

## A study on solar energy forecasting based on time series models

Keunho Lee<sup>a</sup> · Heung-gu Son<sup>b</sup> · Sahm Kim<sup>a,1</sup>

<sup>a</sup>Department of Applied Statistics, Chung-Ang University;

<sup>b</sup>Department of Aviation, The Korea Transport Institute

(Received November 27, 2017; Revised December 11, 2017; Accepted December 11, 2017)

---

### Abstract

This paper investigates solar power forecasting based on several time series models. First, we consider weather variables that influence forecasting procedures as well as compare forecasting accuracies between time series models such as ARIMAX, Holt-Winters and Artificial Neural Network (ANN) models. The results show that ten models forecasting 24hour data have better performance than single models for 24 hours

Keywords: ARIMAX, Artificial Neural Network, weather variables, solar power

---

### 1. 서론

현재 한국의 에너지 소비형태를 살펴보면 전체 에너지수요량의 97%를 해외에 의존하며 총 에너지 소비의 약 85%를 화석연료가 차지하고 있다. 한국의 이러한 화석연료 기반 에너지공급 시스템은 국내 부존 자원의 부재, 해외 에너지 시장의 높은 변동성 및 지정학적 불안정으로 지속가능성이 낮다. 또한 전세계 화석연료의 매장량이 2030년 이후에는 감소할 것으로 예상되며 에너지 수출 국가들이 자국의 에너지 안보 및 경제성 취득을 위해 자원을 독점할 가능성도 배제할 수 없다 (Kim과 Kim, 2016).

최근 이에 대한 해결책으로 신재생에너지 분야가 대안으로 떠오르고 있다. 특히 태양광에너지 분야는 신재생에너지 분야의 핵심 중 하나로, 현재 보급되고 있는 신재생에너지원 중 태양광발전의 발전원가가 가장 비싼 것이 사실이나 기술발전을 통해 매년 낮아지고 있으며, 화석연료의 단가가 인상되고 있어 수년 이내에 화석연료를 재료로 하는 발전원가와 태양광발전원가가 같아지는 그리드 패리티(grid Parity)가 도래할 것으로 전망된다. 설치와 유지보수가 용이하고, 발전과정에 소음이나 오염물질을 배출하지 않아 일반 가정이나 상업시설에도 쉽게 적용이 가능하다. 전체 태양 복사에너지는 전세계 에너지 소비량의 10,000배 이다. 전세계가 소비하는 에너지량은 해가 갈수록 증가하고 있다. 그러나 화석연료는 매장량에 한계가 있으며 자원 채취에 더 많은 비용을 소모해야 하므로 에너지단가는 지속적으로 증가하고 있다. 이러한 흐름은 우리가 왜 신재생 에너지를 선택해야 하는지 분명하게 보여주고 있다. 태양광

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B01014954).

<sup>1</sup>Corresponding author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: [sahm@cau.ac.kr](mailto:sahm@cau.ac.kr)

에너지는 자원의 무한성, 친환경적 에너지전환 및 사용이 가능하다는 장점을 가진다. 하지만 태양광 에너지 자원은 날씨나 기후변화에 영향을 받는 에너지공급의 변동성으로 인하여 다른 자원들과는 다르게 안정적인 에너지 수급이 불가능하다. 따라서 높은 에너지 생산비용, 자원의 변동성 및 지역적 제한성 등 에너지 공급 시스템의 단점을 보완하기 위한 정확한 수요예측이 필요하다. 이러한 정확한 수요예측은 안정적 에너지공급 위해서이며 이를 위해서는 최적의 수요예측 모형을 구상해야 할 것이다.

전력 예측 모형의 경우 크게 장기 예측과 단기 예측으로 나뉘지는데, 태양광에너지의 경우 날씨나 기후변화에 에너지 공급에 많은 영향을 받기 때문에 날씨의 정확한 예측을 위해 단기 예측모델이 장기 예측보다 좋을 것으로 판단하였다. 따라서 본 논문에서는 단기예측 방법에 관해 소개하고자 한다. 다음 2장에서는 태양광 발전량 예측을 위한 신경망 모형과, 시계열 모형을 소개한다. 3장에서는 본 논문에서 제안하는 방법인 연속된 시간자료를 각 시간대 별로 분리하여 재구성하여 24개의 (1시-24시) 일별 자료 형태로 예측에 활용하는 방법을 소개하고 기존방법과 제안된 방법의 성능차이를 비교하기 위해 ARIMAX 모형, 신경망 모형(neural network model)을 이용하여 동일한 모형 및 변수를 가지는 환경에서의 성능차이를 확인할 것이다. 4장에서는 결론과 향후 연구 방향에 대하여 제안할 것이다.

## 2. 태양광 발전량 예측 모형

### 2.1. Holt-Winters 방법

Holt-Winters 방법은 Holt (1957)와 Winters (1960)에 의해 발전된 계절성을 고려할 수 있는 지수평활법의 확장된 형태로 가법형(additive)과 승법형(multiplicative) 두가지 방법이 있다. 본 논문에서는 예측을 위해 가법형을 사용하였다. Holt-Winters 가법형은 예측을 위한 예측식과 수준( $l_t$ ), 추세( $b_t$ ), 계절성( $s_t$ )에 대한 세가지 평활식으로 이루어져 있다. 예측식과 각각의 평활식은 다음과 같다.

$$\begin{aligned}\hat{y}_{t+h|t} &= l_t + hb_t + s_{t-m+h_m^+} \\ l_t &= \alpha(y_t - st - m) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \\ s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}, \\ &\text{(where } h_m^+ = [(h - 1) \bmod m] + 1)\end{aligned}$$

이때,  $\alpha, \beta^*, \gamma$ 는 평활 모수이며,  $m$ 은 계절주기,  $h$ 는 예측하고자하는 미래 시점까지의 시간 간격이다.

### 2.2. ARIMAX 모형

ARIMAX 모형은 auto-regressive intergrated moving average model (ARIMA) 모형에 외생변수를 추가한 모형이다. 평균이  $\mu$ 이고 ARIMA의 차수가  $p, d, q$ 일 때 외생변수  $x_{ij}$ 가  $k$ 개인 경우 ARIMAX 모형의 식은 다음과 같다.

$$\begin{aligned}(1 - B^d)(Y_t - \mu) &= \frac{\theta(B)}{\phi(B)}\varepsilon_t + \sum_{i=1}^k \beta_{it}x_{it}, \\ \text{where } \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q.\end{aligned}$$

이때,  $\phi(B)$ 는 auto-regressive (AR)에 관한 다항식이며,  $\theta(B)$ 는 moving-average (MA)에 관한 다항식이다.  $B$ 는 후진연산자(backward shift operator)이다.  $Y_t$ 는  $t$ 시점 ( $t = 1, 2, \dots, T$ )에서의 관측된 값

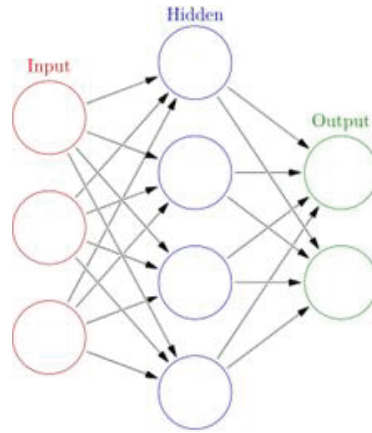


Figure 2.1. Artificial neural network.

이며  $p, d, q$ 는 정수이다.  $\varepsilon_t$ 의 평균은 0, 분산은 일정한 백색잡음과정(white noise process)이다. 그리고  $\beta_{it}$ 는 외생변수  $x_{it}$ 의 계수이다. 본 논문에서는 외생변수로 기온, 습도, 현지기압, 해면기압, 이슬점 온도, 일조량, 일사량을 고려하였다.

### 2.3. 신경망 모형

인공신경망(artificial neural network)은 태양 에너지 예측 분야에서 일반적으로 자주 사용되는 모형으로, 생물학의 신경망에서 영감을 얻은 통계학적 학습 알고리즘이다. 시냅스의 연결로 망(network)를 형성한 노드(node)라고 불리는 인공적인 뉴런이 학습을 통해 시냅스의 결합 세기를 변화시켜 문제 해결 능력을 가지는 모델을 말한다. 다수의 층(layer)로 구성되어 있으며 구체적으로는 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)로 나누어 진다. Figure 2.1에서 보여지는 것과 같이 각각의 화살표는 노드를 나타내며 반복을 통해 도출되는 가중치는 결합의 정도가 된다. 본 논문에서는 입력 변수로 기온, 습도, 현지기압, 해면기압, 이슬점온도, 일조량, 일사량을 고려하였다.

## 3. 예측모형의 성능비교

태양광 발전량은 여러 가지 날씨와 기후환경에 많은 영향을 받는다. 특히 태양에너지에 결정적인 영향을 받기 때문에 해가 진 후에는 태양광 에너지를 이용하는 것이 불가능하기 하다. 따라서 태양광 발전량 자료는 하루 중 일정시간동안 0의 값을 갖는 특징이 있다. 또한 계절별로 일출과 일몰시간에 차이가 생기기 때문에 하루 중 태양광 발전이 이루어지는 시간이 일정하지 않으며 계절에 따라 정확히 고정된 시간을 갖지 않는다. 이러한 특징들 때문에 시간별 태양광 발전량 예측 모형을 적합할 경우 모형 적합이 잘 되지 않는 어려움이 있다.

본 논문에서는 이러한 문제를 해결하기 위하여 연속된 태양광 발전량 자료를 하루 중 일정 시간만을 분리하여 24개의 독립적인 별도의 시계열로 재구성 하였다. 이러한 방식은 Ramanathan 등 (1997)와 Cottet와 Smith (2003)가 제안하였으며, 전력예측 분야에서는 Cancelo 등 (2008)가 전력 수요 예측에 활용한 바 있다. 연속된 시계열 자료를 시간별로 분리하여 각각의 시계열로 취급하는 것에 대해 아직 의견이 분분하지만 많은 학자들이 이 방법을 선호하며 연구에 이용하고 있다.

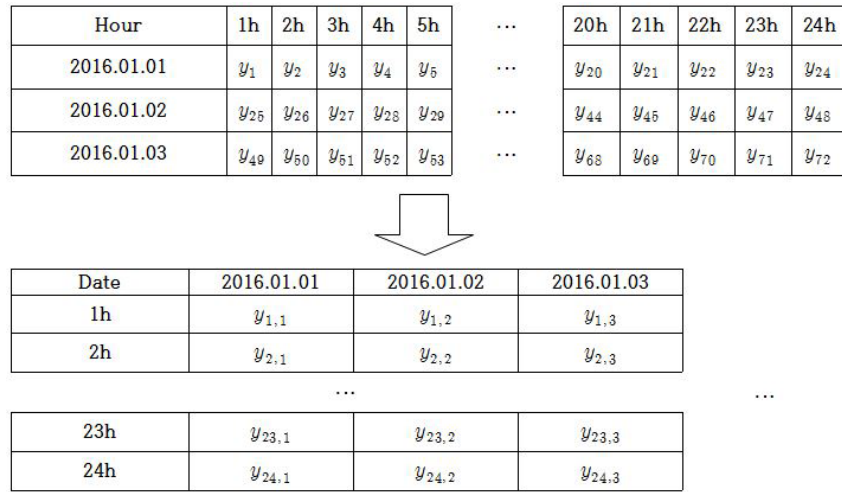


Figure 3.1. Rearrange of time series data.

### 3.1. 예측에 사용된 자료

분석에 활용한 데이터는 2013년 1월 1일부터 2015년 12월 31일까지 1시간 단위로 영암 태양광 발전소에서 관측된 태양광 발전량이며, 2013년 1월 1일부터 2014년 12월 31일까지의 자료를 훈련용 데이터(training data)로 모형 적합에 사용하였다. 그리고 2015년 1월 1일부터 12월 31일까지 자료를 테스트 데이터(test data)로 모형의 성능을 평가하는데 활용하였다. 태양광 발전량 예측을 위한 독립변수는 일반적으로 기상변수들이 주로 사용된다. 국내에서 Won 등 (2011)는 태양광 발전량을 예측하기 위한 입력변수로 일조량과 일사량을 고려 하였고, Lee 등 (2014)은 국내 경상북도 칠곡군의 태양광 발전량을 예측하기 위해 입력변수로 기온, 습도, 일조량, 일사량을 고려하였다. 또한 Yoo 등 (2008)은 태양광 발전량의 주요 변수인 일사량을 계산하기 위해 운량자료를 사용하였다. 국외에서 Inman 등 (2013)은 태양광 발전량과 일사량 예측분야에서 주로 사용되는 기상변수로 기온, 습도, 운량, 대기압 등을 소개한 바 있다. 본 연구에서는 독립변수로 영암 태양광 발전소에서 가장 가까운 장흥, 강진군 기상관측소에서 관측값을 제공하는 기온, 습도, 현지기압, 해면기압, 이슬점온도, 일조량, 일사량을 사용하였으며, 모형 적합 및 예측에는 모두 실제 관측값이 사용되었다. 사용된 기상자료는 장흥, 강진군 기상관측소에서 관측된 자료의 평균을 이용하였다.

