Designing Solar Power Generation Output Forecasting Methods using Time Series Algorithms: Global warming affected weather conditions

EunGyeong Kim[1], M. Shaheer Akhtar[1*], O-Bong Yang[1,2*]

[1]*Graduate School of Integrated Energy-AI, Jeonbuk National University, Jeonju, 54896, Republic of Korea*

[2]*School of Semiconductor and Chemical Engineering, Jeonbuk National University, Jeonju, 54896, Republic of Korea*

Abstract

The present photovoltaic (PV) power generation systems are globally facing the irregularity problem in the distribution of PV generation. In particular, the exact PV power forecasting is critical for grid-connected photovoltaic (PV) systems under unwanted changes in environmental circumstances. The grid energy management, grid operation and scheduling are important factors to forecast the PV power output. Time series analysis is one of the most important aspects of PV output prediction, especially in places (in South Korea) where past solar radiation data or other weather parameters have not been recorded. In this paper, a variety of time-series methods including deep-learning algorithm and machine learning algorithms was used to predict the PV power generation output for quick respond to equipment and panel defects. For designing AI models, the input data were characterized by dividing seasons and choosing the multiple parameters from seasons. In this study, the photovoltaic power generation data was collected from Ansan city, South Korea during January 2017 to June 2021 and the weather data was collected from Suwon city, South Korea during January 2017 to June 2021. In totality, approx. 40,000 hours of operation data from 1.5 MW grid-connected PV system in South Korea was used in this work. PV power generation forecasting was carried out on an hourly basis to test efficacy of various models. Among all models (Holt-Winters, Multivariate Linear Regression, ARIMA, SARIMA, ARIMAX, SARIMAX), LSTM model presented the lowest error rate as compared to other models for quick PV power generation forecasting.

Keywords: Photovoltaic power output, Forecasting methods, Machine learning, Deep learning, Climate change, time series algorithm.

1. Introduction

Past few years, the climate change, global warming, and rising energy demands have prompted the Korean government to look for sustainable energy sources that are both economically and environmentally viable. Excessive industrialization and urbanization increase the power demand, so the power system must become more efficient and stable. When the supply of electricity exceeds the demand, surplus electricity is generated, which is uneconomic. Therefore, insufficient power supply is to meet the demand, a power failure occurs, which may result in a blackout. However, due to the irregular power generation problem of photovoltaic systems, there is a problem that engineers cannot quickly respond to equipment and panel defects[1]. Few studies have been done to PV power generation forecasting [2-4]. Therefore, predicting the amount of PV power generation in advance and stabilizing the power supply increases the efficiency of PV power generation.

The present PV power generation systems are still shown numerous faults and dependencies which normally come from solar irradiance. The electrical power generated is influenced by a number of factors including the quality of the PV cells, the type of solar cells used, the electrical circuit of the module, the angle of incidence, weather conditions, and other parameters. Mainly the temperature of the solar cell in a PV system affects the amount of power produced [5]. Some reports are available to define PV output power forecasting based on the weather classification [6-9]

It is essential to predict PV power output in order to quickly respond to panel and equipment defects. It is demanded to develop model for improving the PV power generation using the artificial intelligence (AI) including machine learning, deep learning etc. Lee et al analyzed the high PV power generation forecasting model by considering meteorological factors[10]. Lee's group selected the factors such as date and time, temperature, wind speed, wind direction, humidity, total cloud cover, and solar insolation for forecasting the PV power generation and meteorological variables. Mellit et al. utilized more than a year of data to anticipate the electricity produced by

a 50 W PV plant using two artificial neural networks [11]. Most of models are focused on the short-term (three-days-ahead) forecasting [12].

For PV power generation forecast, the common AI models are statistical models and Deep learning models[2-5, 13-19]**.** Several mathematical equations in statistical models are used to extract patterns from input data. In general, statistical techniques can be divided two groups: times-series and machine learning (ML) based models [20] . Time-series models such as ARMA [15-18], ARIMA [4, 18, 21], SARIMA [19, 22] and holt-winters model [14, 20] are popularly used because of their exponential smoothing. Nowadays, the machine learning and deep learning models are widely used to predict PV power output forecasting [3, 23-26]. To overcome the aforementioned obstacles, fresh and sophisticated procedures must be used to achieve legitimate and reliable results. Several researchers have reported time series models for PV power generation forecasting using seasons such as 4 seasons [2, 27-29], and the sunny day, cloud day and rainy day [7, 10, 13, 30, 31]. Last few years, deep learning (DL) model driven approaches have been experienced a great deal of interest in the time-series PV generation prediction fields because DL can enable to interpretate highly noisy and incorporate the irrelevant datasets [32]. Among several DL based model, long short-term memory (LSTM) approach has been used for the PV power prediction as LSTM delivers excellent result accuracy with time-series datasets due to remembering the previous information via a complex memory [33]. It is assumed that using LSTM approach, highly efficient model over statistical model can construct. To advance the PV power forecast accuracy, the data was classified into four sub datasets based on the annual seasons.

In this work, a novel PV power generation forecast model using time series algorithms is developed using six statistical and one deep learning time series models. The main reason is to use time series models in this study because they are affected by period, trend, and seasonality [28, 34]. Herein, in order to make database, we divided the data using four seasons

(spring, summer, fall and winter) collected data from two cities of Korea rather than a method using a specific weather. In addition, the seasons are further divided into three methods and conduct the forecasting by seven AI models.

2. Methodology

2.1. Data description

In this work, we have chosen the data from PV power generation plant located in South Korea, as shown in Fig. 1. All PV power generation data were collected from Ansan city, and the weather data was collected from Suwon city, South Korea during January 2017 to June 2021. Furthermore, the collected PV data on various variables (temperature, sunlight, insolation, clouds (total cloudiness, low-middle cloudiness, minimum cloud height), fine dust ([Particulate Matter (PM)] or PM10, PM2.5), precipitation, snow cover)) were used and analyzed in 3 methods (month, seasons, revised season affected by global warming). The seasons were divided into one dataset, the model's trend, seasonal, cycle change. In Fig.1 (a)~(b), the PV power generation output data collected from Yeonseong Water Purification Plant Solar Power, Ansan city which had a capacity of 1.49 MW and consisted of a total 40,898 data. For the weather data, there was no meteorological data provided by Yeonseong Water Purification Plant Solar Power, Ansan city. We have chosen the weather data from Suwon city, which is just 20.15 km away in a straight line to Ansan city. From Suwon city, the collected 408,980 weather data were used on temperature, sunlight, insolation, clouds (total cloudiness, low-middle clouds, minimum cloud height), fine dust (PM10, PM2.5), snow, and precipitation. Before performing the model training, the collected season data (in Fig. 1. (c)) was divided into three cases. Case 1 was referred to seasons by months (like spring, summer, fall, winter), Case 2 addressed the solar term-based subdivision of seasons (according to lunar calendar), and Case 3 described the season affected by global warming. The selection of Case 3 was based on the overall temperature increase due to global warming. According to weather

forecasting department, due to global warming, the significant rise in temperature was recorded in Korean weather for examples, more than 5°C temperature rise in spring, more than 20°C rise in summer, in fall the average daily temperature fell below 20°C and temperature normal in winter. All collected data were taken from government open portal systems managed by Ministry of the Interior and safety, Korea meteorological administration (KMA), and Korea Environment Corporation.

The dataset was first normalized using minmax normalization method, and then omitted the outliers and missing values was filled by using Python's fillna function. After the model was trained, and the error rates were compared to propose a variable and season segmentation method for efficient model construction.

2.2. Time series data procedure

Fig. 2 describes the entire time series modeling for PV power generation forecasting. For each modelling, 80% of the total data was used as training data and 18% was used as validation data. Afterward, only 2% test data was used to predict 5 days. Each dataset (Case 1-3) was resampled using the seasons for models. To predict highly accurate model, we prepared three cases based on several parameters. Only PV data was considered as Case A, and Case B included the PV data, temperature, insolation, sunshine and Case C described the PV data with Pearson correlation parameters. For example, the multiple variables like temperature, fine dust, clouds, were used to models such as ARIMAX, Multivariate Linear Regression, ARIMAX, SAIRMAX, LTSM. By reflecting these seasonal characteristics, the variables to be used for each season were compared with the Pearson correlation coefficient for comparing the linear relationship and afterward applied to the model.

2.3 Pearson Correlation Coefficient (PCC)

The independent variable PV data were calculated by PCC to select the dataset of Case C. PCC generally calculated from the corr function of the Pandas library of Python [34]. The

obtained PCC results are displayed in Figs. 3-5 in terms of Case 1, Case 2 and Case 3. It is known that the ≥0.7, ≥0.3 and ≤0.1 are referred to strong PCC, medium PCC and weak PCC, respectively. From Figs. 3-5, variables such as temperature, sunshine, insolation, cloud level, cloud amount, and cloud ceiling are shown over 0.1 PCC values in spring and winter seasons. It is seen in Case 2 and Case 3 that the independent variables such as dust PM10 having the PCC values of 0.14. Based on PCC values, several variables were characterized into Case 1, Case 2, and Case 3, as shown in Table 1. Therefore, PCC analysis is helpful to choose the independent variables on PV power generation output.

2.4. Performance

$$mMAPE = \begin{cases} \frac{100}{n}\sum_{t=1}^{n}|A_t - F_t|, & |A|_{max} < 1 \\ \frac{100}{n}\sum_{t=1}^{n}\frac{|A_t - F_t|}{|A|_{max}}, & else \end{cases} \quad (1)$$

In the performance evaluation of the model, the model was evaluated using the maximum mean absolute percentage error (mMAPE) index to prevent infinity (Inf) when the actual value was 0 [35]. Using the equation 1, the mMAPE was calculated after simulating two cases. $|A|_{max}$ is the maximum absolute value of the actual data. When the value of $|A|_{max}$ is less than 1, the difference between the actual data and the predicted data is used as the evaluation value, which considerably avoid the slight difference between the actual value and the predicted data. Therefore, the predicted data was amplified by $|A|_{max}$. If $|A|_{max}$ is greater than 1, the error value is calculated by dividing the difference between the actual data and the predicted data by $|A|_{max}$. With MAPE, the magnitude of the error is affected by the difference between the actual and predicted values.

3. Models

3.1. Holt-Winter Model

According to the studies in refs. [36] and [37], the Holt-Winters model is a simple, excellent accuracy and good predictability for energy generation forecast. The Holt-Winters model is highly useful for a constant periodicity and seasonal component. The associated issues like exponential smoothing model make obstacle to estimate the trend and seasonality. However, the use of Holt-Winters model considerably overcome these issues [38]. Holt-Winters model is normally divided into two types according to the data characteristics such as multiplicative seasonal model and additive seasonal model. The multiplicative seasonal model deals the increase in variance of data with the passage of time. The additive seasonal model is referred to highly suitable when the data presents a linear and uniform increasing trend. In this work, PV output data from Case A was applied to suitable additive seasonal (Holt-Winters) model. It was found that PV power output data fluctuations are constant between 0 to 1400 kWh. The additive seasonal model can be expressed by following expressions;

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}), \qquad (2)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}, \qquad (3)$$

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}, \qquad (4)$$

$$F_{t+m} = (L_t + b_t m) + S_{t-s} \qquad (5)$$

Where, L_t = time series mean level at time t, b_t = time series trend component at time t, S_t = time series seasonal component at time t, F_{t+m} = predicted value of time t+m predicted at time t, s = length of seasonal component, and α, β, γ = smoothing parameters.

The time series components such as L_t, b_t, **and** S_t. in Holt-Winters model are predicting PV generation for Case A i.e., seasonal, as presented in Table 5. This model is only use for seasonal not for weather factors (as error rate is very high). Therefore, it is necessary to search another model for predicting the PV power generation which is suitable for both seasonal and weather conditions too.

3.2. Multivariate Linear Regression Model (MLP)

The Multivariate Linear Regression Model (MLP) model is a statistical technique that predicts a independent variable using multiple dependent variables. The purpose of MLP model is developed the linear relationship between multiple dependent variables and target variables. The variables in MLP model are described as;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdot + \beta_p x_{ip} + \epsilon, \quad (6)$$

where, y_i = dependent variable, x_i = explanatory variables, β_0 = y-intercept (constant term), β_p = slope coefficients for each explanatory variable, ϵ = the model's error term (also known as the residuals)

The MLP model generally provide the information about the correlation between the dependent variable and the independent variable. In this work, the dataset of Case 1, Case 2, Case 3, Case B and Case C are used for this model, but Case A cannot use because it is used only for univariate models. From table 5, the error rate is very high, which means not suitable for predict the PV power generation using seasonal. Thus, MLP model is suitable for weather factors.

3.3. ARIMA Model

ARIMA model is consist of the autoregressive (AR) model, moving average (MA) model and difference. The model parameters p, d, q are used in ARIMA model in which p is the AR model's time lag parameter, d is the difference parameter for time lag, and q is MA model's time lag parameter. The time series model is measured values at time (t) and time lag (t-h). In this case, it becomes difficult to apply a general regression operation, so we need to proceed with the difference. Difference refers to the process of replacing the sequence obtained by subtracting the data at (t-h) from the data at t with the subject of time series analysis [39].

Table 2 shows the p,d,q values of ARIMA using the dataset from Case 1, Case 2, and Case 3. To

know the values of p, d, and q, the 'auto.arima' function of the forecast library of RStudio was used. ARIMA model was performed after setting the p, d, q ranging from 0 to 5, and the variable with the lowest AIC (Akaike's Information Criterion). In fall season, the p, d, q values in Case 1, Case 2 and Case 3 are similar because the weather factors are stable. However, the significant difference in p, d, q values for spring, summer and winter are observed, as seen in Table 2. ARIMA model is good for Case 1, Case 2, Case 3, and Case A. Even though, it shows good results in different seasons and weather but still presents high error rate. However, the error rate in ARIMA model is lower than those of above models.

3.4. SARIMA Model

SARIMA model integrates the AR model and the MA model including the seasonal variables. SARIMA model allows to measure the p, d, q and P, D, Q, S values. P, D, Q, S values in SARIMA are more appropriate wherein P is the AR model's seasonal parameter, Q is MA model's seasonal parameter, D is difference seasonal parameter and S is seasonal parameter in dataset. SARIMA model is also a univariate model, thus only Case 1, Case 2, Case 3 and Case A were used.

3.5. ARIMAX Model

ARIMAX is an extension of the ARIMA model to predict the PV power output value using multiple parameters and p, d, q values, as shown in table 2. In this model, the multiple input parameters are based on PCC results in table 1. ARIAMX model is shown the good suitability to season data of Case 1, Case 2, Case 3 and parameter data of Case B and Case C. The error rate in ARIMAX model is similar to SARIMA model.

3.6. SARIMAX Model

SARIMAX model is an extension of the SARIMA/ARIMAX to predict the PV power

data using seasonal and exogenous parameters. SARIMAX p, d, q and P, D, Q, S values are found same as in SARIMA (table 3). Multiple parameters in Table 1 were used as exogenous parameters in SARIMAX model. This model is good for season data like Case 1, Case 2, Case 3 and parameter data of Case B and Case C. It can be seen that all above models are not completely suitable for seasonal (Case 1, Case 2, Case 3) and weather factors (Case A, Case B and Case C).

3.7. Long-Short Term Memory (LSTM) Model

LSTM model is a Neural Network (NN) model in which the whole data can process and predict the future data. The most representative time series NN model is recursive neural network (RNN) but it has gradient vanishing problem. Gradient vanishing is a problem in which old data cannot affect the forecasting values. LSTM model is solving the gradient vanishing problem using three gates such as forget, input, and output gates. The window value and horizon value are required for performing the LSTM model. The window value is a variable that sets number of previous data to predict. Horizon is a variable that sets how many lags to predict the value. Herein, the horizon value is set to 1. Window values are calculated from the average daytimes. The average daytime for each season was calculated using the API for sunset and sunrise times of the Korea Astronomy and Space Science Institute. The window values are summarized in Table 4. The error rate in this model is the lowest as compared to other models used for all cases. Moreover, our model presents the lowest error rate as compared to reported AI models used for forecasting the PV generation in various regions (worldwide), as seen in Table 6.

4. Results and discussions

Among all used models, LSTM model has exhibited the good accuracy along with an excellent correlation coefficient (R^2) of 0.943 and average mMAPE of 5.79 for seasonal and

weather factors. Based on data, Fig. 6 depicts the top 3 models R^2 graphs between actual PV power output data and predict PV power data for seasonal and weather factors considering 5 days. From Fig. 6, the R^2 values are getting bigger from Case 1 to Case 3. In this study, the highest R^2 value of 0.922 records in Case 1 Spring. In Case 3 spring and Case 3 winter, the highest R^2 of 0.944 and 0.943 are obtained respectively. However, the fall and winter in Case 1 present the similar R^2 values. This observation clearly demonstrates the low error rates as the observed high R^2 values. It is well known that R^2 value is ranging from 0 to 1. The high and low error rate is decided by evaluating the R^2 value. The high error rate means R^2 value close to 0, and low error rate means R^2 value nearly 1. For all Cases, herein LSTM model displays over 0.9 R^2 value. The high R^2 values in our results are comprehensively reflected the lower error rate of the system.

Fig. 7 and Fig. 8 displays the graph between actual PV power output data and top 3 model's forecasting data for 5 days. The data gap between actual data and predict data are smaller when we compare Case 1 to Case 3. From these figs., the lowest error rate is observed by C-LSTM model as compared to other models. This explains the seasonal and cycle must be considered rather than simple datasets. Case B presents the lower error rate than that of Case C. It is seen that when large number of variables are used in Case C, the complexity of the model increases, and the error rate rises. Thus, Case B with three variables result in the lowest error rate. The reduction in error rate of graph for actual and predict value is needed to find the good accuracy model. In this regard, the hyperparameter (activation, optimization method) for NN models, optimization number of parameters, study of ensemble model (bagging, stacking, tune model) would be considered in future to predict highly reliable and quick model for PV power generation forecasting.

5. Conclusion

In summary, the PV power output forecast with high accuracy was studied by using seven AI models. In this work, the input data were characterized by dividing seasons and choosing the multiple parameters from seasons for the designing time series models. PV power generation data and weather data were collected from Ansan city and Suwon city during January 2017 to June 2021, respectively. PV power generation forecasting was evaluated on an hourly basis to test the efficacy of various models. As compared to all models (Holt-Winters, Multivariate Linear Regression, ARIMA, SARIMA, ARIMAX, SARIMAX). LSTM model showed the lowest error rate for quick PV power generation forecasting in seasonal and weather factors. In future, the tuning of current LSTM model would carry out by using finetuning and ensemble model to find best efficiency.

References

[1] Chow, S.K., E.W. Lee, and D.H. Li, *Short-term prediction of photovoltaic energy generation by intelligent approach.* Energy and Buildings, 2012. **55**: p. 660-667.

[2] Frederiksen, C.A.F. and Z. Cai, *Novel machine learning approach for solar photovoltaic energy output forecast using extra-terrestrial solar irradiance.* Applied Energy, 2022. **306**: p. 118152.

[3] Jiang, Y., L. Zheng, and X. Ding, *Ultra-short-term prediction of photovoltaic output based on an LSTM-ARMA combined model driven by EEMD.* Journal of Renewable and Sustainable Energy, 2021. **13**(4): p. 046103.

[4] Das, S., *Short term forecasting of solar radiation and power output of 89.6 kWp solar PV power plant.* Materials Today: Proceedings, 2021. **39**: p. 1959-1969.

[5] Al-Nimr, M.d.A., S. Kiwan, and H. Sharadga, *Simulation of a novel hybrid solar photovoltaic/wind system to maintain the cell surface temperature and to generate electricity.* International Journal of Energy Research, 2018. **42**(3): p. 985-998.

[6] Chen, C., et al., *Online 24-h solar power forecasting based on weather type classification using artificial neural network.* Solar energy, 2011. **85**(11): p. 2856-2870.

[7] Shi, J., et al., *Forecasting power output of photovoltaic systems based on weather classification and support vector machines.* IEEE Transactions on Industry Applications, 2012. **48**(3): p. 1064-1069.

[8] