

Solar radiation forecasting by time series models

Yu Min Suh^a · Heung-goo Son^b · Sahm Kim^{a,1}

^aDepartment of Applied Statistics, Chung-Ang University; ^bKorea Power Exchange

(Received October 25, 2018; Revised December 2, 2018; Accepted December 11, 2018)

Abstract

With the development of renewable energy sector, the importance of solar energy is continuously increasing. Solar radiation forecasting is essential to accurately solar power generation forecasting. In this paper, we used time series models (ARIMA, ARIMAX, seasonal ARIMA, seasonal ARIMAX, ARIMA GARCH, ARIMAX-GARCH, seasonal ARIMA-GARCH, seasonal ARIMAX-GARCH). We compared the performance of the models using mean absolute error and root mean square error. According to the performance of the models without exogenous variables, the Seasonal ARIMA-GARCH model showed better performance model considering the problem of heteroscedasticity. However, when the exogenous variables were considered, the ARIMAX model showed the best forecasting accuracy.

Keywords: ARIMA, ARIMAX, seasonal ARIMA, seasonal AIRMAX GARCH, weather variables, solar radiation

1. 서론

최근 원전 피해 위험성과 화석연료의 사용으로 인한 환경문제가 부각되고 있다. 에너지시장에서는 이러한 문제의 해결책으로 신재생에너지의 확대를 제기한다. 재생에너지 보급 확대와 함께 주목받는 기술로는 발전량 예측 기술이다. 기상 변동에 따른 재생에너지 발전량을 좀 더 정확히 예측해야, 원활한 전력 공급 대응이 가능하기 때문이다.

재생에너지의 시간대별 불규칙한 전력 생산에 따른 좀 더 안정적인 수급 대책 마련을 위해서, 정확한 재생에너지 발전량 예측이 필수적이다 (Kim 등, 2016; Kim 등, 2018). 한국에너지공단에서 발표한 ‘2017년도 재생에너지 보급통계(2016년 보급실적, 확정치) 결과 요약’에 따르면, 태양 에너지의 2016년 발전량은 총 512만 MWh로 전년도인 398만 MWh보다 약 28.7% 증가한 수치를 보였으며, 발전에 사용되는 재생에너지 자원들 중 가장 큰 32%의 발전 기여도를 기록하여 재생에너지 중 가장 높은 수치를 보였다. 태양광 발전은 발전 소자 개발 비용 감소 및 반도체 생산 기술이 발달함에 따라 발전량이 늘어가는 추세이며, 재생에너지 중에서 가장 높은 발전 기여도를 보인다. 태양광 발전량 예측은 기상 자료보다 일사량에 의한 영향이 강하므로 일사량을 정확하게 예측하는 것은 태양광 발전의 정확성을 위해 꼭 필요하다.

This research was supported by the Graduate Fellowship in 2018.

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

일사량 예측을 위하여 국내외에서 다양한 방법들이 시도되고 있다. 일사량은 태양 에너지가 지상으로 도착했을 때의 에너지양으로, 일사량 내의 추세를 찾아내기 힘들다. Ji와 Chee (2011)는 더 안정적인 월별 평균 일사량 자료를 찾기 위해 Al-Sadah 모형을 이용한 일사량의 추세를 없애는 과정을 실시하였고, 일사량의 선형 성분과 비선형 성분을 구별하여 선형 성분은 auto-regressive moving average (ARMA) 모형로, 비선형 성분은 time delay neural network (TDNN) 모형으로 예측하는 hybrid 모형을 제시하였다. Colak 등 (2015)는 ARMA 모형과 auto-regressive integrated moving average (ARIMA) 모형을 사용하여 1시간 단위의 일사량을 예측한 후, 다양한 예측 기간에 대한 예측 오차를 비교하였다. 또한 일사량은 시간에 따른 변동성이 큰 성질을 가지고 있기 때문에, 분산이 동일하다고 가정하는 일반 시계열을 이용하면 좋은 예측력을 장담할 수 없다. Grek (2014)는 일사량과 비슷하게 시간에 따라 변동이 나타나는 일별 거래량과 일간 거래량의 영향을 받는 주가지수를 예측하기 위하여 이분산성 모형인 generalized auto-regressive conditionally heteroscedastic (GARCH) 모형을 이용하였다. Liu 등 (2011)는 더 정확한 풍력 에너지 예측을 위해 10가지 시계열 모형을 통해 풍속 예측을 하였으며, 그 결과 ARMA-GARCH(-M) 모형의 예측 성능이 우수함을 보였다. 또한 Sun 등 (2015)는 중국의 베이징과 우루무치 지역의 일사량을 예측하는데 변동성을 고려한 모형인 7개의 다른 ARMA-GARCH 모형으로 일사량 값을 예측하여 그 결과들을 비교하였다. 본 논문에서는 일사량 자료의 분산의 이분산성을 설명할 수 있는 ARIMA-GARCH 모형을 이용하여 일사량 예측 성능을 ARIMA 모형과 비교하고자 한다.

태양 에너지는 지구의 대기권을 지나오면서 수증기, 운량 등과 같은 다양한 기상변수의 영향을 받아 실제 값보다 감소하는 경향이 있어 일사량 예측을 위해 외생변수를 고려해야 할 필요가 있다. 일사량 예측에 주로 고려되는 기상변수는 기온, 상대습도, 운량 등이 있다. Elizondo 등 (1994)는 기온과 습도를 포함한 다양한 기상변수를 이용해 일별 전지구적 일사량을 예측하였다. 그리고 Rehman과 Mohandes (2008)는 사우디 아라비아의 Abha city의 일사량을 예측하기 위해 당일 평균 기온과 최고 기온, 그리고 상대 습도를 이용하여 artificial neural network (ANN) 모형으로 일사량을 예측하였다. Kim (2017)은 일사량 예측을 위해 대기권 밖 일사량의 변수를 활용하여 시계열 모형의 예측 정확도를 비교하였다.

본 논문에서는 태양광 패널이 설치되어 있는 광주와 청주의 기상데이터를 활용하여 여러 시계열 모형을 이용하여 단기 예측 성능을 비교하고자 한다. 제 2장에서는 태양광 발전량 예측을 위한 시계열 모형을 소개하며, 제 3장에서는 활용된 일사량 데이터, 기상변수 데이터에 대하여 설명하고, ARIMA, ARIMA with eXogenous variable (ARIMAX), seasonal ARIMA, seasonal ARIMAX, ARIMA-GARCH, ARIMAX-GARCH, seasonal ARIMA-GARCH, seasonal ARIMAX-GARCH 모형들을 이용하여 일사량 예측 모형을 수립하고 분석 결과를 통해 비교 분석할 것이다. 마지막으로 제 4장에서는 결론과 향후 연구 방향에 대하여 논의할 것이다.

2. 일사량 예측 모형

2.1. ARIMA 모형

시계열에서 ARIMA 모형은 현재의 시계열 값이 과거의 관측값으로 설명된다는 AR 모형과 현재의 시계열 값이 오차로 설명된다는 MA 모형과 함께 차분 절차를 포함하는 모형이다. AR 모형의 차수가 p , MA 모형의 차수가 q , d 차 차분한 ARIMA(p, d, q) 모형식은 다음과 같이 표현된다.

$$\begin{aligned} \phi_p(B)(1-B)^d(Y_t - \mu) &= \theta_q(B)\epsilon_t \\ \text{where } \theta_q(B) &= 1 - \theta_1 B - \dots - \theta_q B^q, \\ \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \end{aligned} \quad (2.1)$$

이때, Y_t 는 t 시점에서 관측된 데이터이며, B 는 후진연산자, μ 는 평균을 의미한다. $\theta_q(B)$ 와 $\phi_p(B)$ 는 각각 q 차, p 차 다항식으로 p, q 는 모형의 차수를 d 는 모형의 차분 차수를 의미한다.

2.2. ARIMAX 모형

ARIMAX 모형은 ARMA 모형에 차분을 가한 후 독립변수로 외생변수를 추가한 모형이다. 이 때 독립변수들과 종속변수는 안정적인 시계열이며, 오차가 ARIMA 모형을 따른다고 가정하는 것이 특징이다. ARIMAX 모형은 아래와 같이 표현된다.

$$\phi_p(B)(1-B)^d \left(Y_t - \sum_{i=1}^k \beta_i X_{it} \right) = \theta_q(B)\epsilon_t$$

where $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$,

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (2.2)$$

Y_t 는 종속변수, X_{it} 는 독립변수이며 ϵ_t 는 ARIMA(p, d, q) 모형을 따르는 오차항에 해당하며 $\epsilon_t \sim WN(0, \sigma^2)$ 를 따른다. ARIMAX 모형에선 외생변수인 X_t 가 다양한 시차를 간격을 두고 독립변수에 영향을 주는데, 이를 후진연산자 B 로 표현하였다. β_i 는 회귀계수에 해당된다. 본 논문에서는 외생변수로 기온, 습도, 강수, 풍향 등 예보가 되는 기상변수들을 이용하였다. $\theta_q(B)$ 와 $\phi_p(B)$ 는 각각 q 차, p 차 다항식을 나타낸다.

2.3. Seasonal ARIMA 모형

Seasonal ARIMA 모형은 ARIMA 모형에 주기성이 존재하는 모형이다. 모형은 ARIMA(p, d, q) \times (P, D, Q) $_S$ 로 표현되며 평균이 μ 인 모형의 식은 다음과 같다.

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D(Y_t - \mu) = \theta_q(B)\Theta_Q(B^S)\epsilon_t$$

where $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$,

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}$$

$$\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS} \quad (2.3)$$

이때, Y_t 는 t 시점에서 관측된 데이터이며, B 는 후진연산자를 의미한다. $\theta_q(B)$ 와 $\phi_p(B)$ 는 각각 q 차, p 차 다항식으로 p, q 는 모형의 차수를 d 는 모형의 차분 차수를 의미한다. $\Theta_Q(B^S)$ 와 $\Phi_P(B^S)$ 는 각각 P 차, Q 차 차수를 가지는 계절 AR, MA를 의미하며, S 는 계절 차분 차수를 의미한다.

2.4. Seasonal ARIMAX 모형

Seasonal ARIMAX 모형은 ARIMA 모형에 주기성이 존재하며, 독립변수로 외생변수를 추가한 모형이다. 모형은 ARIMAX(p, d, q) \times (P, D, Q) $_S$ 로 표현되며 평균이 모형의 식은 다음과 같다.

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D \left(Y_t - \sum_{i=1}^k \beta_i X_{it} \right) = \theta_q(B)\Theta_Q(B^S)\epsilon_t$$

where $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$,

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\begin{aligned}\Theta_Q(B^S) &= 1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}, \\ \Phi_P(B^S) &= 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS},\end{aligned}\quad (2.4)$$

여기서 Y_t 는 종속변수, X_{it} 는 독립변수이며 ϵ_t 는 모형을 따르는 오차항에 해당하며 $\epsilon_n \sim WN(0, \sigma^2)$ 를 따른다. 외생변수인 X_t 가 다양한 시차를 간격을 두고 독립변수에 영향을 주는데, 이를 후진연산자 B 로 표현하였다. β_i 는 회귀계수에 해당된다. 본 논문에서는 외생변수로 기온, 습도, 강수, 풍향 등 예보가 되는 기상변수들을 이용하였다. $\theta_q(B)$ 와 $\phi_p(B)$ 는 각각 q 차, p 차 다항식을 나타낸다. $\Theta_Q(B^S)$ 와 $\Phi_P(B^S)$ 는 각각 P 차, Q 차 차수를 가지는 계절 AR, MA를 의미하며, S 는 계절 차분 차수를 의미한다.

2.5. ARIMA-GARCH 모형

ARIMA-GARCH 모형은 평균 모형이 ARIMA, 분산 모형이 GARCH 모형을 따르며, 이 모형을 통해 독립변수의 영향을 파악하여 종속변수의 변동을 예측할 수 있다. 분산은 이분산성 모형인 GARCH 모형을 통해 이분산성을 분석한다. 기존의 ARIMA 모형에서 가정하였던 등분산성이 아닌 이분산성을 가정하는 모형이 ARCH 모형이다. GARCH 모형은 ARCH 모형에서 현재 분산이 과거 분산의 영향을 고려할 수 있도록 확장한 모형이며, 조건부 분산의 자기회귀형태가 모형에 추가된다. ARIMA-GARCH는 ARIMA의 차수가 p, d, q 이고 GARCH의 차수가 r, m 인 것을 말하며, 모형은 다음과 같다.

$$\begin{aligned}\phi_p(B)(1-B)^d Y_t &= \theta_q(B)\epsilon_t, \\ \epsilon_t &= \sqrt{h_t}\epsilon_t, \\ h_t &= \alpha_0 + \sum_{i=1}^m \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^r \beta_i h_{t-i}\end{aligned}$$

where $\alpha_i, \beta_i \neq 0$ (2.5)

B 는 후진연산자(backward shift operator)이며 d 는 차분 차수를 의미한다. ϵ_t 는 $WN(0, \sigma^2)$ 를 따르는 오차항이다. 여기서 $\phi(B)$ 와 $\theta(B)$ 는 각각 모형의 차수가 p, q 인 AR, MA에 관한 다항식으로 표현되며 식은 다음과 같다.

$$\begin{aligned}\theta_q(B) &= 1 - \theta_1 B - \dots - \theta_q B^q, \\ \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p B^p.\end{aligned}\quad (2.6)$$