분석을 진행하기 위해 이 절의 앞부분에서 설명한 방법대로 영암태양광 발전소의 발전량 자료를 재구성하였다. Figure 3.1은 발전량 자료를 재구성한 방법을 간단히 나타낸 도식이다. 연속된 시간별 시계열 자료를 각 시간대의 일별자료로 재구성 하였다. Figure 3.2는 2013년 한 해 동안의 발전량 자료를 나타낸 시계열 도표이며, Figure 3.3은 태양광 발전량 자료의 패턴을 좀 더 세밀하게 살펴 보기 위해 2013년 1월과 2월의 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. 이 패턴의 주기가 고정되어있지 않으며 하루 각각마다 값이 증가하기 시작하는 시간, 최고점을 이루는 시간, 더 이상 발전이 되지 않는 시간이 다르다.

Figure 3.4는 모형적합에 사용될 2013년과 2014년의 자료를 각각의 시간대별로 분리하여 별도의 시계열 도표로 나타낸 것이다. 전체 1095일 중 대부분의 날에 자료가 존재하는 시간대인 9시부터 18시까지

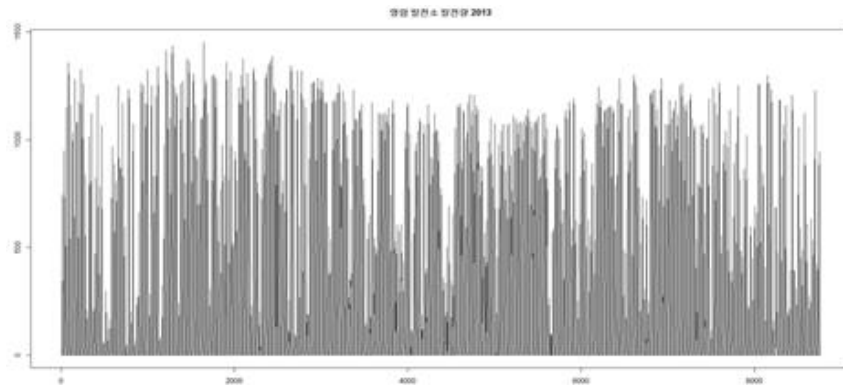


Figure 3.2. 2013.01.01.–2013.12.31. Yeongam Solar Power Plant hourly generation.

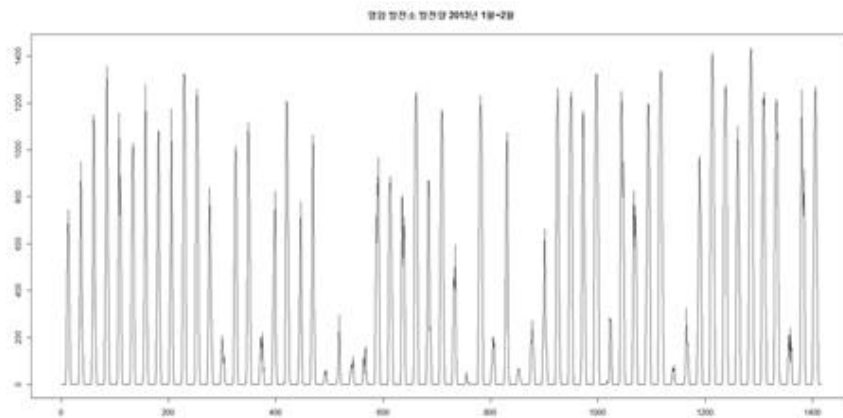


Figure 3.3. 2013.01.01.–2013.02.28. Yeongam Solar Power Plant hourly generation.

10시간의 자료를 이용하였다. 그 결과 연속된 2년간의 시간별 자료가 10개의 2년간의 일별 시계열 자료 형태로 재구성되었다. Figure 3.4에서 가장 이른 아침 시간대 9시, 10시와 저녁 시간대인 17시, 18시에는 일부분에서 분산이 일정하지 않은 모습을 보이지만, 그 외의 시간대에서는 안정적인 시계열 형태를 보여주고 있다. 이렇게 재구성한 자료를 이용하여 10개의 모형을 적합하고, 각각의 시간대에 대한 예측을 진행하여, 전체자료를 이용하여 적합한 모형을 통해 예측한 값과 비교하였다. 비교에 사용된 모형은 Holt-Winters, ARIMAX, 신경망 세 가지이다.

### 3.2. 모형 적합 결과

Holt-Winters 방법을 이용한 예측은 일반적인 시간적 순서에 따른 시계열분석(Model 1)과 자료를 시간별로 정리하여 시계열분석(Model 2)을 각각 수행하였다. Figure 3.5는 일반적 시계열자료를 Holt-Winters 방법을 이용해 구한 예측값을 실제 관측값과 비교한 결과이다. 점선이 실제 관측값이고, 굵은 실선이 모형에 의해 예측된 값이다. 태양광 발전량 자료는 0이하의 값이 나오지 않지만 예측값에서는 많은 부분이 음수로 나타난다. 저녁 시간대에 접어들면서 태양광 발전량이 급감하는 추세가 발전이 끝난 야간 시간대에 영향을 미친 것으로 생각된다.

