

A study on solar irradiance forecasting with weather variables

Sahm Kim^{a,1}

^aDepartment of Applied Statistics, Chung-Ang University

(Received October 17, 2017; Revised October 28, 2017; Accepted October 28, 2017)

Abstract

In this paper, we investigate the performances of time series models to forecast irradiance that consider weather variables such as temperature, humidity, cloud cover and Global Horizontal Irradiance. We first introduce the time series models and show that regression ARIMAX has the best performance with other models such as ARIMA and multiple regression models.

Keywords: solar irradiance, ARIMA, ARIMAX, multiple regression, extraterrestrial radiation

1. 서론

화석에너지 자원의 한계와 화석연료 발전으로 인한 환경문제가 점차 대두됨에 따라 전 세계에서 신재생 에너지(renewable energy)에 대한 관심이 증가하고 관련 산업이 발달함으로 인하여 신재생에너지 발전량 예측에 대한 중요성이 높아지고 있다. 신재생에너지는 신에너지와 재생에너지를 통틀어 이르는 말이다. 기존의 화석연료를 변환하여 이용하거나 물, 강수, 바이오, 햇빛 등을 이용하여 재생이 가능한 에너지로 변환하는 에너지를 신재생에너지라고 한다. 국내에서도 미세먼지와 원자력의 위험에 대한 국민들의 인식 변화와 정부의 관심이 높아지면서 신재생에너지에 대한 정부 정책이 많이 나타나고 있다. 국내에서는 주로 발전되는 신재생에너지는 풍력과 태양광발전이 있는데 신재생에너지는 기존의 화석연료를 이용한 에너지와는 달리 기상과 같은 외부요인에 의한 발전량 변동이 매우 크다. 본 연구에서는 신재생 에너지의 한 종류인 태양광 발전량 예측을 위한 일사량 예측을 실시하고 여러 시계열 모형의 성능을 비교하고자 한다.

일사량의 연구에서 대부분 ARIMA 모형을 이용하여 연구가 이루어졌다. Moreno-Munoz 등 (2008)은 일사량을 예측하는데 승법 ARMA 모델을 사용하였고 ARMA류의 모형이 5분의 초단기 일사량 모형 적합에 적절함을 보였다. Reikard (2009)는 서로 다른 해상도에서 일사량 예측을 실시하였는데, 고해상도에서 운량 지수를 이용한 변이함수가 로그 변환 ARIMA 모형의 성능을 높여준다는 것을 확인하였고 ARIMA 모형이 다른 모형보다 일주주기를 더 효과적으로 포착한다는 것을 밝혀내었다. Martín 등 (2010)은 시계열 분석에 기반한 일사량 예측을 실시하였고 AR 모형 외 다양한 모형들의 성능을 비교하였다. Maleki 등 (2017)은 시간별, 일별, 월별 일사량 추정에 사용되는 다양한 모형들을 소개하였다.

This research was supported by the Chung-Ang University research grant in 2015.

¹Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea.
E-mail: sahm@cau.ac.kr

유사분야에서 ARIMA 모형의 활용 사례로서 Park과 Kim (2016)의 연구에서 태양광과 유사한 신재생 에너지인 풍력발전에서도 ARMAX를 포함한 다양한 시계열 모형을 이용하여 발전량 예측을 실시하였으며, Sohn 등 (2016)은 스마트 그리드에서 전력수요 예측 연구에서 시계열군집분석을 활용한 연구에서도 ARIMA 모형이 활용되었다.

본 연구에서는 태양광 발전량이 여러 기상 변수에 영향을 받지만 그 중 일사량에 의하여 가장 높게 영향을 받음에 따라서, 대기권 밖 일사량을 변수로 활용하여 연구를 실시하였다. 일사량을 예측하기 위하여 많은 연구자들이 다양한 통계적 방법을 이용하여 연구를 수행하고 있다. Dazhi 등 (2012)는 GHI, DNI, DHI, 운량지수를 각각 다르게 입력변수로 적용하고 ARIMA 모델을 사용하여 일사량을 예측하였고 GHI를 적용한 방법, DHI, DNI를 적용한 방법과 비교하여 운량지수를 이용한 방법이 더 좋은 예측력을 우수함을 보였다. Park 등 (2014)는 일사량과 대기 투과율 사이의 특성을 대기권 밖 수평면 일사량을 이용하였다. Huang과 Davy (2016)은 시간 내 일사량 변동을 예측하는데 청명지수와 500–850hPa 지오포텐셜 두께(geopotential thickness)와 같은 기상변수를 주요 예측자로 사용하였고 선형회귀를 이용하여 태양광 변동을 예측하였다.