3. 예측모형의 성능비교

3.1. 데이터 소개

본 연구에 적용한 데이터는 2015년 1월 1일부터 2018년 1월 7일까지 1시간 단위로 기상포털자료에서 제공되는 기상자료이다. 후에 태양광 예측에 사용하기 위해 태양광 발전량 패널이 존재하는 광주, 청주 2개 지역을 선정하여 분석에 이용하였으며, 2015년 1월 1일부터 2017년 12월 31일까지의 자료를 훈련용 데이터(training data)로 이용하여 모형을 적합에 사용하였다. 그리고 나머지 2018년 1월 1일부터 1월 7일까지 자료를 테스트 데이터(test data)로 모형의 성능을 평가하는데 활용하였다. 사용된 독립변수들은 기상청에서 제공하는 기상자료개방포털을 통해 기상관측소에서 관측된 자료를 이용하여, 기온, 습도, 운량, 적설량 등 8개의 기상변수와 Kim (2017)에서 제안한 대기권 밖 일사량을 모형에 활용했다. 분석을 진행하기 위해 이 절의 앞부분에서 설명한 방법대로 2개 지역의 일사량 자료를 재구성하였다. Figure 3.1은 2015년 한 해 동안의 서울 일사량 자료를 나타낸 시계열 도표이며, Figure 3.2는 일사량

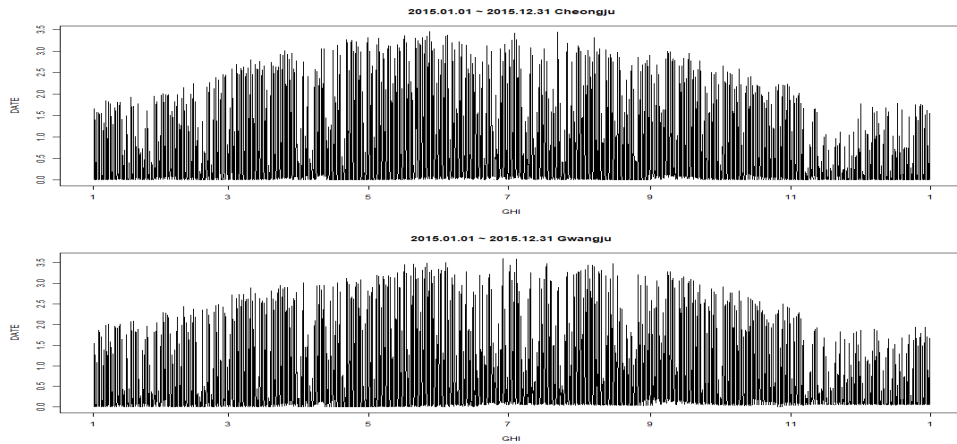


Figure 3.1. Hourly solar radiation of Cheongju and Gwangju in 2015.

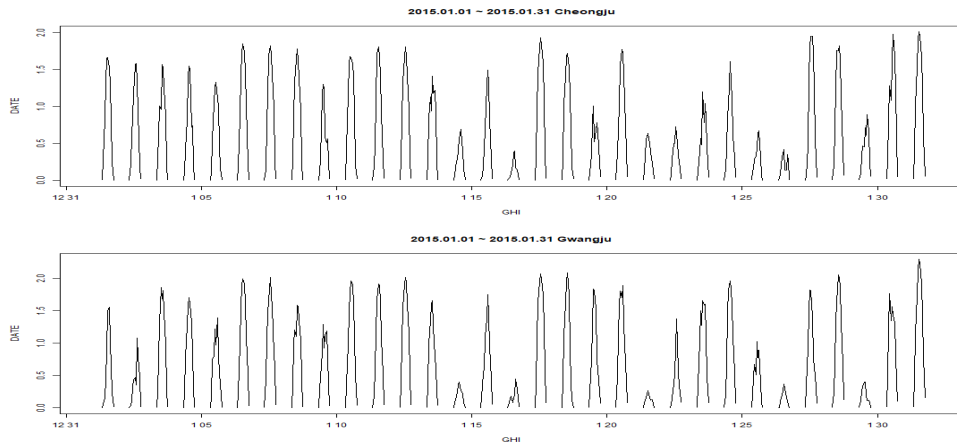


Figure 3.2. Hourly solar radiation of Cheongju and Gwangju in January 2015.

자료의 패턴을 좀 더 세밀하게 살펴보기 위해 2015년 1월의 자료만을 이용한 시계열 도표이다. 자료 중 0의 값을 가지는 부분이 다수 존재하며 일출 후부터 발전이 시작되어 최고점을 형성한 후 일몰 후에는 다시 0의 값을 갖는 패턴이 하루마다 반복된다. 이러한 특징으로 시간별 일사량 예측 모형을 적합할 경우 모형 적합이 정확하게 이루어지지 않으므로, 본 논문에서는 문제를 해결하기 위하여 연속된 일사량 자료를 일몰과 일출 시간을 이용하여 일정 시간을 분리하여 예측 모형에 이용하였다. 이러한 데이터를 활용하여, ARIMA, ARIMAX, seasonal ARIMA, seasonal ARIMAX, ARIMA-GARCH, ARIMAX-GARCH, seasonal ARIMA-GARCH, seasonal ARIMAX-GARCH의 시계열 모형을 이용하여 적합하고 예측을 진행하여 모형 간 비교하였다.

3.2. 모형 적합 결과

일사량 자료는 일출과 일몰시간에 의해 하루 중 일정한 시간동안 0의 값을 갖는다. 그러므로 일출과 일몰시간을 고려하여 07시부터 21시까지의 시간만을 사용하여 예측하도록 한다. 일사량 예측을 위한 과

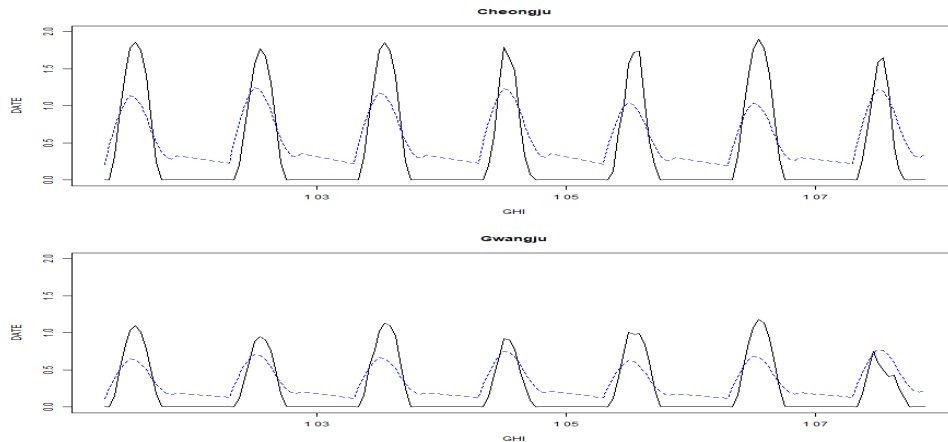


Figure 3.3. Result of forecasting 7-Days ahead, ARIMA model.

정은 다음과 같다. 먼저 전체 데이터 중 2015년 1월 1일부터 2017년 12월 31일까지 07시부터 21시까지의 데이터를 사용하여 예측 모형을 수립한다. 예측 모형이 수립되면 모형 수립에 사용되지 않은 첫 번째 날인 2018년 1월 1일의 07시부터 21시까지의 일사량을 예측한다. 이 예측값을 추가하여 새로 모형을 수립한 뒤 다음 날인 1월 2일의 07시부터 21시까지의 일사량을 예측한다. 이러한 과정을 반복한 뒤 실제 일사량 관측값과 일사량 예측값을 비교하여 모형의 정확도를 비교하였다. ARIMA, ARIMAX, seasonal ARIMA, seasonal ARIMAX, ARIMA-GARCH, ARIMAX-GARCH, seasonal ARIMA-GARCH, seasonal ARIMAX-GARCH 모형은 Akaike's information criterion (AIC)를 기준으로 AIC 값이 가장 작은 모형을 적합하였다.

R의 `auto.arima` 함수를 이용해 각각의 지역의 자료를 ARIMA 모형으로 적합한 결과 AIC기준으로 청주와 광주 모두 ARIMA(5, 1, 1) 모형으로 적합되었다. Figure 3.3은 적합된 모형의 예측 결과를 보기 위해 2018년 1월 1일부터 2018년 1월 7일까지 7일간의 예측을 진행한 결과이다. 실선이 관측값이며, 파란 점선이 모형에 의한 예측값이다. 예측된 그래프들은 하루 중 07시부터 21시까지의 값이며, 1월 그래프다보니 일사 관측시간이 다른 날보다 짧게 나타났다.

두 지역에서 일사량의 변동을 민감하게 예측하지는 못하지만 일사량의 전체적인 ARIMA를 이용한 예측 결과 일사량의 전체적인 경향은 잘 맞추는 것으로 보인다. 하지만, 하루의 최대값 근방에서의 예측 성능이 우수하지 않으며, 이러한 예측 오차를 줄이기 위하여 다음은 기상변수를 고려한 ARIMAX를 이용한 결과이다. ARIMAX 모형을 적용하기 위해서는 적절한 설명변수가 선정되기 위해 일사량을 제외한 10개의 변수와 일사량과의 상관관계를 확인한 뒤에 상관관계가 높은 변수들을 설명변수로 사용한다. Figure 3.4는 8개의 변수와 일사량 간의 상관계수를 도식화한 그림이다. 시간대별, 월별 모두 기온, 습도, 운량 그리고 대기권 밖 일사량이 일사량과 높은 상관관계를 가지고 있는 것을 확인할 수 있다. 상관관계가 높은 기온, 습도, 운량(4분위) 그리고 대기권 밖 일사량 4가지 변수를 사용해 ARIMAX 모형을 적용하였다.

Table 3.2는 4개 변수들의 모든 조합에 대해 강릉 한 지역을 이용하여 ARIMAX 모형을 적용한 결과를 요약한 표이다. 모든 조합에 대해 MAE를 비교해보았을 때 청주와 광주 지역 모두 대기권 밖 일사량과 기온, 습도, 운량(4분위) 4개를 설명변수로 사용했을 때 가장 예측력이 좋다는 것을 확인할 수 있다. 따라서, 대기권 밖 일사량과 기온, 습도, 운량(4분위), 네 변수를 선택한 후 각각의 지역의 자료를

Table 3.1. Fitted models using ARIMA

Region	Model
Cheongju	ARIMA(5, 1, 1)
Gwangju	ARIMA(5, 1, 1)

ARIMA = auto-regressive integrated moving average.

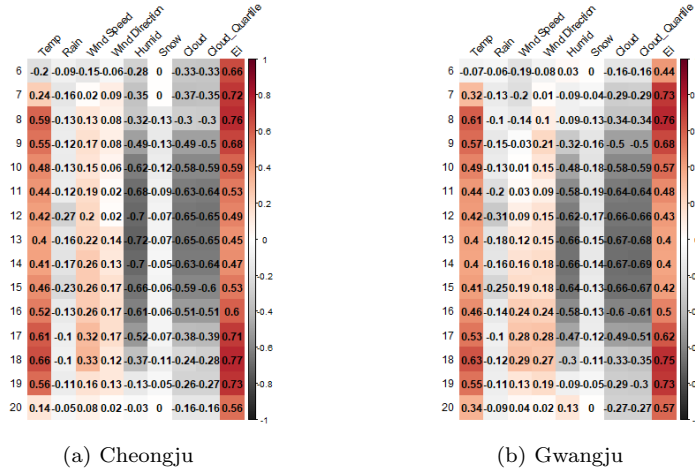


Figure 3.4. Correlation of Cheongju and Gwangju.

Table 3.2. Result of variable selection using ARIMAX model

Variable selection	Cheongju	Gwangju
ei	0.1515	0.1791
ei + temp	0.1866	0.2186
ei + humid	0.1929	0.2037
ei + cloud4	0.1357	0.1755
ei + temp + humid	0.1965	0.2185
ei + temp + cloud4	0.1725	0.1824
ei + humid + cloud4	0.1742	0.1958
ei + temp + humid + cloud4	0.1238	0.1731

ARIMAX = auto-regressive integrated moving average with eXogenous variable.

Table 3.3. Fitted models using ARIMAX

Region	Model
Cheongju	ARIMAX(2, 1, 2)
Gwangju	ARIMAX(2, 1, 5)

ARIMAX = auto-regressive integrated moving average with eXogenous variable.

ARIMAX 모형으로 적합한 결과 AIC기준으로 청주의 최적 모형은 ARIMAX(2, 1, 2)이고 광주의 최적 모형은 ARIMAX(2, 1, 5)이다. 이 두 가지 변수를 사용해 ARIMAX 모형에 적합한 후 예측한 결과는 Figure 3.5와 같다. 실선이 관측값이며, 파란 점선이 모형에 의한 예측값이다. ARIMAX의 결과, 일사량만 이용한 것보다 다른 외생변수를 고려했을때 좀 더 실제값에 가까이 나타나는 것으로 훨씬 나은 성능이 나타나는 것으로 보인다.

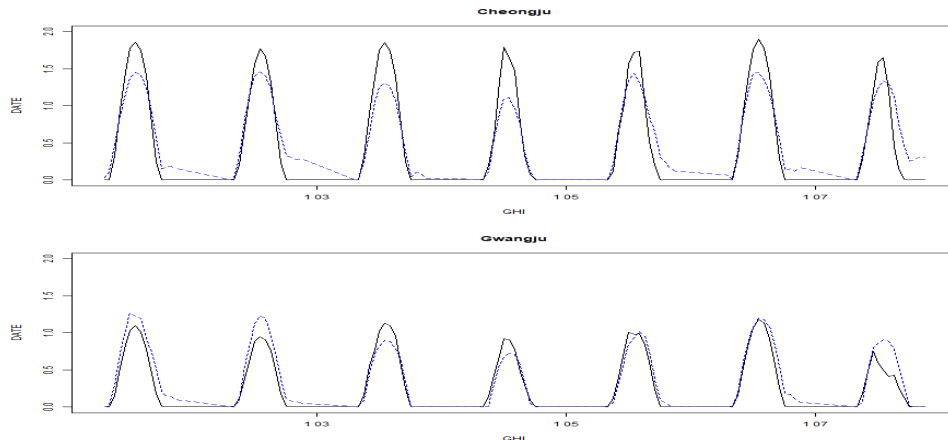


Figure 3.5. Result of forecasting 7-Days ahead, ARIMAX model. ARIMAX = auto-regressive integrated moving average with eXogenous variable.

