

## 전처리 방법과 인공지능 모델 차이에 따른 대전과 부산의 태양광 발전량 예측성능 비교: 기상관측자료와 예보자료를 이용하여

심채연<sup>1)</sup> · 백경민<sup>1)</sup> · 박현수<sup>1)</sup> · 박종연<sup>1),2)\*</sup>

<sup>1)</sup>전북대학교 지구환경과학과, <sup>2)</sup>전북대학교 환경에너지융합학과

(접수일: 2024년 3월 21일, 수정일: 2024년 4월 15일, 게재확정일: 2024년 4월 25일)

### Comparison of Solar Power Generation Forecasting Performance in Daejeon and Busan Based on Preprocessing Methods and Artificial Intelligence Techniques: Using Meteorological Observation and Forecast Data

Chae-Yeon Shim<sup>1)</sup>, Gyeong-Min Baek<sup>1)</sup>, Hyun-Su Park<sup>1)</sup>, and Jong-Yeon Park<sup>1),2)\*</sup>

<sup>1)</sup>Department of Earth and Environmental Sciences, Jeonbuk National University, Jeonju, Korea

<sup>2)</sup>Department of Environment and Energy, Jeonbuk National University, Jeonju, Korea

(Manuscript received 21 March 2024; revised 15 April 2024; accepted 25 April 2024)

**Abstract** As increasing global interest in renewable energy due to the ongoing climate crisis, there is a growing need for efficient technologies to manage such resources. This study focuses on the predictive skill of daily solar power generation using weather observation and forecast data. Meteorological data from the Korea Meteorological Administration and solar power generation data from the Korea Power Exchange were utilized for the period from January 2017 to May 2023, considering both inland (Daejeon) and coastal (Busan) regions. Temperature, wind speed, relative humidity, and precipitation were selected as relevant meteorological variables for solar power prediction. All data was preprocessed by removing their systematic components to use only their residuals and the residual of solar data were further processed with weighted adjustments for homoscedasticity. Four models, MLR (Multiple Linear Regression), RF (Random Forest), DNN (Deep Neural Network), and RNN (Recurrent Neural Network), were employed for solar power prediction and their performances were evaluated based on predicted values utilizing observed meteorological data (used as a reference), 1-day-ahead forecast data (referred to as fore1), and 2-day-ahead forecast data (fore2). DNN-based prediction model exhibits superior performance in both regions, with RNN performing the least effectively. However, MLR and RF demonstrate competitive performance comparable to DNN. The disparities in the performance of the four different models are less pronounced than anticipated, underscoring the pivotal role of fitting models using residuals. This emphasizes that the utilized preprocessing approach, specifically leveraging residuals, is poised to play a crucial role in the future of solar power generation forecasting.

**Key words:** Solar power generation, Residual analysis, Forecast data, Machine learning

\*Corresponding Author: Jong-Yeon Park, Earth and Environmental Sciences, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do 54896 Korea.  
Phone: +82-63-270-3398, Fax: +82-63-270-3316  
E-mail: jongyeon.park@jbnu.ac.kr

## 1. 서론

NOAA (National Oceanic and Atmospheric Administration)에 따르면, 2020년 전 지구 평균 이산화탄소 농도는 412.5 ppm이며, 산업혁명 이전과 비교했을 때 148.4% 수준이다(NOAA, 2020). 이는 산업혁명 이후 화석연료 사용의 폭발적인 증가에 따라 온실가스 배출량이 지속적으로 증가했기 때문으로 알려져 있다(IPCC, 2021). 이에 전 세계적으로 탄소배출과 에너지 문제의 중요성이 대두되며 친환경 에너지의 관리 및 운영 체계가 중요해졌고 이를 위해 필요한 기술이 에너지 발전량 예보이다(Son, 2023).

친환경 에너지 중 태양광 패널의 효율 상승 및 패널 가격인하 등 관련 기술의 발전으로 에너지 생산 효율이 증가함에 따라 태양광 발전이 주목받고 있다. 한국에너지공단이 제공한 자료에 따르면 2021년 기준, 대한민국의 태양광 발전량은 친환경 에너지 총발전량의 48.8%를 차지했다(KEA, 2021). 이는 대한민국 전체 에너지 발전량의 4%에 해당하며, 주요 에너지 중 하나이다(KIER, 2023). 따라서 산업통상 자원부와 한국 전력 거래소는 에너지의 효율적 관리와 탄소 중립 이행을 위해 2020년에 에너지 발전량 예측 제도를 도입하고 '전력시장운영규칙 개정안'을 제시했다(KPX, 2020).

이러한 추세에 발맞추어 기상변수와 기계학습 모델을 이용한 여러 연구가 진행되었다. 선행연구는 다음과 같은 3가지 특징을 갖는다. 첫째, 기상관측자료를 주로 사용하였으며, 관측자료와 기상청 예보자료를 함께 이용한 연구는 부족한 실정이다(Lee et al., 2018). 둘째, 태양광 발전량의 지점별 위치정보와 설비용량 정보의 부족으로 선행연구에서는 한 지점의 정보에 대한 예측을 본 것이 대부분이었다(Kim, 2019; Kim et al., 2020). 마지막으로, 단위에 따라 편향된 학습을 방지하기 위해 자료의 전처리 과정에서 최소-최대 스케일링(Min-Max Scaling) 혹은 Z-점수 스케일링을 이용한 경우가 대부분이었다(Kang and Yun, 2023; Ki et al., 2023).