본 연구에서는 기상청 기상자료개방포털에서 제공하는 기상자료를 이용하였으며, 일사량에 영향을 미치는 기상변수들과 기상변수 외에 일사량에 큰 영향을 미치는 대기권 밖 일사량 변수를 함께 고려한 시계열 자료의 예측에 널리 사용되는 ARIMA 방법, ARIMA 모형에서 독립변수를 추가하여 확장한 ARIMAX 방법과 일사량의 특성을 활용한 다중회귀모형을 이용하여 일사량의 예측 성능을 비교한다.

2. 시계열 모형

2.1. 다중회귀 모형

다중회귀모형(multiple linear regression model)은 종속변수의 변화를 설명하기 위하여 두 개 이상의 독립변수가 사용되는 선형회귀모형이다. 독립변수의 개수가 k (≥ 2)이고 관측치의 개수가 n 인 경우 다중회귀모형은 식 (2.1)과 같이 표현된다.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \varepsilon_t, \quad i = 1, 2, \dots, n. \quad (2.1)$$

2.2. ARIMA 모형

Auto-regressive integrated moving average (ARIMA) 모형은 auto-regressive (AR) 모형과 moving average (MA) 모형을 동시에 포함하는 시간의 흐름에 따른 관찰값들의 확률구조를 고려한 모형으로, 과거의 관측값과 오차를 사용해서 현재의 시계열 값을 설명한다. 평균이 μ 이고 차수가 p, d, q 인 ARIMA(p, d, q) 모형은 식 (2.2)와 같이 표현된다.

$$\begin{aligned} \phi(B)\dot{W}_t &= \theta(B)\varepsilon_t, \\ \text{where } \dot{W}_t &= (1 - B)^d(Z_t - \mu), \quad \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q. \end{aligned} \quad (2.2)$$

이 때, $\phi(B)$, $\theta(B)$ 는 각각 AR, MA에 대한 다항식이며 B 는 후진연산자(back shift operator)이고 d 는 Z_t 의 차분 차수이다.

2.3. ARIMAX 모형

Auto-regressive integrated moving average with exogenous variables (ARIMAX) 모형은 ARIMA

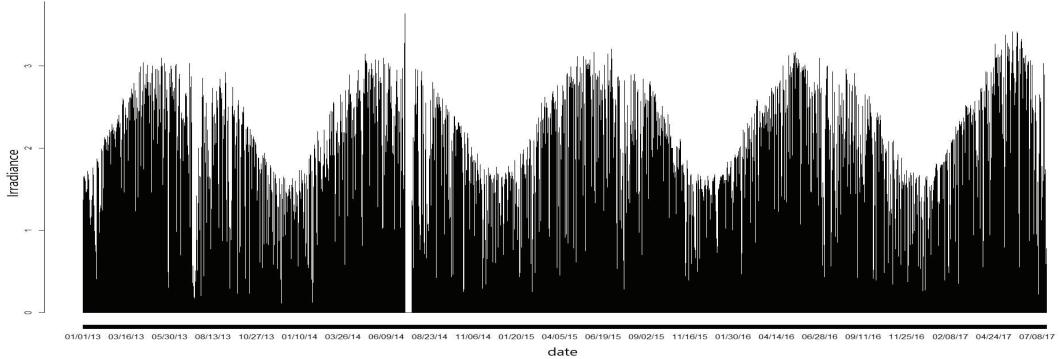


Figure 3.1. Solar irradiance time series plot of Seoul.