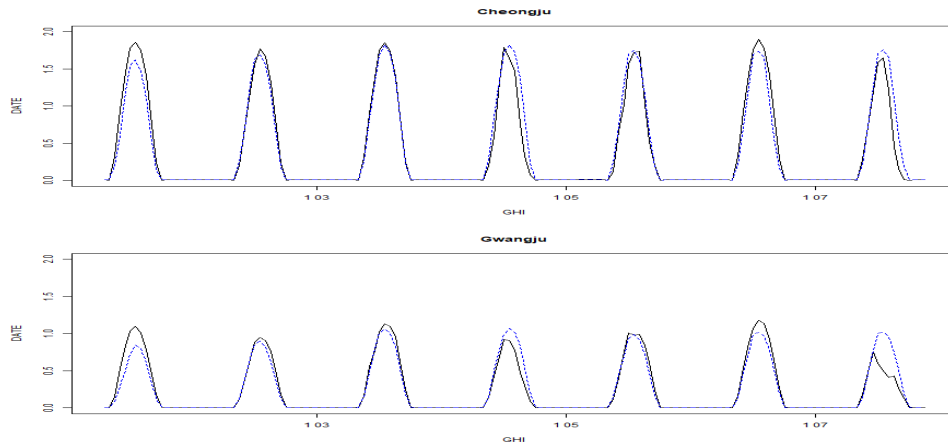


Figure 3.6. Result of Forecasting 7-Days ahead, seasonal ARIMA model. ARIMA = auto-regressive integrated moving average.

Table 3.4. Fitted models using seasonal ARIMA

Region	Model
Cheongju	ARIMA(2, 0, 1)(2, 1, 0)[15]
Gwangju	ARIMA(2, 0, 4)(2, 1, 0)[15]

ARIMA = auto-regressive integrated moving average.

Table 3.4는 Seasonal ARIMA 결과이다. Seasonal ARIMA 모형으로 적합한 결과 AIC기준으로 청주는 ARIMA(2, 0, 1)(2, 1, 0)[15]와 광주는 ARIMA(2, 0, 4)(2, 1, 0)[15] 모형으로 적합되었다. Figure 3.6는 적합한 모형의 예측 결과이다. 일사량은 주기성을 지니고 있으므로, ARIMA 모형이나 ARIMAX 모형보다 예측 성능이 우수하게 나타난다.

Table 3.5는 seasonal ARIMAX 결과이다. Seasonal ARIMAX 모형으로 적합한 결과 AIC기준으로 청

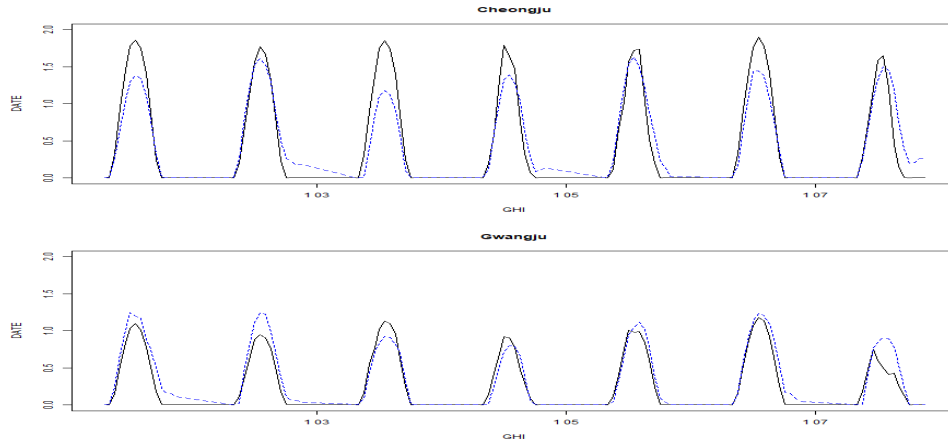


Figure 3.7. Result of forecasting 7-Days ahead, seasonal ARIMAX model. ARIMAX = auto-regressive integrated moving average with eXogenous variable.

Table 3.5. Fitted models using seasonal ARIMAX

Region	Model
Cheongju	ARIMAX(5, 1, 0)(2, 0, 0)[15]
Gwangju	ARIMAX(0, 1, 5)(2, 0, 0)[15]

ARIMAX = auto-regressive integrated moving average with eXogenous variable.

주는 ARIMAX(5, 1, 0)(2, 0, 0)[15]와 광주는 ARIMAX(0, 1, 5)(2, 0, 0)[15] 모형으로 적합되었다. Figure 3.7은 적합된 모형의 예측 결과이다. 실선이 관측값이며, 파란 점선이 모형에 의한 예측값이다. Seasonal ARIMAX의 결과, 일사량만 이용한 것보다 다른 외생변수를 고려했을때 좀 더 실제값에 가까이 나타나는 것으로 훨씬 나은 성능이 나타나는 것으로 보인다. 하지만 seasonal ARIMA 보다 과소 추정 되는 것으로 보인다.

일사량 자료를 ARIMA-GARCH로 적합한 결과 ARIMA(5, 1, 1) × GARCH(1, 1) 모형이 적합 되었다. 적합된 모형의 예측 경향을 알아보기 위해 2018년 01월 01일 이후 7일을 예측하였다. Figure 3.8과 같이 ARIMA-GARCH 모형은 등분산성이 아닌 이분산성을 가정을 하므로 일사량에서 생기는 분산을 고려하기 때문에 ARIMA 모형보다 향상된 성능을 나타낸다.

ARIMAX-GARCH 모형은 ARIMAX 모형에 이분산 문제를 고려한 GARCH 모형을 추가한 모형이다. ARIMAX-GARCH 모형도 ARIMA-GARCH 모형과 마찬가지로 ‘ugarchspec’, ‘ugarchfit’ 함수를 이용하여 모형을 적합시킨 후 2018년 01월 01일 이후 7일을 예측하였다. 또한 ARIMAX와 동일하게 앞서 선정된 기상변수와 대기권밖 일사량 변수를 활용하여 모형을 적합시켰다. 그 결과 청주의 ARIMX-GARCH 모형의 최적 모형은 ARIMAX(2, 1, 2)-GARCH(1, 1)이고 광주의 최적모형은 ARIMAX(2, 1, 5)-GARCH(1, 1)이다. 두 지역 모두 대기권 밖 일사량 변수와 기온, 습도, 운량 3가지 기상변수를 외생변수로 고려하였다. 청주와 광주의 최적 모형을 사용하여 예측한 결과는 Figure 3.9와 같다.