이에 본 연구에서는 다음과 같은 차별점을 두고 태양광 발전 모델을 개발하였다. 첫째, 에너지 예보에서 기상예보자료의 효용성을 확인하고자, 관측자료를 이용해 모델을 학습시킨 후, 기상청의 1일 전, 2일 전 예보자료를 이용하여 모델의 예측성능을 평가했다. 둘째, 대전과 부산 내의 모든 지점을 고려하여 지역별 평균적인 예측을 하고자 했다. 셋째, 시계열 자료에서 체계적 성분을 제외한 불규칙성분만을 남겨 각 모델이 불규칙성분을 학습하는데 집중하도록 하였다. 이를 통해 태양광 발전량 예측에서 중요한 전처리 방안을 제안하였다.

## 2. 자료 및 실험방법

### 2.1 자료

기상청의 기상관측자료와 기상예보자료를 독립변수로, 한국전력거래소의 태양광 발전량 자료를 종속변수로 사용하였다. 한국전력거래소에서 제공하는 태양광 발전량 자료는 광역시 또는 도 규모로 제공되는데, 도 수준에서는 지리적 영역이 넓어 기상변수와 태양광 발전량을 대응시켜 연구하는데 한계가 있다고 판단하였다. 이에 본 연구에서는 광역시 수준에서 연구를 진행하면서 지역의 기후적 특성에 따른 성능 차이를 확인하고자 했다. 대표적인 기후구분법 중 하나인 트레와다 구분법을 기준으로 온대내륙성기후(Dca)에 속하며 전형적인 내륙지방의 기후를 보여주는 대전광역시와 아열대습윤기후(Cfa)에 속하며 해안지역의 기후를 갖는 부산광역시를 선정하였다(Kang et al., 2021).

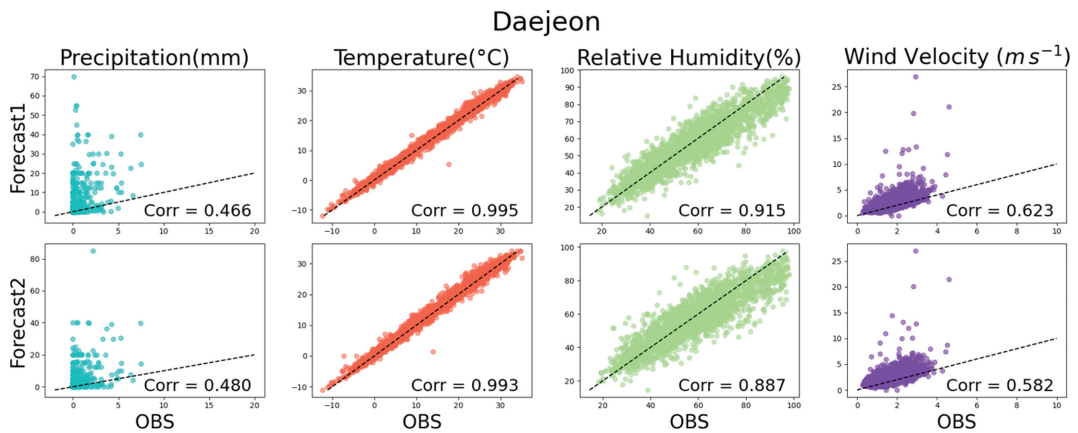
한국전력거래소에서 일부 제공된 지점정보와 범지구위치결정시스템(Global Positioning System, GPS) 영상을 이용해 태양광 발전패널이 대전과 부산 내에서 비교적 고르게 분포되어 있음을 확인한 후 각 도시의 평균 태양광 발전량을 계산했다. 이에 따라 독립변수의 기상관측자료 또한 도시 전체의 평균값을 이용하기 위해 ASOS (Automated Synoptic Observing System) 뿐만 아니라 AWS (Automatic Weather System)도 활용하였으며, 기상예보자료는 모두 동네예보를 사용하였다(Table 1). 기상청의 예보자료가 기상 변수별로 3시간 단위, 6시간 단위로 제공되어, 시간 단위의 자료가 아닌 태양광 발전이 주로 일어나는 9시부터 18시 까지를 선정한 후 일평균하여 사용했다.

기상변수는 기상관측자료와 기상예보자료 모두에서 공통으로 제공되는 변수를 추렸다. 선행연구처럼 태양광 발전량 예측에서 일사량을 주요 변수로 사용할 수도 있었지만(Lee et al., 2019; Kim et al., 2020), 본 연구에서는 기상관측자료와 예보자료의 비교를 통해 효용성을 확인하고자 했기 때문에 예보자료가 제공되지 않는 일사량을 고려하지 않았다. 또한 하늘상태와 같은 범주형 변수는 기계학습 시 숫자로 변환되어 이용되는데 이 경우 가중치로 잘못 인식되어 값의 왜곡으로 인한 예측 성능 저하를 유발할 수 있다는 단점이 있어 변수에서 제외하였다. 또한 태양광 패널의 특성을 고려하였을 때 상대습도와 풍속이 패널 효율을 결정하는 주요 변수이기 때문에 태양광 발전량 예측에 활용할 기상변수로서 상대습도, 온도, 강수량, 풍속을 최종적으로 선정하였다(Mekhilef et al., 2012).