모형에 외생변수를 추가한 모형이다. 평균이 μ 이고 차수가 p, d, q 일 때, 외생변수 x_{at} 가 k 개인 경우 ARIMA(p, d, q) 모형은 식 (2.3)과 같이 표현된다.

$$\phi(B)\dot{W}_t = \theta(B)\varepsilon_t + \sum_{a=1}^k \gamma_a x_{at}$$

where $\dot{W}_t = (1 - B)^d(Z_t - \mu)$, $\phi(B) = 1 - \phi_1B - \dots - \phi_pB^p$,
 $\theta(B) = 1 - \theta_1B - \dots - \theta_qB^q$. (2.3)

이 때, $\phi(B), \theta(B)$ 는 각각 AR, MA에 대한 다항식이며 B 는 후진연산자이고 d 는 Z_t 의 차분 차수이다. γ_a 는 외생변수 x_{at} 의 계수이다.

3. 모형의 예측력 비교 결과

본 연구에서는 기상청에서 운영하는 기상자료개방포털(data.kma.go.kr)에서 2013년 1월 1일 00시부터 2017년 7월 24일 23시까지의 1시간 단위로 측정된 서울 108번 기상관측소의 기상관측 자료를 이용하였다. 기상자료 내에서도 예보가 가능하여 미래 예측에 사용할 수 있으면서 일사량과 연관이 높은 운량, 기온, 습도 변수를 이용하였다. 13년 1월 1일 00시부터 17년 7월 24일 23시까지의 일사량 시계열 도표는 Figure 3.1과 같다.

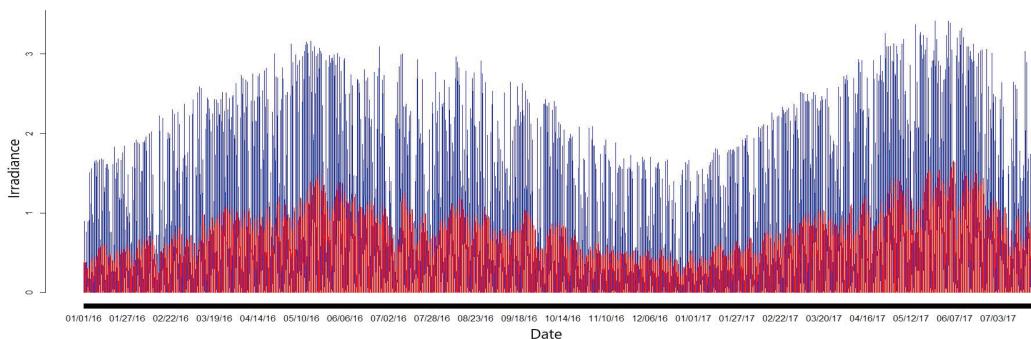
본 연구에서는 운량, 기온, 습도의 기상변수에 더하여 대기권 밖 일사량을 고려하였다. 대기권 밖 일사량을 계산하는데 필요한 모형과 모수들에 대한 설명은 Duffle과 Beckman (2013)의 Solar Engineering of Thermal Process (4th ed)을 참조하였다. 대기권 밖 일사량은 지구의 대기 상층 수평면에서 단위시간 동안 단위면적에 내리쬐는 일사량으로 태양고도에 따라 계절의 증감이 이루어고 1년 주기를 가진다. 대기권 밖 일사량(I_o)은 1년 중 시간변화에 따라 지구 상층권에서의 일사량으로 다음과 같이 표현된다.

$$I_o = \frac{12 \times 3600}{\pi} \cdot G_{sc} \cdot \left(1 + 0.033 \cdot \cos \frac{360 \cdot n}{365} \right) \\ \cdot \left(\cos \phi \cdot \cos \delta \cdot (\sin \omega_2 - \sin \omega_1) + \frac{\pi \cdot (\omega_2 - \omega_1)}{180} \cdot \sin \phi \cdot \sin \delta \right), \quad (3.1)$$