Seasonal ARIMA-GARCH 모형은 seasonal ARIMA 모형에 이분산 문제를 고려한 GARCH 모형을 추가한 모형이다. 모형을 적합시킨 결과 청주의 seasonal ARIMA-GARCH 모형의 최적 모형은 ARIMA(2, 0, 1)(2, 1, 0)[15]-GARCH(1, 1)이고 광주의 최적모형은 ARIMA(2, 0, 4)(2, 1, 0)[15]-

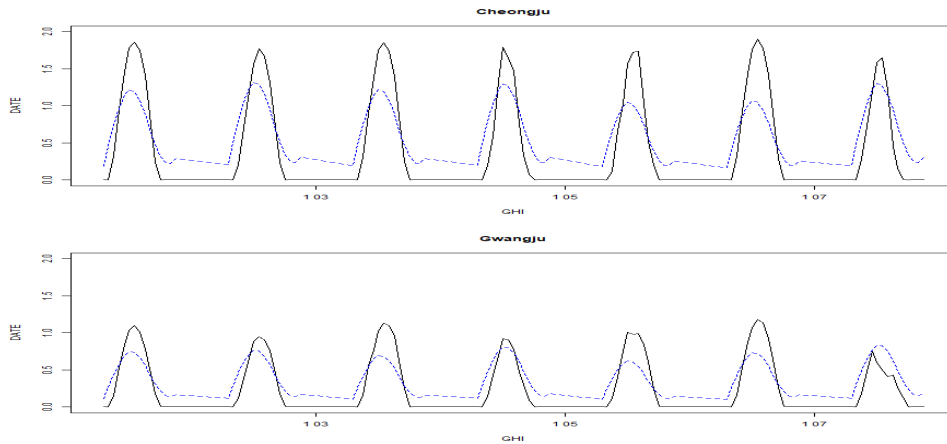


Figure 3.8. Result of forecasting 7-Days ahead, ARIMA-GARCH model. ARIMA = auto-regressive integrated moving average; GARCH = generalized auto-regressive conditionally heteroscedastic.

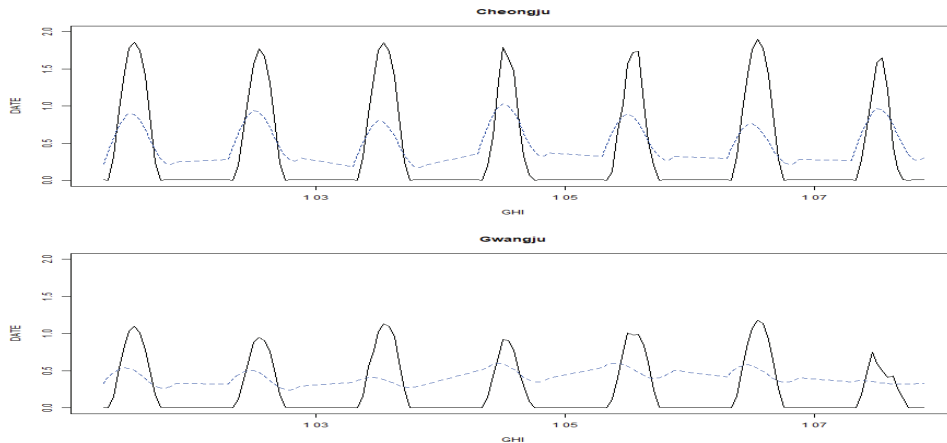


Figure 3.9. Result of forecasting 7-Days ahead, ARIMAX-GARCH model. ARIMAX = auto-regressive integrated moving average with eXogenous variable; GARCH = generalized auto-regressive conditionally heteroscedastic.

GARCH(1,1)이다. 청주와 광주의 최적 모형을 사용하여 예측한 결과는 Figure 3.10과 같다.

Seasonal ARIMAX-GARCH 모형은 seasonal ARIMAX 모형에 이분산 문제를 고려한 GARCH 모형을 추가한 모형이다. 모형을 적합시킨 결과 청주의 seasonal ARIMAX-GARCH 모형의 최적 모형은 ARIMAX(5, 1, 0)(2, 0, 0)[15]-GARCH(1, 1)이고 광주의 최적모형은 ARIMAX(0, 1, 5)(2, 0, 0)[15]-GARCH(1, 1)이다. 청주와 광주의 최적 모형을 사용하여 예측한 결과는 Figure 3.11과 같다. Seasonal ARIMAX와 비교해서 변동성을 더 정확하게 예측하는 것으로 나타난다.

3.3. 모형의 성능 비교

앞의 절에서 적합한 모형들을 비교하기 위한 오차의 척도로 MAE와 RMSE를 이용하였다. 일반적으로 전력수요예측에서는 mean absolute percentage error (MAPE)를 많이 사용하지만 태양광 발전의 경우

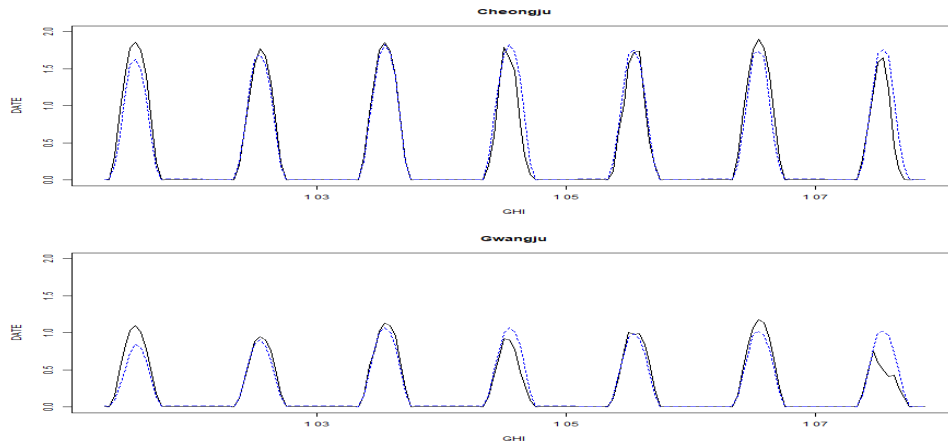


Figure 3.10. Result of forecasting 7-Days ahead, seasonal ARIMA-GARCH model. ARIMA = auto-regressive integrated moving average; GARCH = generalized auto-regressive conditionally heteroscedastic.