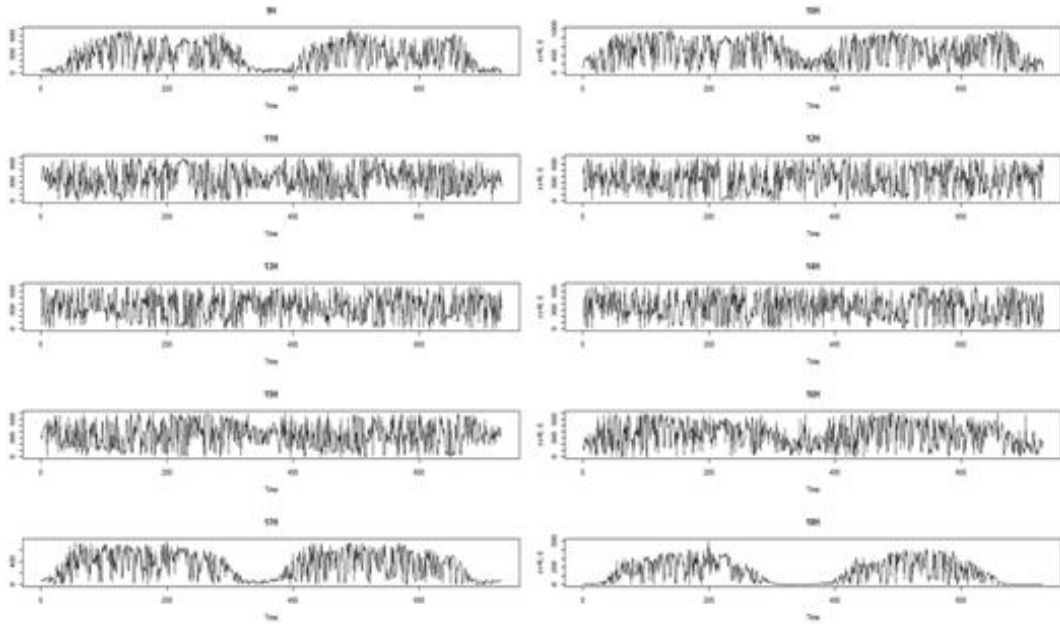


Figure 3.4. 2013.01.01.–2014.12.31. Yeongam Solar Power Plant generation, rearranged by hour.

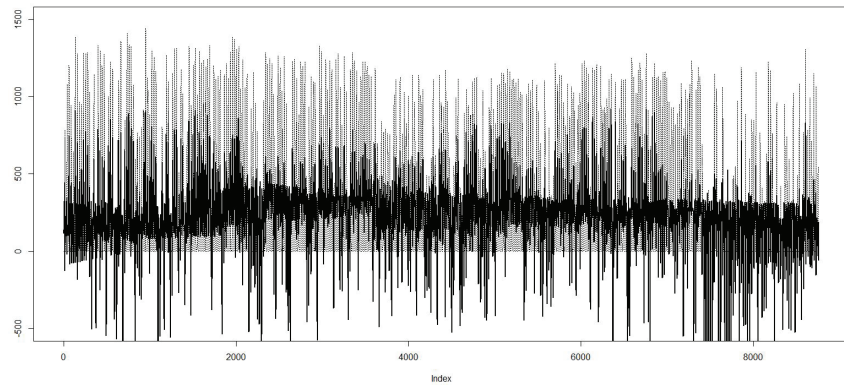


Figure 3.5. Forecasting result 2015.01.01–2015.12.31, Holt-Winters Model 1.

Figure 3.6부터 Figure 3.8은 시간별로 정리된 시계열자료를 Holt-Winters 방법을 이용해 구한 예측값을 각각 시간대 별로 실제 관측값과 비교한 결과이다. 점선이 실제 관측값이고, 굵은 파선이 모형에 의해 예측된 값이다. 대부분의 시간대에서 태양광 발전량의 변동을 민감하게 예측하지 못하지만 평균 수준에서 크게 벗어나지 않는 모습이 나타난다.

ARIMAX 모형의 적합은 일반적 시간순서에 따른 자료(Model 3)와 3절에서 재구성한 자료(Model 4)를 이용하여 진행하였다. 가공하지 않은 원래의 자료를 ARIMAX로 적합한 결과 ARIMA(0,0,0) (1,0,0)<sub>24</sub> 모형이 적합 되었다. 각 독립변수에 대한 계수의 추정값은 Table 3.1과 같다.

$\beta_1$ 부터  $\beta_7$ 은 순서대로 각각 기온, 습도, 이슬점온도, 현지기압, 해면기압, 일조량, 일사량 변수에 대한

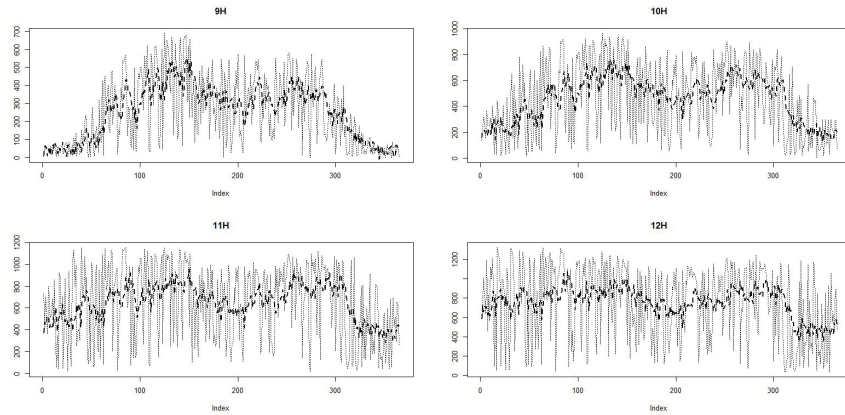


Figure 3.6. Forecasting result 2015.01.01–2015.12.31, Holt-Winters Model 2 (9–12).

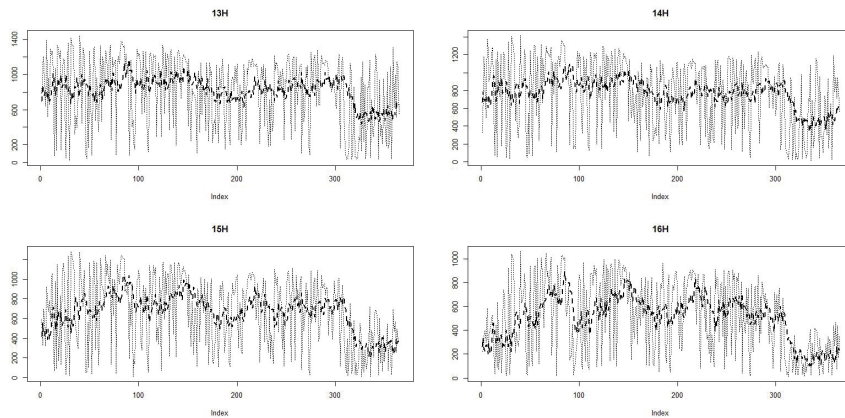


Figure 3.7. Forecasting result 2015.01.01–2015.12.31, Holt-Winters Model 2 (12–16).