여기서 G_{sc} 는 태양상수($1,367 \text{ W/m}^2$)이며 δ 는 적위, ω 는 시간각, ϕ 는 추정하고자하는 지역의 위도이다. 지구 표면에 내리는 수평면 전 일사량(global horizontal irradiance; GHI)는 기본적으로 대기권

Table 3.1. Parameter Estimate of ARIMA(4, 1, 2) model

Parameter	Estimate	Standard error	p-value
ϕ_1	2.1222	0.0085	< 0.0001
ϕ_2	-1.3219	0.0161	< 0.0001
ϕ_3	0.0436	0.0146	0.0028
ϕ_4	0.1147	0.0068	< 0.0001
θ_1	-1.8518	0.0057	< 0.0001
θ_2	0.8549	0.0058	< 0.0001

**Figure 3.2.** Predicted value and real value of ARIMA(4, 1, 2) model.

밖 일사량에 근간을 두고 있다. 대기권 밖 일사량이 지구 대기를 통과하면서 대기 중의 수증기, 공기 분자, 먼지 등에 의해 산란 및 흡수되고 남은 에너지가 지상에 도달하는데 이것을 직달 일사량(direct irradiance)이라고 한다. 이 직달 일사량에 산란 일사량을 더한 것이 전일사량이다. 따라서 전일사량을 예측하는데 대기권 밖 일사량을 활용하는 것이 더 좋은 예측 성능을 나타낼 것이라고 추론할 수 있다.

13년 1월 1일 00시부터 17년 7월 24일 23시까지의 기상 변수 중에서 13년 1월 1일 00시부터 15년 12월 31일까지의 기상자료를 training set으로 이용하고 17년 1월 1일 00시부터 17년 7월 24일 23시까지의 기상자료를 test set으로 이용하였다. 동 기간의 대기권 밖 일사량은 (3.1)을 이용하여 서울 108번 기상 관측소의 경위도 값을 입력하여 계산하였다. 예측력 비교를 위한 모형 적합에는 통계 소프트웨어 R의 arima 모형, arimax 모형을 적합하기 위한 auto.arima 함수와 다중회귀모형을 적합하기 위한 lm 함수를 사용하였고 미래 예측에는 predict 함수를 사용하였다. 모형 적합 및 미래 예측 방법은 다음과 같다. 13년 1월 1일 00시부터 15년 12월 31일 23시까지의 자료를 이용하여 ARIMA 모형, ARIMAX 모형, 다중회귀모형을 적합한 후 16년 1월 1일 06시부터 20시의 기상자료를 이용해 16년 1월 1일 06시부터 20시의 일사량을 예측한다. 그 후 다시 13년 1월 1일 00시부터 16년 1월 1일 23시까지의 자료를 이용하여 모형을 갱신한 후 16년 1월 2일 06시부터 20시의 일사량을 예측한다. 이 과정을 571번 반복하여 16년 1월 1일부터 17년 7월 24일까지의 06 20시 일사량 예측값 8565개를 구하였다. 같은 기간의 실제 일사량과 비교해 MAE를 이용하여 각 모형의 예측 정확도를 평가하였다.

3.1. ARIMA 모형 분석 결과

ARIMA 모형은 종속변수의 과거 시점의 자료를 독립변수로 하여 만들어진다. 따라서 ARIMA 모형은 일사량 변수만을 사용하여 적합되었다. auto.arima 함수를 이용하여 Akaike's information criterion (AIC)를 기준으로 해당 값이 최소인 모형을 선택하였다. 그 결과, ARIMA 모형의 최적 모형은 ARIMA(4, 1, 2)이다. ARIMA(4, 1, 2) 모형의 모수 추정치는 Table 3.1과 같다. 해당 모형을 사용하여

Table 3.2. Parameter Estimate ARIMAX(2, 1, 3) model

Parameter	Estimate	Standard error	p-value
ϕ_1	1.8969	0.0185	< 0.0001
ϕ_2	-0.9623	0.0202	< 0.0001
θ_1	-1.8440	0.0098	< 0.0001
θ_2	0.8559	0.0156	< 0.0001
θ_3	0.0323	0.0088	0.0002
Temperature	-0.1258	0.0112	< 0.0001
Humidity	-0.0035	0.0006	< 0.0001
Cloud cover	-0.0195	0.0011	< 0.0001
Extraterrestrial radiation	0.2643	0.0067	< 0.0001
Drift	-0.00003	0.0007	0.9626

Table 3.3. Parameter estimate of ARIMAX(2, 0, 3) model

Parameter	Estimate	Standard error	p-value
ϕ_1	1.7750	0.0071	< 0.0001
ϕ_2	-0.8563	0.0058	< 0.0001
θ_1	-0.5678	0.0104	< 0.0001
θ_2	0.0057	0.0087	0.5119
θ_3	-0.0789	0.0077	< 0.0001
Temperature	0.0245	0.0159	< 0.0001
Humidity	-0.0131	0.0008	< 0.0001
Cloud cover	-0.0159	0.0003	< 0.0001
Intercept	1.0545	0.0008	< 0.0001

서울 지역 일사량을 예측한 결과는 다음 Figure 3.2와 같다. 파란색 실선은 일사량 실제 관측값을 나타내고 빨간색 실선은 모형을 통해 계산한 예측값을 나타낸다.

3.2. ARIMAX 모형 분석 결과

ARIMAX 모형은 ARIMA 모형에 외생변수 X를 추가한 모형이다. 본 연구에서는 외생변수로 3가지 기상변수(운량, 기온, 습도)와 대기권 밖 일사량을 이용하였고 대기권 밖 일사량의 영향을 살펴보기 위해 4가지 변수를 모두 이용한 모형과 대기권 밖 일사량을 제외하고 3가지 기상변수만을 사용한 모형을 비교하였다. ARIMAX 모형도 ARIMA 모형과 마찬가지로 auto.arima 함수를 이용하여 AIC를 기준으로 해당 값이 최소인 모형을 선택하였다. 그 결과 4가지 변수를 모두 이용한 ARIMAX 모형의 최적 모형은 ARIMAX(2, 1, 3)이고 대기권 밖 일사량을 제외한 3가지 기상변수를 이용한 ARIMAX 모형의 최적 모형은 ARIMAX(2, 0, 3)이다. ARIMAX(2, 1, 3) 모형과 ARIMAX(2, 0, 3)의 모수 추정치는 다음 Table 3.2, Table 3.3과 같다.

4가지 변수를 모두 이용한 ARIMAX(2, 1, 3) 모형을 사용하여 예측한 결과와 대기권 밖 일사량을 제외한 3가지 기상변수만 이용한 ARIMAX(2, 0, 3) 모형을 사용하여 예측한 결과는 각각 Figure 3.3, Figure 3.4와 같다.

3.3. 다중회귀모형 분석 결과

다중회귀모형의 종속변수는 일사량, 독립변수는 운량, 기온, 습도, 대기권 밖 일사량을 사용하였다.

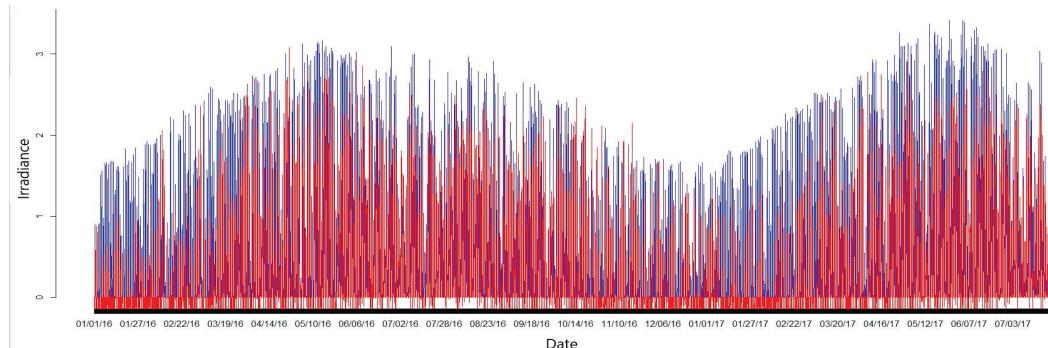


Figure 3.3. Predicted value and Real value of ARIMAX(2, 1, 3) model.

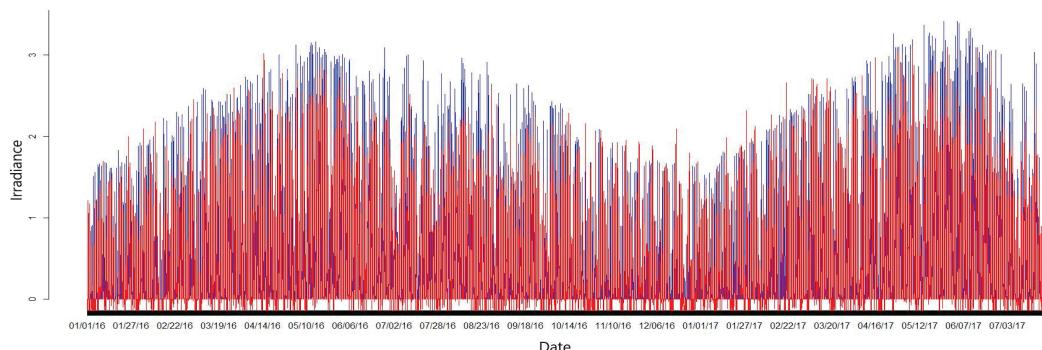


Figure 3.4. Predicted value and Real value ARIMAX(2, 0, 3) model.

Table 3.4. Parameter estimate of multiple regression model 1

Parameter	Estimate	Standard error	p-value
Temperature	0.0035	0.0003	< 0.0001
Humidity	-0.0009	0.0002	< 0.0001
Cloud cover	-0.0578	0.0009	< 0.0001
Extraterrestrial radiation	0.4053	0.0023	< 0.0001
Intercept	0.7449	0.0107	< 0.0001

Table 3.5. Parameter estimate of multiple regression model 2

Parameter	Estimate	Standard error	p-value
Temperature	0.0288	0.0004	< 0.0001
Humidity	-0.0208	0.0003	< 0.0001
Cloud cover	-0.0244	0.0015	< 0.0001
Intercept	1.6903	0.0154	< 0.0001

ARIMAX 모형과 마찬가지로 다중회귀모형도 4가지 변수를 모두 사용한 모형과 대기권 밖 일사량을 제외한 3가지 기상변수만을 사용한 모형을 비교하였다. 4가지 변수를 모두 사용한 모형과 대기권 밖 일사량을 제외한 3가지 기상변수만을 사용한 모형의 모수 추정치는 Table 3.4, Table 3.5와 같다.

4가지 변수를 모두 이용한 다중회귀모형 일사량 예측 결과와 3가지 변수만을 이용한 다중회귀모형 일사

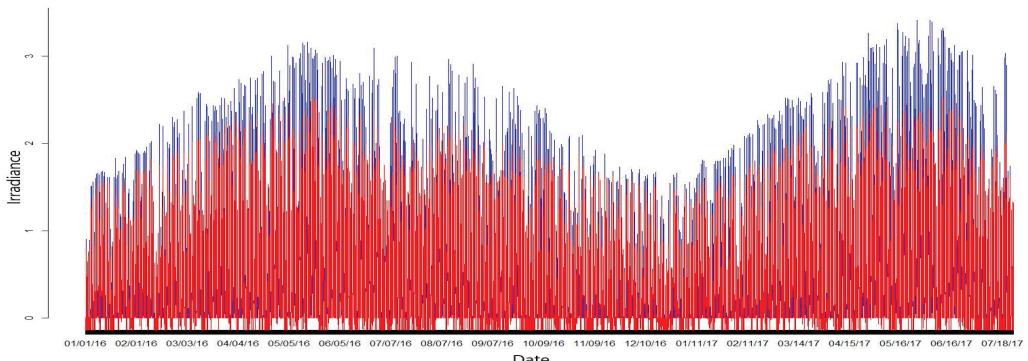


Figure 3.5. Predicted value and real value of multiple regression model 1.