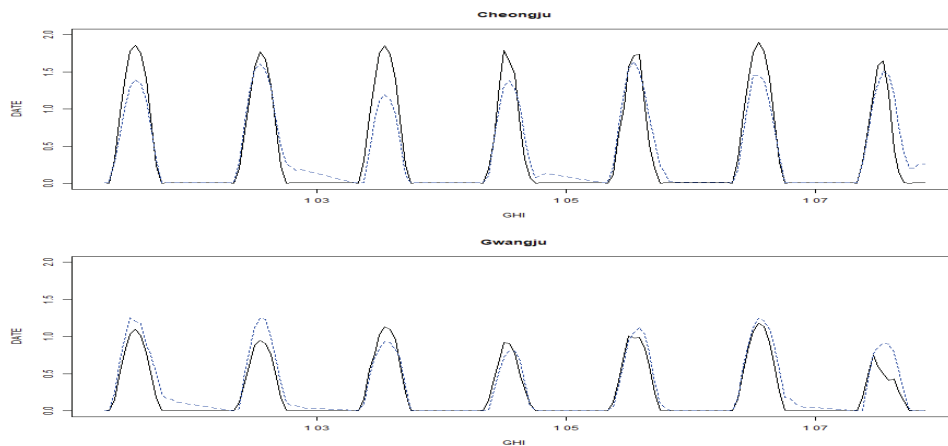


Figure 3.11. Result of forecasting 7-Days ahead, seasonal ARIMAX-GARCH model. ARIMAX = auto-regressive integrated moving average with eXogenous variable; GARCH = generalized auto-regressive conditionally heteroscedastic.

자료가 0인 경우가 많아 MAPE를 계산할 수 없는 경우가 많다. MAE는 실제 관측값과 예측값의 차이의 절대값을 전체 자료의 개수로 나눈 척도로 아래 식과 같이 정의된다.

$$\text{MAE} = \frac{\sum_{t=1}^m |Y_t - \hat{Y}_t|}{m} \quad (3.1)$$

MSE는 실제 관측값과 예측값의 차이의 제곱합 평균으로 정의된다. RMSE는 MSE의 제곱근이며 아래 식과 같이 정의된다.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^m (Y_t - \hat{Y}_t)^2}{m}} \quad (3.2)$$

Table 3.6. Model comparison

Region	Model	RMSE	MAE
Cheongju	ARIMA	0.4173	0.3736
	ARIMAX	0.2586	0.1937
	Seasonal ARIMA	0.1669	0.0977
	Seasonal ARIMAX	0.2572	0.1824
	ARIMA-GARCH	0.3889	0.3406
	ARIMAX-GARCH	0.5007	0.4222
	Seasonal ARIMA-GARCH	0.1666	0.0973
	Seasonal ARIMAX-GARCH	0.2543	0.1803
Gwangju	ARIMA	0.2436	0.2162
	ARIMAX	0.1509	0.1092
	Seasonal ARIMA	0.1348	0.0759
	Seasonal ARIMAX	0.1480	0.1073
	ARIMA-GARCH	0.2279	0.1994
	ARIMAX-GARCH	0.3629	0.3262
	Seasonal ARIMA-GARCH	0.1345	0.0750
	Seasonal ARIMAX-GARCH	0.1509	0.1097

RMSE = root mean square error; MAE = mean absolute error; ARIMA = auto-regressive integrated moving average; ARIMAX = ARIMA with exogenous variable; GARCH = generalized auto-regressive conditionally heteroscedastic.

위의 두 척도 모두 MAPE와 달리 자료의 크기에 영향을 받는 척도이며, 작을수록 모형의 성능이 우수하다는 것을 의미한다. 모형의 비교결과는 Table 3.6, Table 3.7와 같다.

모형 비교 결과, 청주의 MAE 기준으로 ARIMA-GARCH 모형의 MAE가 0.3406이고 ARIMA 모형의 MAE가 0.3736로 약 0.03 정도의 차이가 나타남을 알 수 있다. 분산의 이분산 모형을 가정하였을 때, 기존의 모형보다 예측 정확도를 더 향상시킴을 확인할 수 있었다. 모형별 비교에서는 기상변수와 대기권 밖 일사량을 고려한 ARIMAX 모형의 MAE가 0.1937로 ARIMA 모형보다 모형의 성능을 높였다. 가장 성능이 좋은 모형은 MAE가 0.0977인 seasonal ARIMA 모형이다. 반면 외생변수를 고려한 seasonal ARIMAX는 seasonal ARIMA에 비해 모형의 정확도가 떨어지는 것으로 나타났다. seasonal ARIMA 모형보다 Seasonal ARIMA-GARCH 모형의 예측오차가 개선됨을 보였지만, 차이는 매우 작았으며 시간이 오래걸린다는 점에서 seasonal ARIMA 모형이 더 적합한 모형으로 볼 수 있다. 광주도 동일하게 ARIMA 모형 보다 ARIMA-GARCH 모형의 예측 성능이 우수하였고, seasonal ARIMA-GARCH 모형의 예측 오차가 가장 낮았다. 하지만, ARIMAX 모형에서 분산의 이분산 모형을 가정한 ARIMAX-GARCH 모형의 MAE가 청주에서는 0.3164 광주는 0.5390으로 오히려 모형의 성능을 저하시켜 이분산 문제를 보완하지 않았을 때 더 좋은 예측 성능을 보인다.

일사량의 경우 자료의 특성상 일출과 일몰 시간대에 자료값이 0 가깝이에 나타나 MAPE를 계산 할 수 없는 경우가 있어, 10시부터 16시까지의 피크 시간대만 고려하여 자료를 이용하여 계산하였다. 그 결과는 Table 3.7과 같다. 청주의 경우 평균적으로 seasonal ARIMA 모형의 성능이 가장 우수하게 나왔지만 2018년 1월 1일이나 7일과 같이 MAPE가 크게 나오는 경향이 보이는데 그 원인 중에 하나는 기상 요인으로 인하여 일사량 출력이 감소하면서 편차가 크게 발생했다. 광주의 경우는 ARIMAX 모형이 다른 모형들에 비해 실제값과 편차가 작게 나타났다. 하지만 흐린 날씨에는 여전히 예측 성능을 저하시킬 수 있다.

Table 3.7. Model comparison of peak hour using mean absolute percentage error

Region	Model	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Average
Cheongju	ARIMA	47.2650	29.4263	40.9100	34.8224	41.0737	62.9476	34.0887	41.5048
	ARIMAX	19.8567	11.4872	33.2028	33.1779	18.3497	18.5958	30.4866	23.5938
	Seasonal ARIMA	27.6952	7.7431	3.2290	24.0290	11.1588	22.7771	25.6299	17.4660
	Seasonal ARIMAX	32.6295	7.5297	65.3726	24.8055	15.5418	28.0608	29.6922	29.0903
	ARIMA-GARCH	16.5643	17.0336	17.2961	27.4833	27.6187	22.9549	34.4632	23.3449
	ARIMAX-GARCH	34.2862	26.9510	42.7261	33.3060	33.3026	53.7572	33.2845	36.8019
	Seasonal ARIMA-GARCH	6.8847	20.2207	18.6048	37.4787	29.7156	8.1529	41.3636	23.2030
	Seasonal ARIMAX-GARCH	9.5727	18.6762	21.4747	22.7059	25.2635	6.1135	32.3176	19.4463
Gwangju	ARIMA	46.4582	27.4027	53.0787	20.6318	55.7974	52.9428	28.2293	40.6487
	ARIMAX	19.9435	24.5671	19.5145	23.2756	16.0069	9.2008	33.2434	20.8217
	Seasonal ARIMA	43.8726	10.9220	9.5644	22.3077	11.7917	16.6670	34.9274	21.4361
	Seasonal ARIMAX	17.6462	24.2292	18.0499	27.8457	14.0420	10.4230	34.3397	20.9394
	ARIMA-GARCH	14.6097	23.1338	16.7562	35.2435	16.4209	14.9719	50.7543	24.5558
	ARIMAX-GARCH	26.2592	23.0909	42.3104	26.0282	24.2462	28.5083	25.0416	27.9264
	Seasonal ARIMA-GARCH	11.0332	26.4734	21.9058	43.3024	8.6714	3.2907	43.0287	22.5294
	Seasonal ARIMAX-GARCH	18.1922	24.7870	17.2519	26.8925	13.6975	11.0473	34.5419	20.9158