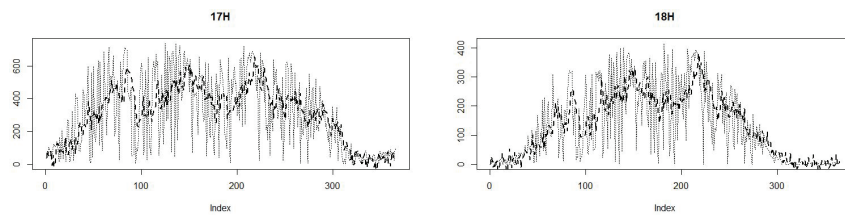


Figure 3.8. Forecasting result 2015.01.01–2015.12.31, Holt-Winters Model 2 (17–18).

계수를 나타낸다. 습도( $\beta_2$ )와 현지기압( $\beta_4$ ), 해면기압( $\beta_5$ )의 계수가 유의하지 않게 나타나지만, 모형의 목적이 예측이고, 제안된 방법과 동일한 변수를 가지는 상황에서의 성능비교인 만큼, 해당 모형을 그대로 사용하기로 하였다. 본 논문에서 제안하는 방법으로 재구성한 자료를 이용하여 ARIMAX를 적합한 결과 각각 Table 3.2와 같은 모형이 적합되었다.

Figure 3.9는 적합한 모형의 예측경향을 알아보기 위해 일반적 시계열자료에 ARIMAX 모형을 적합한 후(Model 3), 2015년 1월 1일부터 10일 후까지 예측을 진행한 결과이다. 예측에 사용된 독립변수는 실

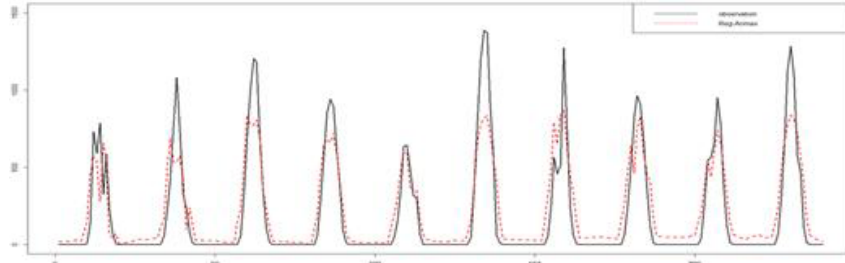


**Table 3.1.** Parameter estimate of ARIMA(0, 0, 0)(1, 0, 0)<sub>24</sub> model

Parameter	Estimate	Standard error
$\Phi_{24}$	0.609	0.007
Intercept	-3078.962	349.476
$\beta_1$	6.586	1.350
$\beta_2$	0.112	0.283
$\beta_3$	-2.965	1.178
$\beta_4$	35.767	47.643
$\beta_5$	-32.619	47.476
$\beta_6$	138.083	6.076
$\beta_7$	198.492	2.727

**Table 3.2.** Fitted models using rearranged data

Hour	Model
9h	ARIMA(1, 1, 1) with regressors
10h	ARIMA(1, 1, 1) with regressors
11h	ARIMA(1, 1, 2) with regressors
12h	ARIMA(0, 1, 1) with regressors
13h	ARIMA(0, 1, 1) with regressors
14h	ARIMA(0, 1, 1) with regressors
15h	ARIMA(2, 1, 3) with regressors
16h	ARIMA(0, 1, 1) with regressors
17h	ARIMA(0, 1, 1) with regressors
18h	ARIMA(0, 1, 1) with regressors

**Figure 3.9.** 10days ahead forecasting, ARIMAX Model 3.

제 관측값을 이용하였다. 실선이 관측값, 점선이 모형에 의한 예측값이다.

일반적 시계열자료를 이용한 예측 결과 발전량의 전체적인 경향은 어느 정도 잘 맞추는 모습을 보이지만, 하루의 최대값과 그 근방에서 예측력이 떨어지는 모습을 볼 수 있다. 이는 독립변수 중 태양광 발전량에 결정적인 영향을 미치는 일조량과 일사량이 포함되어 있어 하루 중 발전량이 증가하기 시작하는 부분과 더 이상 발전하지 않는 부분을 잘 예측한 것으로 예상된다.

Figures 3.10–3.12는 적합한 모형의 예측경향을 알아보기 위해 시간별로 정리된 시계열자료에 ARIMAX 모형을 적합한 후(Model 4), 2015년 1월 1일부터 10일 후까지 각각 시간대 별로 예측을 진행한 결과이다. 예측에 사용된 독립변수는 실제 관측값을 이용하였다. 실선이 관측값, 점선이 모형에 의한 예측값이다.



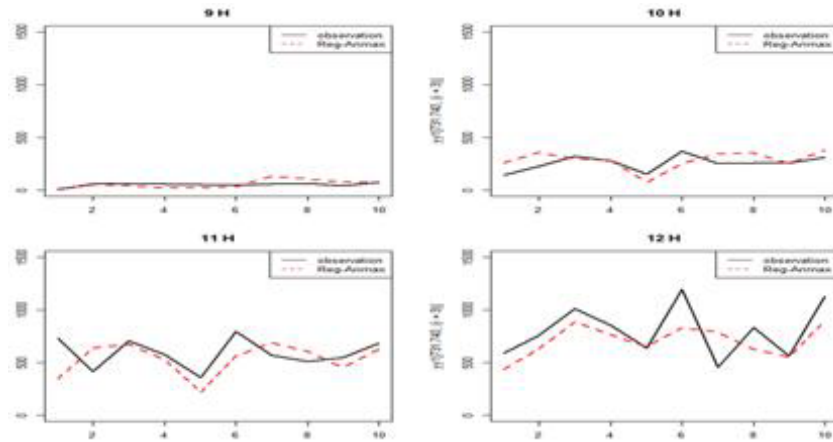


Figure 3.10. 10days ahead forecasting, ARIMAX Model 4 (9-12).

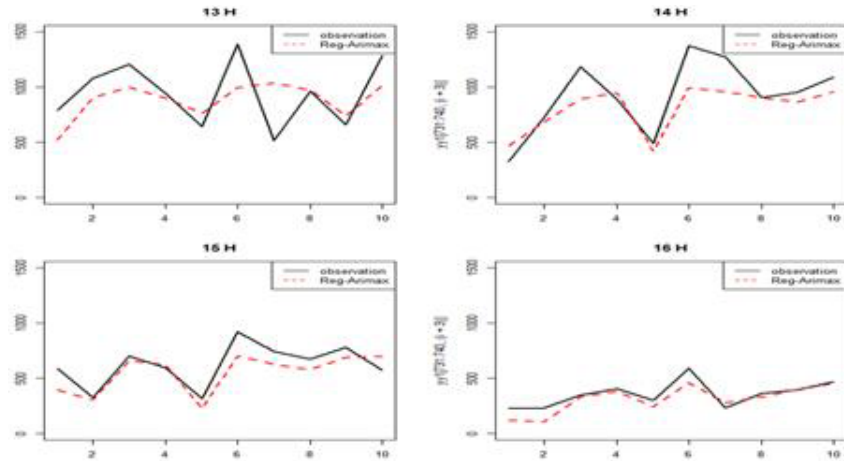


Figure 3.11. 10days ahead forecasting, ARIMAX Model 4 (13-16).

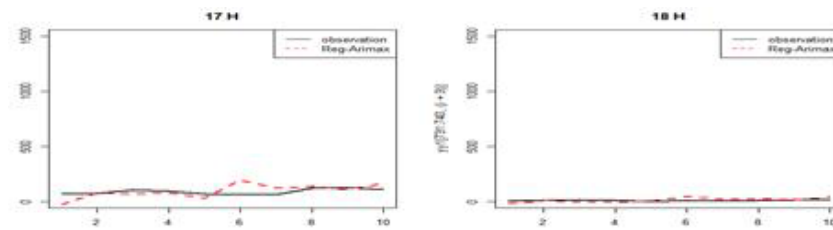


Figure 3.12. 10days ahead forecasting, ARIMAX Model 4 (17-18).

시간별로 정리된 시계열자료를 이용한 예측 결과 실제 관측 값이 대부분 0인 9시, 17시, 18시에서 관측 값과 별로 차이가 나지 않는 것처럼 보이지만, 현재 예측하고 있는 2015년 1월 1일부터 1월 10일까지의 자료는 해당 시간대의 값이 거의 0에 가까워 태양광 발전량을 예측하는데 있어서 큰 의미가 없다. 오전

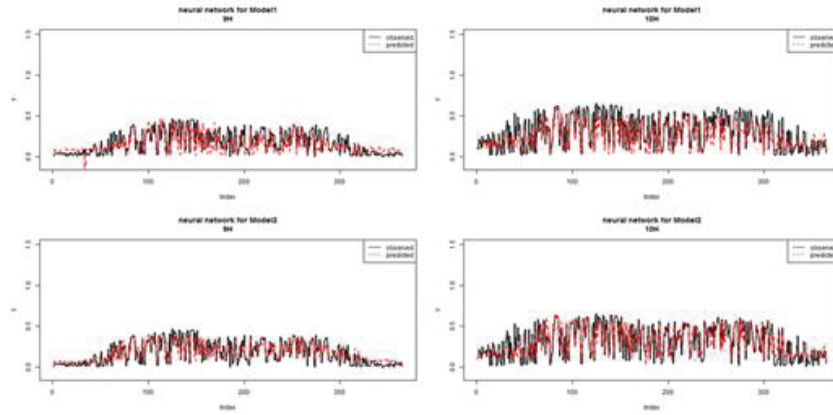


Figure 3.13. Forecast comparison of ANN Model 5 and ANN Model 6 (9–10).

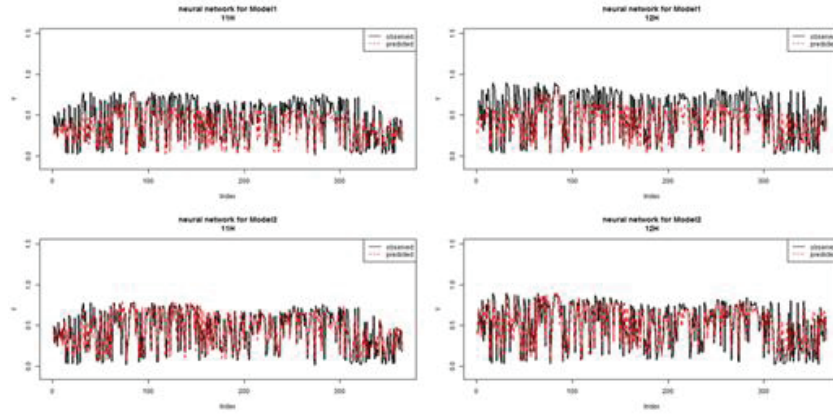


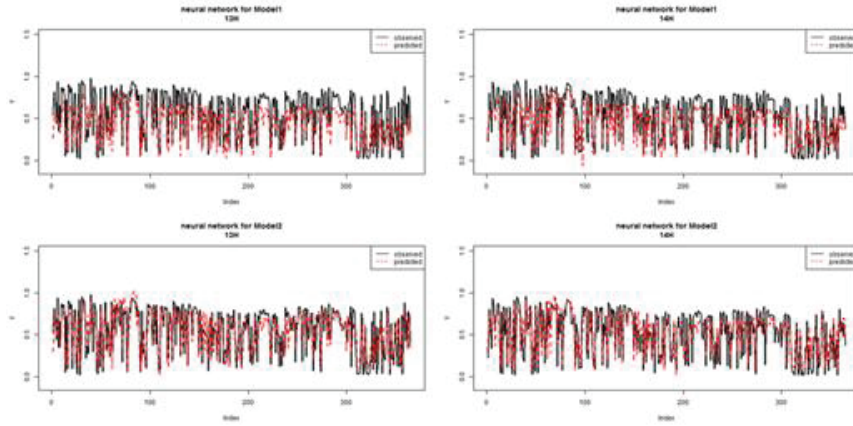
Figure 3.14. Forecast comparison of ANN Model 5 and ANN Model 6 (11–12).

에서 오후로 넘어가면서 약 12시부터 15시 사이에 하루 중 최대값을 이루게 되는데, 일반적 자료를 이용해 적합한 결과(Figure 3.9)에서 하루 중 최대값에 대한 예측값이 실제값과 크게 차이는 것과 달리 어느 정도 비슷한 수준을 유지하며 예측하는 모습을 보여주고 있다. 이와 같은 결과로 동일한 시간대의 자료만을 사용하여 고정적이지 않은 주기의 영향이 어느 정도 제외되었음을 예상해 볼 수 있다.