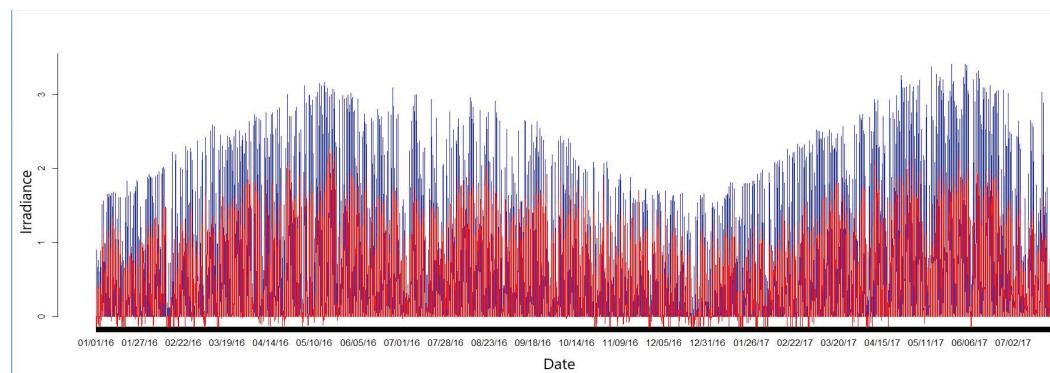


Figure 3.6. Predicted value and real value of multiple regression model 2.

량 예측 결과 도표는 Figure 3.5, Figure 3.6과 같다.

3.4. 각 모형의 MAE비교

모델 성능에 대한 평가기준은 mean absolute error (MAE)를 사용하였다. MAE와 통계량은 식 (3.2)와 같이 나타난다.

$$\text{MAE} = \frac{\sum_{i=1}^n |Y_t - F_t|}{n}. \quad (3.2)$$

이때, n 은 자료의 개수, 즉 모형의 기간을 나타내며, Y_t 는 실제값, F_t 는 예측값을 나타낸다. MAE 값이 작을수록 실제 일사량과 예측 일사량의 차이가 평균적으로 작다는 것을 의미한다. ARIMA, ARIMAX, 다중회귀모형의 모형 별 MAE는 Table 3.6과 같다. 모형 뒤의 1은 4가지 변수를 모두 사용한 모형을 의미하고 2는 대기권 밖 일사량을 제외한 3가지 기상변수만을 사용한 모형을 의미한다.

MAE 값을 살펴보면 대기권 밖 일사량을 고려한 모형의 MAE가 0.1895, 0.1793이고 그렇지 않은 모형의 MAE가 0.4137, 0.3209로 약 0.16–0.23정도의 차이가 나타남을 알 수 있다. 모형별 비교에서는 ARIMAX 모형과 다중회귀 모형의 MAE가 크게 차이나지 않고 ARIMAX 모형을 이용한 일사량 예측값을 나타낸 Figure 3.3과 다중회귀 모형을 이용한 일사량 예측값 Figure 3.5를 살펴보면 ARIMAX 모형이 다중회귀모형보다 실제 일사량의 패턴에 적합함을 확인 할 수 있다.