선정된 예측 인자변수들과 태양광 발전량 사이의 관련성을 상관관계 분석을 통해 살펴보았을 때, 상대습도와 태양광 발전량 사이의 상관계수 값(r-value)은 대전에서 -0.706, 부산에서 -0.569로 95% 신뢰수준에

**Table 1.** Weather observation and forecast data in Daejeon and Busan from KMA (Korea Meteorological Administration).

Region	Data Type (Number of Stations)	Stations	Variables	Period (Interval)
Daejeon	ASOS (1)	대전	Relative Humidity [%] Temperature [°C] Precipitation [mm] Wind Velocity [m s <sup>-1</sup> ]	20180101~20230531 (hourly)
	AWS (3)	장동, 세천, 오월드		
	Forecast (79)	대전 동네예보		
Busan	ASOS (1)	부산	Relative Humidity [%] Temperature [°C] Precipitation [mm] Wind Velocity [m s <sup>-1</sup> ]	20180101~20230531 (hourly)
	AWS (12)	금정구, 기장, 북구, 동래, 해운대, 사상, 부산진, 부산(레), 부산남구, 사하, 영도, 가덕도		
	Forecast (205)	부산 동네예보		



**Fig. 1.** Comparison between observations and forecasts of four meteorological variables. The first row for each variable represents the comparison of observation with forecast at a 1-day lead time. The second row of each region represents the comparison between observation and forecast values at a 2-day lead time. The  $y = x$  line is depicted in black dashes, indicating the similarity between observed and forecasted values.

서 유의한 음의 상관관계를 가졌다. 그 외에 기온, 강수량, 풍속의 상관계수는 대전에서 0.147, -0.363, 0.175, 부산에서 0.098, -0.317, 0.026으로 나타나 상대적으로 상관관계 값이 낮았지만 단순 선형관계가 약하다고 해서 태양광 발전량 예측에서의 기상변수 효용성이 없는 것이 아니기에 본 연구에서는 4가지 변수를 모두 사용하였다(Kim, 2019). 선정된 4가지 기상변수를 이용해 태양광 발전량을 예측하는 회귀문제를 풀고자 하는 것이 목적이므로, 독립변수 간의 다중 공선성을 먼저 확인 하였고, 이를 위해 독립변수 간의 상관계수와 분산팽창지수인 VIF (Variance Inflation

Factor)를 계산하였다. 대전과 부산 모두에서 상관계수가 0.5 이하였으며 VIF 또한 모두 1 이하로 다중 공선성 문제가 없다고 판단하였다.

본 연구에서는 태양광 발전량 예측에 기상예보자료 또한 활용하기에 모델을 학습시키기 전 기상예보자료와 기상관측자료가 얼마나 유사한지 확인하고자 Figs. 1, 2의 산점도를 그렸다. 각 그림에서 검정색 점선은  $y = x$ 을 나타내며, 이 선에 가까울수록 예보자료와 관측자료가 유사한 것을 나타낸다. 대전과 부산 모두 기온과 상대습도에서 예보자료의 정확성이 높음을 확인할 수 있었고, 강수와 풍속은 과대모의를

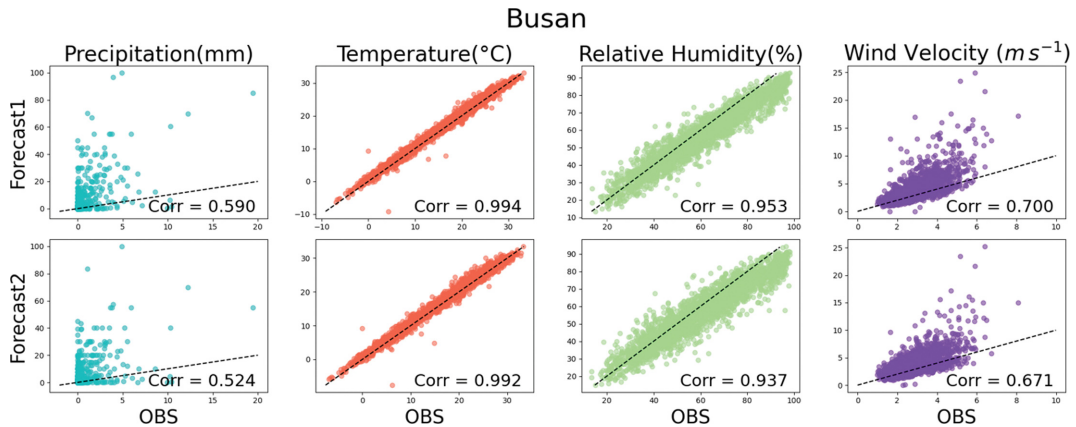


Fig. 2. Same as Fig. 1. but in Busan.

하고 있다. 또한 두 지역 모두에서 1일 전 예보가 2일 전 예보보다 관측과 더 유사하게 나타났으나, 그 차이는 크지 않았다. 따라서 기상변수 기반 태양광 발전량 예측 모델의 신뢰성이 보장된다면 2일까지는 기상예보자료로 발전량 예측이 가능할 수 있을 것으로 판단되었다.

2.2 실험방법  
2.2.1 전처리

모델 학습 시 시계열 자료의 체계적 성분을 제외한 불규칙성분만 학습에 이용하고자 했다. 독립변수와 종속변수 모두 추세와 계절성을 제거한 불규칙성분만 남겼고 최종적으로 선택된 기상변수들을 독립변수로 설정하여 회귀 문제를 해결하고자 했다. 기존의 연구에서는 스케일링 과정을 통한 전처리가 일반적이었으나 본 연구에서는 추세와 계절성을 제외한 태양광 발전량을 예측하여 성능을 높이고자 하였다. 추세와 계절성은 선형회귀를 이용해 제거하는 방법을 선택하였다. 계절성 제거를 고려하기 위해 월별로 데이터를 나누고 각 월별 회귀선을 구했다. 원데이터의 값과 회귀선으로 구해진 값의 차이를 산출해 추세와 계절성이 제거된 값을 학습에 이용했다.

기상변수의 변수 중요도를 파악하기에 용이하고, 단위의 차이로 편향된 학습을 방지하기 위해 독립변수에 식(1)과 같이 최소-최대 스케일링을 진행했다(Kim, 2019). 또한 종속변수가 회귀 문제를 풀기 위한 가정인 등분산성을 만족하지 못하는 것을 확인한 후, 등분산성을 만족하도록 종속변수에 제곱 변환하여 가중 처리했다(Voyant et al., 2020).

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{1}$$

이러한 전처리 과정의 유의미한 성능 차이를 확인하기 위해 최종적으로 전처리 수준을 3단계로 나누어 학습하고 그 성능을 비교하였다. 독립변수에 단순히 최소-최대 스케일링만 진행한 것을 No pre, 독립변수와 종속변수에 추세와 계절성을 제거한 후 독립변수에 최소-최대 스케일링을 진행한 것을 Partial pre, Partial pre에서 종속변수에 가중 처리를 해 준 것을 Full pre로 명명했다.

- (No pre): 독립변수(최소-최대 스케일링), 종속변수(-)
- (Partial pre): 독립변수(detrend+de-seasonality, 최소-최대 스케일링), 종속변수(detrend+de-seasonality)
- (Full pre): 독립변수(detrend+de-seasonality, 최소-최대 스케일링), 종속변수(detrend+de-seasonality, weighted)

본 연구에서 사용한 여러 예측 모델 중 하나인 Recurrent Neural Network (RNN)에 이용되는 자료는 시계열 형식의 자료 배열이 필요하기에 추가적인 전처리를 실시했다. 각 자료를 순차적으로 7개씩 중첩해 묶어준 후 새로운 차원을 구성했다. 이를 통해 과거 6일과 현재의 정보를 이용하여 RNN이 학습할 수 있는 형태로 자료를 변환하였다.

2.2.2 예측 모델

예측 모델에는 Random Forest (RF), Multiple Linear Regression (MLR), Deep Neural Network (DNN), RNN을 사용했다. RF는 의사결정나무를 다수 이용하는 앙상블 모델이며, MLR는 회귀분석기법 중 하나로 여러 개의 독립변수와 하나의 종속변수 사이의 인과 관계를 찾고 모델을 만들어 변수를 예측하는 기법이다. RF와 MLR는 대표적인 기계학습 기법으로 널리 사용된다. DNN은 여러 개의 은닉층으로 이루어진 인

공신경망으로 가장 간단한 구조의 딥러닝 모델이다. RNN은 DNN에서 시계열 자료를 처리하는 데 중점을 둔 모델로 과거의 자료를 이용하여 현재의 학습에서 가중치 업데이트에 사용하는 특징을 지닌다. 즉, RNN은 구성 자료에 과거의 정보를 담은 차원을 추가하여 시간적 정보를 학습하도록 한다는 점에서 DNN과 다른 특징을 가지는 딥러닝 모델이다. 본 연구의 주요 목적 중 하나가 자료 전처리의 효율성 확인이므로 기존 연구에서 사용한 여러 모델 중 복잡성이 낮은 알고리즘을 선택하였다. 기계학습 모델로 기초적인 MLR, RF 모델을 선택했고 딥러닝 모델로는 가장 기초적인 알고리즘인 DNN과 시계열 자료 처리에 적합한 모델인 RNN을 선정하였다.

선택된 모델에 기상 관측 자료를 학습자료로 부여하고 태양광 발전량 자료를 정답 자료로서 학습시켰다. 학습 시에 자료는 훈련세트와 평가세트로 나뉘고 평가세트는 학습이 완료된 모델의 적합성(훈련세트에 대한 과대적합 혹은 과소적합)을 검증하기 위한 검증세트와 학습과 검증이 완료된 모델의 성능을 평가하기 위한 평가세트로 나뉘었다. 훈련세트는 2017년 1월

부터 2021년 5월까지, 검증세트는 2021년 6월부터 2022년 5월까지, 그리고 평가세트는 2022년 6월부터 2023년 5월로 검증과 평가 기간을 각각 1년씩으로 설정했다. 기상관측자료(OBS)로 학습과 예측을 진행했고, 이는 대조군으로서 reference 값으로 설정했다. 학습된 모델에 1일 전 예보자료세트(foe1)와 2일 전 예보자료세트(foe2)를 만들어 각각 예측을 진행했다. 성능을 평가할 때, 실제 발전량 수치와 비교하기 위해 모델별 예측값을 전처리 전 스케일로 역변환하여 비교하였다.

### 3. 결 과

태양광 발전량 예측 결과는 r-value와 RMSE (Root Mean Squared Error)를 평가 척도로 사용하여 성능을 비교했으며 신뢰도 95%에서 유의함을 확인했다(Chai and Draxler, 2014). 4가지 모델에 관측자료, 1일 전 예보자료, 2일 전 예보자료를 사용해 태양광 발전량 예측값을 생산하였으며, 각각 reference, foe1, foe2로 표기하였다(Figs. 3-4).

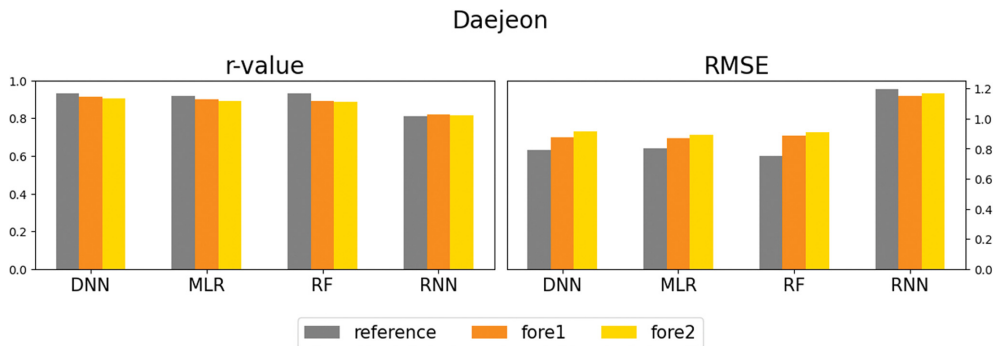


Fig. 3. Comparison of prediction performance by model in Daejeon. All of the r-values are statistically significant at a 95% confidence level, and the unit for RMSE is [MW h<sup>-1</sup>].

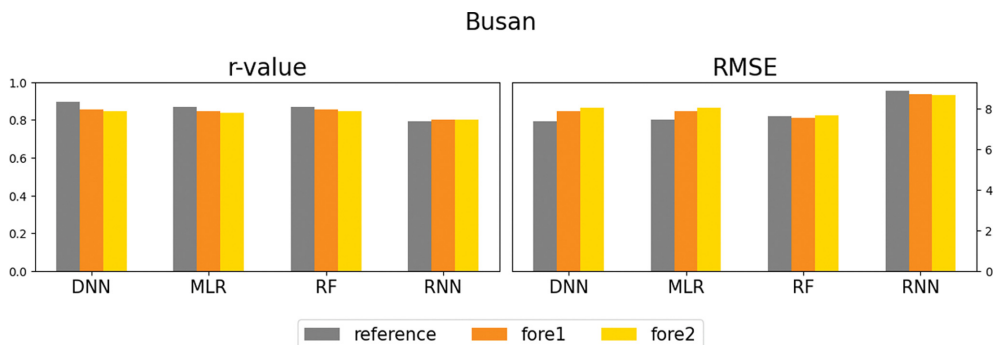
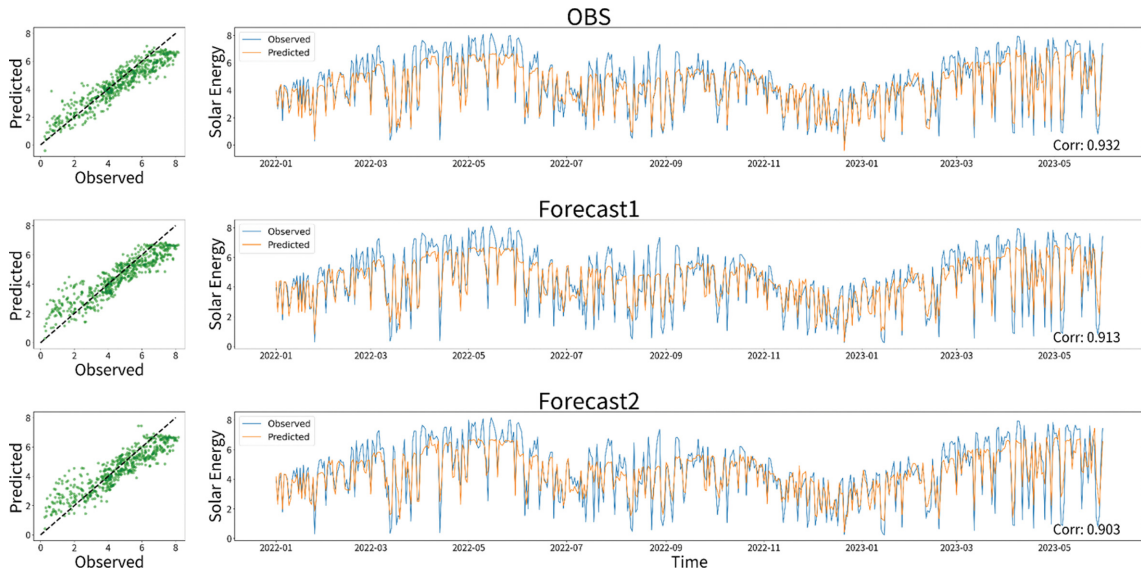


Fig. 4. Same as Fig. 3. but in Busan.



**Fig. 5.** Scatter and time series plots comparing observed and predicted solar power generation using DNN in Daejeon. In the scatter plot, the x-axis represents actual solar power generation, and the y-axis represents predicted solar power generation. The black dotted line signifies the  $y = x$  line, indicating similarity between observed and predicted values. The time series plot displays time on the x-axis solar power generation [ $\text{MW h}^{-1}$ ] on the y-axis. The blue line is observed solar power generation, while the orange line shows predicted solar power generation.

**3.1 대전**

대전의 경우 DNN의 r-value 값이 fore1, fore2에서 각각 0.9130, 0.9025로 가장 높았으며 RNN이 0.819, 0.8139로 가장 낮았다(Fig. 3). RF와 MLR는 fore1에서 각각 0.8914, 0.8980, fore2에서 0.8856, 0.8934로 비슷한 수준을 보였다. fore1의 성능이 fore2보다 더 좋게 예측되었으며 RNN에서는 fore1과 fore2가 reference보다 더 높게 나타나는 현상을 보였다. 예보자료를 통한 예측이 관측자료를 통한 예측과 큰 차이를 보이지 않았고 이는 태양광 발전량 예측에서 예보자료의 효용성을 보여주는 결과라고 할 수 있다.

Figure 5에서 좌측의 산점도를 확인하면 실제 값이 클 때 과소예측이 되고 실제 값이 작을 때 과대예측이 되는 현상이 나타난다. 이는 시계열 자료의 극값 예측에서도 확인할 수 있는데, 예측된 태양광 발전량을 보면 최댓값과 최솟값을 예측하지 못하는 현상이 보인다. 이는 Fig. 1에서의 예보자료가 극값을 정확하게 예측하지 못하는 것이 원인으로 생각된다.

**3.2 부산**

부산의 경우 RF의 r-value 값이 fore1, fore2에서 각각 0.8565, 0.8461로 가장 높았으며 RNN이 0.7996, 0.8002로 가장 낮았다(Fig. 4). MRL와 DNN은 fore1

에서 각각 0.8464, 0.8536, fore2에서 0.8373, 0.8445로 비슷한 수준을 보였다. 하지만 성능이 가장 높은 RF와 MRL, DNN의 성능이 유의미한 차이를 나타내지 않았으며 비슷한 수치의 값을 보여주었다. RNN의 예측성능은 Reference, fore1, fore2모두 유사하게 나타났다. 부산도 대전과 마찬가지로 reference 값과 fore1, fore2 값의 차이가 크게 나지 않았으며 유사한 시계열 개형을 보여주는 것을 보아 예보자료의 효용성을 보여주었다.

Figures 5, 6에서 좌측 산점도를 확인하면 부산이 대전보다 전반적으로 예측의 과소모의와 과대모의가 크게 나타났다. 즉, 시계열 자료에서 Reference, fore1, fore2 모두 최댓값과 최솟값에 대한 예측이 잘 이뤄지지 않고 있음이 두드러지게 나타난다(Fig. 6). 이는 대전과 마찬가지로 예보자료가 극값을 예측하는데 한계가 있다는 것을 보여준다. 추가로 2022년 3월 이후로 부산의 태양광 발전량이 급격하게 증가한 것 또한 원인으로 볼 수 있다.

**3.3 가중치 처리의 효용성**

각 지역에서 전처리 수준에 따른 예측성능 차이를 확인하기 위해, 전체적으로 모델 성능이 좋게 나온 DNN 모델의 단계별 성능을 r-value, RMSE로 비교했다. 이때 시계열 성분의 체계적 성분(추세성분, 계절

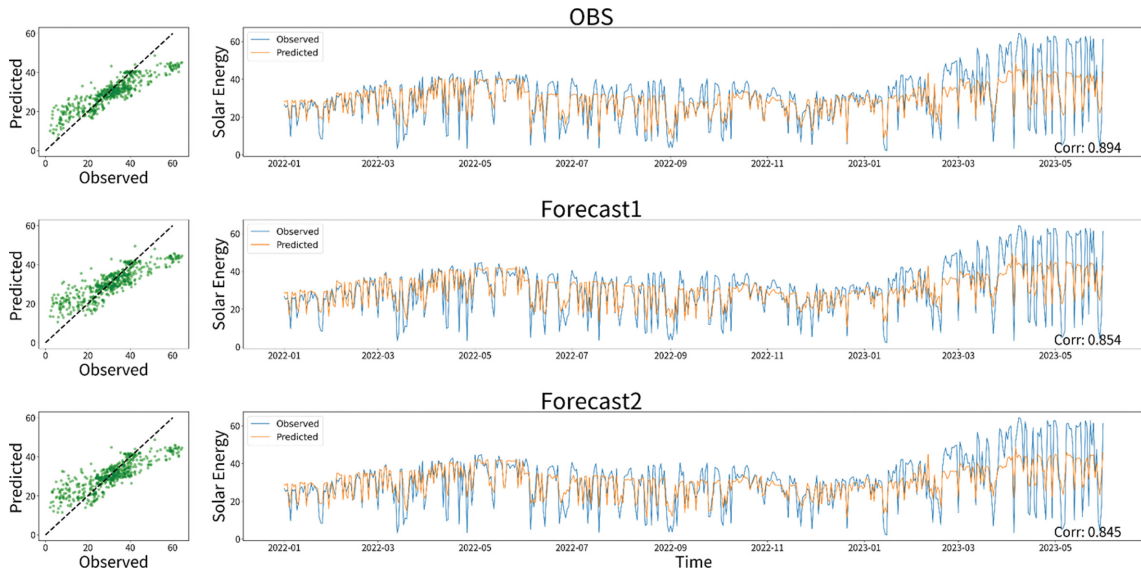


Fig. 6. Same as Fig. 5. but in Busan.