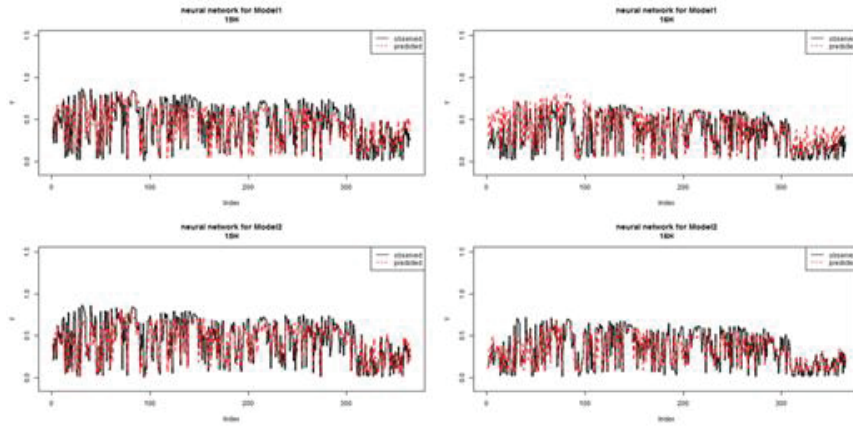
인공신경망을 활용한 분석을 하기 위해서는 자료의 표준화가 필요하다. 일반적으로 0에서 1사이의 값을 가지는 방식으로 아래와 같은 식을 통해 변환한다.

$$\frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}.$$

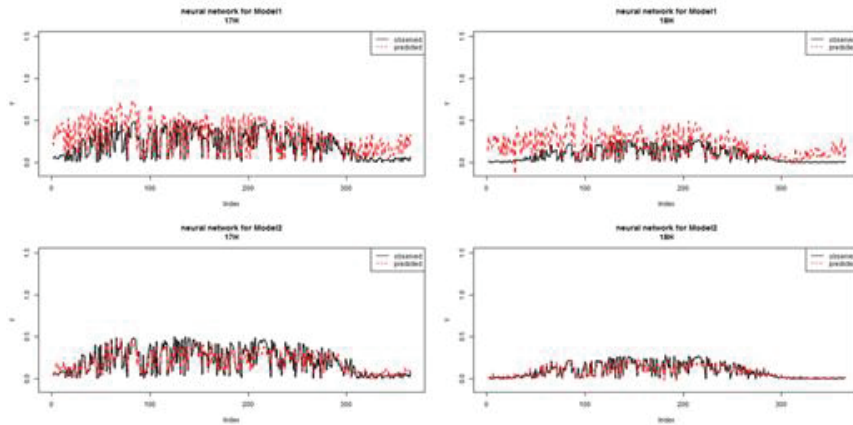
신경망 모형을 이용한 모형 적합은 일반적인 시간적 순서에 따른 시계열분석(Model 5)과 자료를 시간 별로 정리하여 시계열분석(Model 6)을 각각 수행하였다. 인공신경망 모형 분석 또한 다른 기계학습 알고리즘과 마찬가지로 학습(train)과정과 검증(test)과정으로 나누어진다. 본 논문에서는 학습을 위해 2013년 1월 1일부터 2014년 12월 31일의 자료를 사용하였고, 나머지 2015년 1월 1일부터 2015년 12월 31일까지의 자료로 검증하였다. Figure 3.14에서 Figure 3.17은 각각의 시간대에 대해 일반적 시계열자



**Figure 3.15.** Forecast comparison of ANN Model 5 and ANN Model 6 (13–14).



**Figure 3.16.** Forecast comparison of Model 5 and Model 6 (15–16).



**Figure 3.17.** Forecast comparison of ANN Model 5 and ANN Model 6 (17–18).

료와 시간별로 정리된 시계열자료의 예측값을 실제 관측값과 비교한 결과이다. 검정색 실선이 실제 관측값이고, 붉은색 점선이 모형에 의해 예측된 값이다. 각각 상단의 도표가 일반적 시계열자료, 하단의 도표가 시간별로 정리된 시계열자료에 의한 결과이다.

신경망 모형으로 예측된 결과 일반적 시계열자료를 사용한 경우 9시부터 14시까지의 경우 모형에 의한 예측값이 과소 추정된 것을 확인할 수 있다. 그 이후인 15시부터 18시까지의 경우는 모형에 의한 예측값이 오히려 과대 추정 되어있다. 반면 시간별로 정리된 시계열자료를 사용한 경우 과소, 과대 추정되는 경향이 없고, 어느 정도 잘 예측되고 있는 모습을 볼 수 있다.

### 3.3. 모형의 성능 비교

앞의 절에서 적합한 모형들을 비교하기 위한 오차의 척도로 mean absolute error (MAE)와 root mean squared error (RMSE)를 이용하였다. 일반적으로 전력수요예측에서는 MAPE를 많이 사용하지만 태양광 발전의 경우 자료의 특성상 자료값이 0으로 나타나 mean absolute percentage error (MAPE)를 계산할 수 없는 경우가 많아 제외하였다. MAE는 실제 관측값과 예측값의 차이의 절대값을 전체 자료의 개수로 나눈 척도로 아래 식과같이 정의된다.

$$MAE = \frac{\sum_{t=1}^m |Y_t - \hat{Y}_t|}{m}$$

Mean squared error (MSE)는 실제 관측값과 예측값의 차이의 제곱합 평균으로 정의된다. RMSE는 MSE의 제곱근이며 아래 식과같이 정의된다.

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

위의 두 척도 모두 MAPE와 달리 자료의 크기에 영향을 받는 척도이며, 작을수록 모형의 성능이 우수하다는 것을 의미한다. 모형 비교를 위하여 time series cross-validation 기법을 이용하였다. time series cross-validation이란,  $k$ 차 이후 시점에 대한 모형의 예측력을 알아보기 위해 훈련자료를 전부 사용할 때까지 훈련자료를 한 시점씩 추가해 나가면서 예측값을 구해 오차를 구하는 모형 평가 방식이다. 본 논문에서는 적합된 모형의 1일후 시점에 대한 예측력을 비교하였다. 모형의 비교결과는 Tables 3.3-3.5와 같다.

모형비교 결과, Holt-Winters 방법을 이용한 경우와 ARIMAX 모형을 이용한 경우, 신경망 모형을 이용한 경우 본 논문에서 제안한 재구성된 자료를 이용하는 경우가 모두 우수하게 나타났다. ARIMAX를 이용한 경우에서 아침 9시에서 기존방법이 더 우수하게 나타나지만 그 외의 모든 시간대에서 제안된 방법이 우수하게 나타나고 있다. 비교에 사용된 두 척도 모두 자료의 크기에 영향을 받기 때문에 각 시간 대별로만 비교가 가능하다.