Table 3.6. Mean absolute error (MAE) of each model

Model	MAE
ARIMA	0.4121
ARIMAX1	0.1895
ARIMAX2	0.4137
다중회귀1	0.1553
다중회귀2	0.3146

4. 결론

본 연구는 태양광 발전량 예측에 필요한 일사량에 대한 예측에 다양한 시계열 모형과 외생변수를 고려하여 미래 일사량 예측을 하였다. 예측에 사용된 모형은 ARIMA, ARIMAX, 다중회귀모형이며 외생변수로는 다양한 기상변수들 중 실제 예보가 이루어지고 있어 향후 실무에 사용 가능한 기상변수 중 일사량과 연관성이 높은 변수들을 사용하였다. 또한 일사량에 근간이 되는 대기권 밖 일사량 변수를 추가로 사용하여 모형의 성능을 높였다. 예측의 평가 방법으로는 MAE를 사용하였고 모형 별 MAE 비교 결과 대기권 밖 일사량을 고려한 모형의 MAE가 가장 낮음을 알 수 있었다. 대기권 밖 일사량을 고려한 모형에서는 다중회귀모형과 ARIMAX 모형의 성능 차이가 크지 않았으나 예측값과 실제값 도표를 살펴보았을 때 ARIMAX 모형이 실제 일사량 패턴을 다중회귀보다 잘 맞추고 있었다. 이를 통해 일사량 예측에서 대기권 밖 일사량을 외생변수로 고려하는 것이 향후 일사량 예측에 있어 예측 정확도를 더 향상시키는 방안 중 하나임을 확인할 수 있었다. 향후 본 연구에 대한 확장으로 ARIMA, ARIMAX 이외에 ARMA-GARCH, Neural Network와 같은 모형을 고려하고 현재 사용한 기상변수 이외에 일사량에 영향을 미칠 수 있는 요인들을 탐색하여 모형에 추가하는 방안을 고려하고자 한다.

References

- Dazhi, Y., Jirutitijaroen, P., and Walsh, W. M. (2012) Hourly solar irradiance time series forecasting using cloud cover index, *Solar Energy*, **86**, 3531–3543.
- Duffle, J. A. and Beckman, W. A. (2013). *Solar Engineering of Thermal Process* (4th ed), John Wiley & Sons, New York, 3–41.
- Huang, J. and Davy, R. J. (2016). Predicting intra-hour variability of solar irradiance using hourly local weather forecasts, *Solar Energy*, **139**, 633–639.
- Maleki, S. A. M., Hizam, H., and Gomes, C. (2017). Estimation of hourly, daily, and monthly global solar radiation on inclined surfaces: models re-visited, *Energies* **2017**, **10**, 134–161.
- Martín, L., Zarzalejo, L. F., Polo, J., Navarro, A., Marchante, R. and Cony, M. (2010). Prediction of globar solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning, *Solar Energy*, **84**, 1772–1871.
- Moreno-Munoz, A., de la Rosa, J. J. G., Posadillo, R., and Bellido, F. (2008). Short term forecasting of solar radiation, *IEEE, International Symposium on Industrial Electronics*, 1537–1541.
- Park, J. K., Kim, B. S., and Park, J. H. (2014). Characteristic analysis of solar radiation and atmospheric transmissivity at Chupungryeong, *Journal of Agricultural Science*, **41**, 149–155.
- Park, S.-H. and Kim, S. (2016) A study on short-term wind power forecasting using time series models, *The Korean Journal of Applied Statistics*, **29**, 1373–1383.
- Reikard, G. (2009). Predicting solar radiation at high resolutions : A comparison of time series forecasts, *Solar Energy*, **83**, 342–349.
- Sohn, H., Jung, S., and Kim, S. (2016). A study on electricity demand forecasting based on time series clustering in smart grid, *The Korean Journal of Applied Statistics*, **29**, 190–203.

기상변수를 활용한 일사량 예측 연구

김삼용^{a,1}

^a응용통계학과, 중앙대학교

(2017년 10월 17일 접수, 2017년 10월 28일 수정, 2017년 10월 28일 채택)

요약

본 연구에서는 태양광 발전량 예측에 필요한 일사량을 예측하기 위해 다양한 기상변수를 활용한 다중회귀, ARIMA, ARIMAX 모형을 사용하여 각 모형의 예측 성능을 비교하고자 한다. 예측에 사용된 변수와 시계열 모형에 대해 소개하고, 실제 일사량 예측에 적용하여 일사량을 예측한 결과 운량, 기온, 습도, 대기권 밖 일사량을 활용한 ARIMAX 모형의 성능이 가장 우수하였다.

Keywords: 일사량, ARIMA, ARIMAX, 다변량 회귀분석, 외기 복사

이 논문은 2015년도 중앙대학교 연구년 결과물로 제출됨.

¹(06974) 서울시 동작구 흑석로 84, 중앙대학교 응용통계학과. E-mail: sahm@cau.ac.kr