Table 2. Performance of the DNN model on different levels of preprocessing in Daejeon.

Daejeon		(No pre)		(Partial Pre)		(Full Pre)	
		r-value	RMSE	r-value	RMSE	r-value	RMSE
DNN	reference	0.8662	1.2857	0.8823	1.0226	0.9320	0.7908
	fore1	0.8305	1.2727	0.8385	1.1082	0.9130	0.8734
	fore2	0.7840	1.3527	0.8051	1.1993	0.9025	0.9127

Table 3. Performance of the DNN model on different levels of preprocessing in Busan.

Busan		(No Pre)		(Partial Pre)		(Full Pre)	
		r-value	RMSE	r-value	RMSE	r-value	RMSE
DNN	reference	0.7115	16.0775	0.8244	8.6700	0.8943	7.3350
	fore1	0.5788	15.2030	0.7529	9.6221	0.8536	7.8529
	fore2	0.5378	15.3606	0.7221	9.8811	0.8445	8.0305

성분)과 불규칙성분을 합쳐 원자료의 단위로 복원한 후 비교하였다.

대전의 reference는 r-value와 RMSE 두 지표 모두 Full pre, Partial pre, No pre 순으로 성능이 좋게 나타난다(Table 2). reference뿐만 아니라 fore1과 fore2도 모두 reference와 동일한 순서로 성능이 좋은 것을 알 수 있다. 대전의 Full pre의 경우 reference, fore1, fore2 모두 r-value가 0.9를 넘는 우수한 성능을 보여준다. Partial pre보다 Full pre의 성능이 두드러지게 상승한 것을 보아 등분산성 가정을 충족시킨 전처리가 성능을 향상시키는 중요한 요인으로 생각된다.

부산에서도 대전과 같이 모든 지표가 Full pre, Partial

pre, No pre 순으로 성능이 좋게 나타났다(Table 3). 게다가 부산은 대전에 비해 수준에 따른 성능 차이가 더 큰 것을 알 수 있다. 특히 No pre와 Partial pre의 차이가 두드러진다. reference 기준으로 No pre와 Partial pre를 비교해 보면 No pre의 RMSE는 16.0775인데 반해 Partial pre는 8.6700으로 RMSE가 크게 감소했다. 부산에서는 추세와 계절성을 제거한 전처리 과정이 성능을 크게 향상시켰고 Full pre의 가중치 처리도 Partial pre의 전처리 못지않게 r-value와 RMSE의 향상에 도움을 준다는 것을 알 수 있다.

결론적으로 DNN 모델에서 대전, 부산 모두 모든 전처리과정을 포함한 Full pre에서의 성능이 가장 우

수한 것을 알 수 있다. 전처리 수준에 따른 성능 차이가 두드러지는 이러한 결과는 전처리 과정이 기계 학습 기반의 태양광 발전량 예측 성능 향상에 중요한 요소로 작용할 수 있음을 나타낸다.

#### 4. 결론 및 논의

대전과 부산 모두 RNN을 제외한 세 모델에서 예보자료를 이용한 태양광 발전량 예측이 0.85-0.90 수준으로 비슷한 성능을 보였다. 모델의 r-value 뿐만 아니라 예측된 시계열의 개형 역시 reference와 fore1, fore2가 유사하게 예측됐다. 특히 reference 실험에 비해 기상예보자료를 사용한 fore1, fore2에서 태양광 발전량 예측성능이 많이 떨어지지 않는 결과는 예보자료의 효용성이 있음을 시사한다. 또한 이러한 결과는 대전과 부산 모두에서 유사하게 나타나는데 이를 통해 내륙지방과 해안지방의 지역별 기후특성에 관계없이 태양광 발전량 예측에서 예보자료의 효용성이 있음을 보여준다. 다만, RNN의 경우 상대적으로 낮은 0.8 수준의 성능을 보였으며 reference보다 fore1, fore2에서 더 좋은 예측성능을 보였다. 이는 모델의 가중치 처리방식에 따른 학습문제라고 생각된다. 즉, 알고리즘 자체의 중요성보다 회귀문제를 풀 때 회귀의 가정을 고려해 등분산성을 조정하는 전처리 과정이 더 중요하다고 판단된다. 결론적으로 태양광 발전량 예측 모델의 성능 향상에서 잔차 분산 분석의 효용성을 확인한 점에 의의가 있다.

그러나 여전히 자료 자체의 한계점이 존재한다. 첫 번째로 태양광 발전량 자료가 광역시/도 수준의 평균 자료만 제공되기 때문에 기상 자료도 전체 평균을 사용했다. 이 과정에서 각 지점의 기상학적 시그널이 약화되었을 수 있다. 따라서 태양광 발전량의 지점별 자료가 제공된다면 높은 해상도의 기상자료를 사용한 예측성능 향상이 기대된다. 추가로 기상청의 예보자료가 기상 변수별로 3시간 단위, 6시간 단위로 제공되어, 시간 단위의 자료가 아닌 일평균 자료로 변환하여 사용한 점 또한 예측성능의 저하에 기여했을 가능성이 있다. 또한, Figs. 5, 6에서 나타났듯이 현재 기상청의 기상 예보자료는 극값을 과소/과대모의 하는 경향이 있기 때문에, 극값의 예보 성능이 올라갈 경우 태양광 발전량 예측성능도 향상될 여지가 있다. 즉, 자료의 한계점을 보정한 후에는 기계학습 관점에서 성능 향상도 기대해볼 수 있다. 모델 자체가 특정 유형의 자료에 대해 편향되어 있는 가능성을 고려하여, 이를 보정하는 바이어스보정(bias correction)을 실행할 수도 있을 것이다. 이러한 작업이 추가된다면 모델의 공정성과 안정성을 향상시키고 극값 모의 성능 또한 개선할 수 있을 것으로 예상된다.

