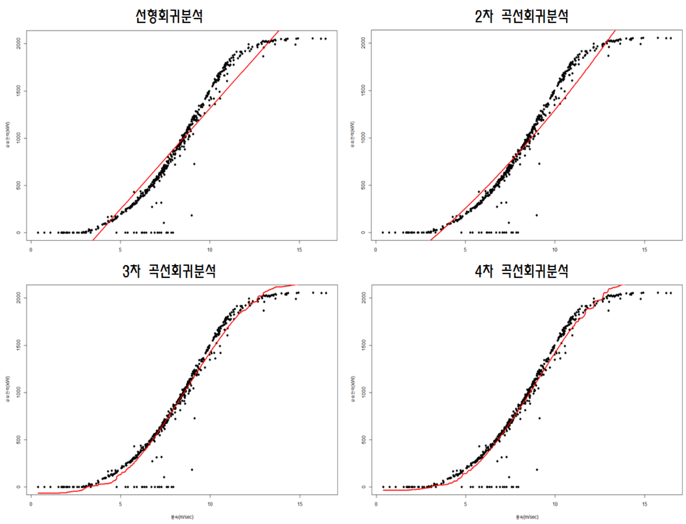


Fig. 4. 회귀모형 그래프.

$$y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \cdots + \beta_m x^m + \varepsilon \tag{13}$$

이때 $\beta_0, \beta_1, \ldots, \beta_m$은 다항회귀계수를 의미하며, ε는 오차 항을 의미한다.

III. 모형 검정 : 결정계수 R^2

결정계수 R^2는 독립변수 x가 종속변수 y를 얼마나 잘 설명하고 있는지를 나타내는 지수이다. 실제 값 y_i가 그 평균 $\bar{y}$과 떨어진 정도를 Eq. 14로 표현할 수 있다.

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}) \tag{14}$$

이때, $y_i - \bar{y}$는 총 편차(total deviation)의 분해식이며, $(y_i - \hat{y}_i)$는 실제 값 y_i와 회귀식을 통해 추정한 값 $\hat{y}$과의 차이로 잔차(residual)라고 하고 $(\hat{y}_i - \bar{y})$는 그 나머지인 회귀오차(regression error)이다. 여기서 총 편차의 제곱 합을 SST (Total Sum of Squares), 회귀에 의해 설명되는 편차의 제곱 합은 SSE (Error sum of squares)로 나타내며, 자료 전체의 분산 정도를 나타내는 SST 중에서 SSR이 클수록 회귀 모형이 적합하다는 것을 의미한다. 이를 결정계수 R^2라고 하며, 다음과 같이 표현한다.

Table. 1. 회귀모형별 회귀식

모형	회귀식
선형회귀모델	$\hat{y} = 183.4x - 605.4$
2차 곡선회귀모델	$\hat{y} = 133.332x - 4.989x^2 - 536.052$
3차 곡선회귀모델	$\hat{y} = -425.436x + 83.329x^2 - 3.201x^3 + 548.170$
4차 곡선회귀모델	$\hat{y} = -239.257x + 40.4535x^2 + 0.5951x^3 - 0.1135x^4 + 310.1574$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{15}$$

R^2는 0과 1사이의 값이며, 1에 가까울수록 더 적합한 회귀 모형임을 의미한다. R^2값을 통해 종속변수 y를 더 잘 설명하는 독립변수 x를 도출해 낼 수 있다.

IV. 시뮬레이션

A. 데이터 분석

회귀분석을 이용한 풍력발전 출력 예측을 위해 제주도에 위치한 A풍력발전단지(기술데이터의 보안을 위해 해당 발전소의 이름을 명시하지 않음)의 2016년 2월 1~3주의 1시간 단위 출력과 풍속 실측데이터를 이용하였다.

회귀분석을 수행하기에 앞서, 실측 된 데이터의 분포에 대한 분석을 수행하였다. Fig. 3을 통해 알 수 있듯이, 풍력발전 출력의 평균은 973.8 kW이었으며, 풍속의 평균은 8.61 m/s이었다.