ARIMA = auto-regressive integrated moving average; ARIMAX = ARIMA with eXogenous variable; GARCH = generalized auto-regressive conditionally heteroscedastic.

4. 결론

본 연구는 태양광 발전량 예측과 관련이 높은 일사량에 대하여 시계열 모형을 이용하여 예측을 실시하였다. 기존 연구에서는 일사량 예측에 대하여 시계열 모형의 기본적인 ARMA 모형이나 Neural Network 모형을 위주로 발전되어 왔다. 고려된 독립변수 또한 기온과 습도만을 고려하였다. 본 연구에서는 모든 기상변수를 일사량과 고려하여 가장 영향력 있는 변수들을 선택하여 모형의 성능이 가장 좋게 나타난 변수를 외생변수로 고려하였다. 일사량과 기상변수 및 대기권 밖 일사량의 상관분석 결과, 대기권 밖 일사량, 기온, 습도, 구름 4분위가 높은 상관관계를 갖는 것을 확인할 수 있었다.

시계열 자료의 분석과 예측에 있어서 많이 이용되는 ARIMA 모형, 외생변수를 고려한 ARIMAX 모형, 주기성을 고려한 seasonal ARIMA 모형, 주기성과 외생변수를 고려한 seasonal ARIMAX 모형을 이용하였다. 추가적으로 이분산 문제를 보완한 GARCH 모형을 활용한 ARIMA-GARCH 모형, 기상변수를 활용한 ARIMAX-GARCH 모형, seasonal ARIMA-GARCH 모형, 그리고 seasonal ARIMAX 모형 총 8개의 모형을 사용하여 청주와 광주의 시간별 일사량에 대한 예측을 시행하였다. 청주 지역을 예측한 경우 계절성과 외생변수를 고려한 seasonal ARIMA-GARCH 모형의 성능이 더 좋은 예측 결과를 나타냈으며, 광주 지역도 동일하게 seasonal ARIMA-GARCH 모형의 성능이 우수하게 나타났다. 이때, 독립변수를 고려하지 않았을 때는 seasonal ARIMA-GARCH 모형이 성능이 우수한 결과를 나타남을 확인할 수 있었다. 반면에 기상 데이터 및 대기권 밖 일사량을 고려한 경우에는 이분산 문제를 고려하지 않았을 때 향상된 예측 성능이 도출되었다. 따라서 기상변수를 고려했을 때 더 좋은 예측력을 얻을 수 있지만 이분산 문제를 고려했을 때는 성능이 하락하는 결과를 알 수 있었다. 향후 본 연구에 대한 확장으로 일사량에 영향을 미치는 기상변수 외의 다른 변수들을 탐색하여 흐린 날과 맑은 날의 구분하여 일사량 예측에 고려하는 방안이 필요하다고 생각한다.

References

- Colak, I., Yesilbudak, M., Genc, N., and Bayindir, R. (2015). Multi-period prediction of solar radiation using ARMA and ARIMA models, *In Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on* (pp. 1045-1049). IEEE.

- Elizondo, D., Hoogenboom, G., and McClendon, R. W. (1994). Development of a neural network model to predict daily solar radiation, *Agricultural and Forest Meteorology*, **71**, 115–132.
- Grek, A. (2014). Forecasting accuracy for ARCH models and GARCH (1, 1) family: Which model does best capture the volatility of the Swedish stock market?.
- Ji, W. and Chee, K. C. (2011). Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN, *Solar Energy*, **85**, 808–817.
- Kim, S. (2017). A study on solar irradiance forecasting with weather variables. *The Korean Journal of Applied Statistics*, **30(6)**, 1005-1013.
- Kim, S., Lee, K., and Son, H. (2018). A study on solar energy forecasting based on time series models. *The Korean Journal of Applied Statistics*, **31(1)**, 139-153.
- Kim, S. and Park S. (2016). A study on short-term wind power forecasting using time series models, *The Korean Journal of Applied Statistics*, **29**, 1373–1383.
- Liu, H., Erdem, E., and Shi, J. (2011). Comprehensive evaluation of ARMA-GARCH (-M) approaches for modeling the mean and volatility of wind speed. *Applied Energy*, **88**, 724–732.
- Rehman, S. and Mohandes, M. (2008). Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy*, **36**, 571–576.
- Sun, H., Yan, D., Zhao, N., and Zhou, J. (2015). Empirical investigation on modeling solar radiation series with ARMA-GARCH models. *Energy Conversion and Management*, **92**, 385-395.

시계열 모형을 활용한 일사량 예측 연구

서유민^a · 손흥구^b · 김삼용^{a,1}

^a중앙대학교 응용통계학과, ^b한국전력거래소

(2018년 10월 25일 접수, 2018년 12월 2일 수정, 2018년 12월 11일 채택)

요약

신재생에너지 산업이 발전함에 따라 태양광 발전에 대한 중요성이 확대되고 있다. 태양광 발전량을 정확히 예측하기 위해서는 일사량 예측이 필수적이다. 본 논문에서는 태양광 패널이 존재하는 청주와 광주 지역을 선정하여 기상포털에서 제공하는 시간별 기상 데이터를 수집하여 연구하였다. 일사량 예측을 위하여 시계열 모형인 ARIMA, ARIMAX, seasonal ARIMA, seasonal ARIMAX, ARIMA-GARCH, ARIMAX-GARCH, seasonal ARIMA-GARCH, seasonal ARIMAX-GARCH 모형을 비교하였다. 본 연구에서는 모형의 예측 성능을 비교하고자 mean absolute error와 root mean square error를 사용하였다. 모형들의 예측 성능 비교 결과 일사량만 고려하였을 때는 이분산 문제를 고려한 seasonal ARIMA-GARCH 모형이 우수한 성능을 나타냈고, 외생변수를 활용한 ARIMAX 모형으로 일사량 예측을 한 경우가 가장 좋은 예측력을 나타냈다.

주요용어: ARIMA, ARIMAX, seasonal ARIMA, GARCH, 기상변수, 일사량

이 논문은 2018년도 대학원생지원장학금의 지원에 의해 작성되었음.

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