#### 감사의 글

본 논문의 개선을 위해 좋은 의견을 제시해 주신 두 분의 심사위원께 감사를 드립니다. 이 연구는 한국연구재단(NRF)의 지원(RS-2023-00207866, 2020R1C1C1008631)으로 수행되었습니다.

#### REFERENCES

- Chai, T., and R. R. Draxler, 2014: Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, **7**, 1247-1250, doi:10.5194/gmd-7-1247-2014.
- IPCC, 2021: *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, Intergovernmental Panel on Climate Change, 2391 pp.
- Kang, B.-B., and J.-H. Yun, 2023: A study on the prediction model of photovoltaic power generation using deep learning algorithm. *J. Institute of Electron. Inf. Engineers*, **60**, 119-125, doi:10.5573/ieie.2023.60.2.119.
- Kang, K.-M., Y.-G. Choi, Y.-J. Kim, S.-J. Min, D.-S. Choi, K.-Y. Kim, and D.-Y. Lee, 2021: The classification of climate types and the delineation of their climatic characteristics using new normals (1991~ 2020) in the Republic of Korea. *J. Climate Res.*, **16**, 179-195, doi:10.14383/cr.2021.16.3.179.
- Ki, J.-H., S.-J. Baek, J.-Y. So, H.-G. Eom, and J.-H. Shin, 2023: Solar power generation prediction using machine learning and study of power generation using solar tracking panels. *Trans. Korean Soc. Mech. Eng. B*, **47**, 55-62, doi:10.3795/ksme-b.2023.47.1.055.
- KEA, 2021: 2021 Renewable Energy Distribution Statistics, Korea Energy Agency, Accessed October 23, 2023, [Available online at <https://www.knrec.or.kr/biz/pds/statistic/view.do?no=170>].
- KIER, 2023: Analyzing 2022 Power Generation in South Korea through Electricity Statistics. Accessed December 20, 2023, Korea Institute of Energy Research Tech-policy Platform, [Available online at <https://www.kier.re.kr/tpp/tppBoard/view/25?menuId=MENU00962>].
- Kim, B.-C., S.-H. Jung, M.-S. Kim, J. G. Kim, H.-S. Lee, and S.-S. Kim, 2020: Solar power generation forecasting based on LSTM considering weather conditions. *J. Korean Institute of Intelligent Syst.*, **30**, 7-12, doi:10.5391/jkiis.2020.30.1.7.
- Kim, J.-W., 2019: A solar power prediction scheme based on machine learning algorithm from weather fore-



- casts. *J. Korean Institute of Inf. Technol.*, **17**, 83-89, doi:10.14801/jkiit.2019.17.9.83.
- KPX, 2020: Amendment to the Electricity Market Operation Rules, Accessed November 21, 2023, Korea Power Exchange, [Available online at [https://new.kpx.or.kr/board.es?mid=a10205010000&bid=0030&act=view&list\\_no=68951&tag=&nPage=1](https://new.kpx.or.kr/board.es?mid=a10205010000&bid=0030&act=view&list_no=68951&tag=&nPage=1)].
- Lee, D.-H., and G.-H. Kim, 2019: Deep learning based prediction method of long-term photovoltaic power generation using meteorological and seasonal information. *J. Soc. e-Business Studies*, **24**, 1-16, doi:10.7838/jsebs.2019.24.1.001.
- Lee, K.-H., H.-G. Son, and S. Kim, 2018: A study on solar energy forecasting based on time series models. *J. Appl. Stat.*, **31**, 139-153, doi:10.5351/KJAS.2018.31.1.139.
- Mekhilef, S., R. Saidur, and M. Kamalisarvestani, 2012: Effect of dust, humidity and air velocity on efficiency of photovoltaic cells. *Renewable and Sustainable Energy Rev.*, **16**, 2920-2925, doi:10.1016/j.rser.2012.02.012.
- NOAA, 2020: Despite pandemic shutdowns, carbon dioxide and methane surged in 2020. Accessed July 10, 2023, National Oceanic and Atmospheric Administration, [Available online at <https://research.noaa.gov/2021/04/07/despite-pandemic-shutdowns-carbon-dioxide-and-methane-surged-in-2020/>].
- Son, W.-B., 2023: Solar power generation forecasting technology status for anticipating the expansion of renewable energy deployment. *J. Korean Institute of Communication Sciences*, **40**, 20-26.
- Voyant, C., G. Notton, J.-L. Duchaud, J. Almorox, and Z. M. Yaseen, 2020: Solar irradiation prediction intervals based on Box-Cox transformation and univariate representation of periodic autoregressive model. *Renewable Energy Focus*, **33**, 43-53, doi:10.1016/j.ref.2020.04.001.
- Yu, J.-H., S.-J. Kim, B.-H. Jang, and S.-M. Woo, 2022: Renewable energy prediction market participation type photovoltaics prediction model development. *J. Korean Solar Energy Soc.*, **42**, 1-13, doi:10.7836/kses.2022.42.6.001.