B. 회귀분석을 이용한 풍력발전 출력예측

해당 데이터를 기반으로 선형회귀분석과 곡선회귀분석을 이용하여 풍력발전 출력 예측을 수행하였다. 예측 수행을 위해 회귀분석의 가장 기본이 되는 1차의 회귀모델을 산정하는 선형회귀분석 및 풍력발전과 풍속의 특성을 잘 반영할 수 있는 다차 회귀모델을 산정하는 곡선회귀분석을 이용하여 회귀모델을 구축했다.

Fig. 4와 같이 회귀모델을 구축하였으며, 회귀식을 산정하였다. 각각의 회귀모델에 대한 회귀식은 Table 1과 같다.

위 표와 같이 회귀식을 산출한 이후, 해당 식을 이용하여 A풍력발전단지의 2월 4주차의 풍력발전 출력 예측을 수행하였다.

C. 회귀분석을 이용한 풍력발전 출력예측 결과 비교

B절에서 2월 1~3주차의 풍속, 출력 데이터를 이용하여 구축한 회귀모델을 통해 2월 4주차의 출력 예측을 실시하였으며, 그 결과는 Fig. 5와 같다.

Fig. 5와 같이 예측 데이터가 실제 데이터와 서로 유사한 패턴을 보이는 것을 확인할 수 있었다. 예측 데이터의 정확도를 확인하기 위하여 실제 데이터와 예측 데이터의 회귀분석을 수행하였으며, 다음과 같은 결과를 얻을 수 있었다.

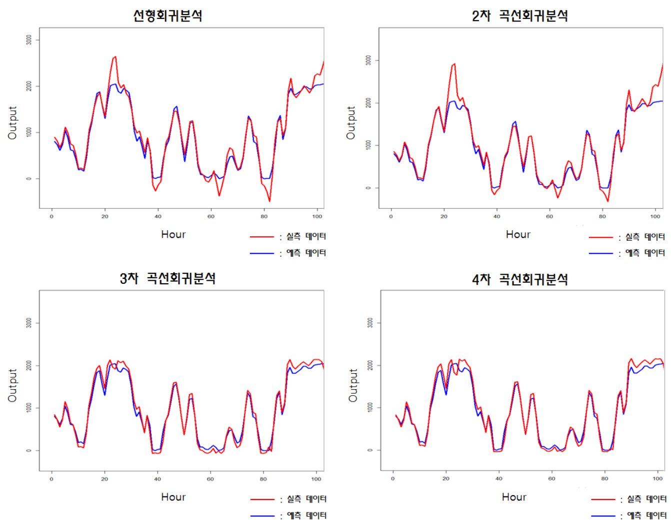


Fig. 5. 선형 회귀분석을 이용한 풍력발전 출력 예측 결과 그래프.

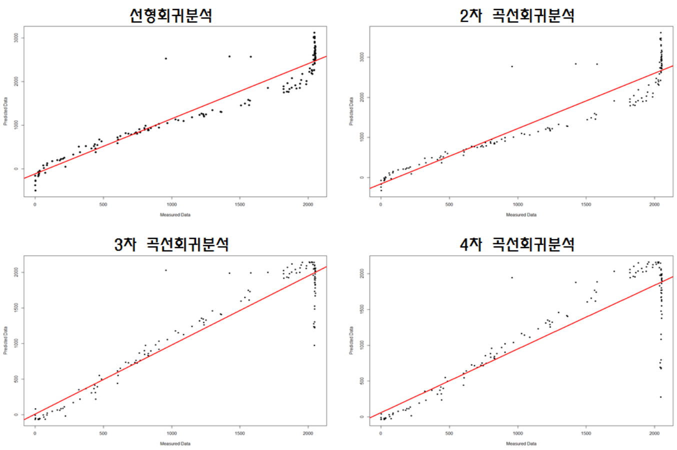


Fig. 6. 실측 데이터와 예측 데이터의 회귀분석 그래프.

두 데이터를 선형회귀분석을 이용하여 그래프로 나타내었을 때, 선형의 특성이 나타나는 경우 두 데이터 사이의 상관관계가 높고, 서로 유사함을 뜻한다. 이를 보다 명확히 파악할 수 있는 Correlation과 R-Squared값을 산출하였으며, 이는 Table 2에 표시되어있다.

Correlation과 R-Squared값 1에 가까울수록 상관관계와 유사성이 높다는 것을 의미한다. 따라서 위에서 수행한 네 가지 방법을 이용한 풍력발전 출력예측 데이터는 실제 측정된 데이터와 서로 상관관계와 유사성이 높은 것을 확인할 수 있다. 그 중 풍력발전의 출력과 풍속의 관계는 3차식으로 표현이 되기 때문에, 3차 곡선회귀분석을 이용한 출력 예측의 Correlation과 R-Squared값이 가장 높게 나타났다.

V. 결론

본 논문에서는 제주 지역 실제 풍력발전 출력과 풍속 데이터를 사용하여 단순선형회귀분석과 곡선회귀분석을 수행하여 회귀식을 도출하였으며, 이를 통해 풍력발전 출력 예측을 수행하였다. 수행결과 회귀분석을 통한 예측 결과와 실제 출력이 높은 상관관계와 유사성을 가지고 있음을 확인하였다. 특히 3차 곡선회귀분석의 경우 0.981의 높은 상관관계와 0.946의 높은 유사성이 나타남을 확인할 수 있었다.

향후에는 독립변수로 사용한 풍속 데이터 이외에 추가적인 독립변수를 이용하여 다중회귀분석을 수행하여 종속변수인 풍력발전 출력에 영향을 미치는 독립변수를 분석하고, 예측을 수행하고자 한다.

Table. 2. 예측 데이터와 실측 데이터의 Correlation 및 R-Squared 값

	선형회귀	2차 곡선회귀	3차 곡선회귀	4차 곡선회귀
Correlation	0.962	0.951	0.981	0.973
R-Squared	0.927	0.905	0.946	0.946

ACKNOWLEDGEMENT

This research was supported by Korea Electric Power Corporation through Korea Electrical Engineering & Science Research Institute. (Grant Number : R15XA03-26)

이 논문은 한국전력공사의 재원으로 기초전력연구원의 2015년 선정 기초연구개발과제의 지원을 받아 수행된 것임. (과제번호 : R15XA03-26)

REFERENCES

[1] 최종근, “지구통계학”, p110, 시그마프레스, 2007.
[2] Gareth James, Daniela Witten, Trevor Hasiem, “An Introduction to Statistical learning with Applications in R”, Springer, 2015.
[3] 위영민, “풍력발전 설비 효율화를 위한 다변량 분석을 이용한 풍력발전단지 단기 출력 예측 방법”, 조명전기설비학회논문지, 2015.7.
[4] Almoataz Youssef Abdelaziz1, Short Term Wind Power Forecasting Using Autoregressive Integrated Moving Average Approach, Journal of Energy and Power Engineering 7, 2013.
[5] 정솔영, 허진, 최영도, “공간모델링 기반의 풍력발전출력 예측 모델에 관한 연구”, KEPCO Journal, Vol 1, 2015.9.
[6] Nynke Hofstra, Mark New, “Short Communication Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe”, INTERNATIONAL JOURNAL OF CLIMATOLOGY, 1872~1880, 2009
[7] P. J. Woolcock and R. C. Brown, “A review of cleaning technologies for biomass-derived syngas”, Biomass and Bioenergy, Vol. 52, 2013, pp. 54-84.
[8] 최승용, 한건연, 김병현, 실시간 수위 예측을 위한 다중선형회귀 모형의 비교, 대한토목학회논문집, 2012.
[9] 김원정, 류준형, 박찬우 외 2명, “지속가능한 미래를 위한 에너지 신재생에너지”, pp258~312, 한티미디어, 2009.
[10] 황종대, 정종윤, 정윤교, “다항식회귀곡선을 통한 임펠러의 역공학 적용에 관한 연구”, 한국생산제조시스템학회, 2004.