## 4. 결론

매장된 지하자원의 한계, 환경 보호의 중요성이 증가로 인하여 신재생에너지에 대한 관심이 증가하고 있다. 특히 태양에너지를 이용한 산업이 성장하고 있다. 태양광발전 의존율이 높아질수록 발전량의 정확한 예측이 중요해질 것이다. 이에 대비하여 기상자료를 이용하여 신경망 모형과 시계열 모형을 통한 태양광 발전량의 예측을 실시하였다. 이 때, 시계열 자료를 그대로 사용하는 기존 방법과, 연속된 시계열

**Table 3.3.** Model comparison, Holt-Winters

Hour	Holt-Winters 일반자료		Holt-Winters 시간별자료	
	RMSE	MAE	RMSE	MAE
9h	199.71	157.06	174.78*	122.44*
10h	342.87	259.56	290.86*	219.70*
11h	508.26	337.39	469.15*	284.90*
12h	597.62	382.06	564.54*	323.94*
13h	653.48	398.98	626.08*	335.88*
14h	600.62	397.95	567.96*	339.96*
15h	496.97	347.45	454.78*	291.01*
16h	349.97	288.07	308.48*	241.99*
17h	217.02	184.42	169.92*	146.67*
18h	200.50	90.37	163.02*	66.82*

RMSE = root mean squared error; MAE = mean absolute error.

**Table 3.4.** Model comparison, ARIMAX

Hour	ARIMAX 일반자료		ARIMAX 시간별자료	
	RMSE	MAE	RMSE	MAE
9h	55.73*	43.67*	86.97	63.12
10h	179.83	147.56	134.20*	99.33*
11h	237.34	199.00	168.76*	118.36*
12h	309.25	247.86	186.75*	138.13*
13h	327.73	268.24	187.24*	138.94*
14h	278.10	225.57	198.75*	147.77*
15h	242.78	188.03	177.71*	134.67*
16h	196.63	148.80	148.38*	110.92*
17h	191.08	152.82	108.39*	79.85*
18h	178.76	148.69	52.63	37.30*

RMSE = root mean squared error; MAE = mean absolute error.

**Table 3.5.** Model comparison, ANN

Hour	ANN 일반자료		ANN 시간별자료	
	RMSE	MAE	RMSE	MAE
9h	0.09	0.07	0.07*	0.05*
10h	0.12	0.09	0.10*	0.08*
11h	0.16	0.12	0.13*	0.10*
12h	0.17	0.14	0.15*	0.11*
13h	0.18	0.15	0.15*	0.11*
14h	0.17	0.13	0.14*	0.11*
15h	0.13	0.10	0.13*	0.10*
16h	0.14	0.11	0.11*	0.08*
17h	0.17	0.13	0.08*	0.06*
18h	0.15	0.12	0.04*	0.03*

RMSE = root mean squared error; MAE = mean absolute error.

자료를 하루 중 일정 시간만을 분리하여 24개의 독립적인 별도의 시계열로 재구성 하는 방법의 성능을 비교하였다.

Time series cross-validation 기법을 통한 1년간의 테스트 데이터에서 예측 모형을 평가한 결과, ARI-MAX 모형에서는 오전 9시를 제외한 모든 시간대에서 시계열 자료를 재구성한 방법이 우수하게 나타났고 Holt-Winters, 신경망 모형에서는 모든 시간대에서 시계열 자료를 재구성한 방법이 우수하게 나타났다.

## References

- Cancelo, J. R., Espasa, A., and Grafe, R. (2008). Forecasting the electricity load from one day to one week ahead for the Spanish system operator, *International Journal of Forecasting*, **24**, 588–602.
- Cottet, R and Smith, M. (2003). Bayesian modeling and forecasting of intraday electricity load, *Journal of the American Statistical Association*, **98**, 839–849.
- Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages, *ONR Memorandum*, **52**. Pittsburgh, PA: Carnegie Institute of Technology. Available from the Engineering Library, University of Texas at Austin.
- Inman, R. H., Pedro, H. T., and Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration, *Progress in Energy and Combustion Science*, **39**, 535–576.
- Kim, K. H. and Kim, J. Y. (2016). The optimal design and economic evaluation of a stand-alone RES energy system for residential, agricultural and commercial sectors, *Korean Chemical Engineering Research*, **54**, 470–478.
- Lee, S. H., Kim, H. D., and Cho, C. (2014). Study on the variation characteristic of the photo-volatic power generation due to regional meteorological elements, *Journal of Environmental Science International*, **23**, 1943–1951.
- Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., and Brace, C. (1997). Short-run forecasts of electricity loads and peaks, *International Journal of Forecasting*, **13**, 161–174.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages, *Management Science*, **6**, 324–342.
- Won, J. M., Doe, G. Y., and Heo, N. R. (2011). Predict solar radiation according to forecast report, *Journal of Korean Navigation and Port Research*, **35**, 387–392.
- Yoo, H. C., Lee, G. H., and Park, S. H. (2008). Analysis of data and calculation of global solar radiation based on cloud data for major cities in Korea, *Journal of the Korean Solar Energy Society*, **28**, 17–24.

# 시계열 모형과 기상변수를 활용한 태양광 발전량 예측 연구

이근호<sup>a</sup> · 손흥구<sup>b</sup> · 김삼용<sup>a,1</sup>

<sup>a</sup>중앙대학교 응용통계학과, <sup>b</sup>한국교통연구원 항공교통본부

(2017년 11월 27일 접수, 2017년 12월 11일 수정, 2017년 12월 11일 채택)

---

## 요약

최근 정부의 친환경 정책에 따라 태양광 발전 설비가 지속적으로 증가하고 있다. 태양광 발전량은 에너지원인 태양의 특성상 계절에 따라 하루 중 발전이 이루어지는 시간이 일정하지 않다. 이러한 특성으로 인해 태양광 발전량 예측에서는 연속된 시간간격으로 수집된 자료에 적용할 수 있는 시계열 모형 적용에 어려움이 있다. 본 논문에서 제안하는 방법은 연속된 시간자료를 각 시간대 별로 분리, 재구성하여 24개의 (1시-24시) 일별 자료 형태로 예측에 활용하는 방법이다. 강원도 영암 태양광 발전소의 시간별 발전량 자료를 공공데이터포털에서 수집하여 연구하였다. 기존방법과 제안된 방법의 성능차이를 비교하기 위해 ARIMAX, 신경망(neural network model) 모형을 동일한 모형과 변수를 가지는 환경에서 성능차이를 확인하였다.

주요용어: ARIMAX, 신경망, 기상변수, 태양광 발전

---

---

이 논문은 2016년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016R1D1A1B01014954).

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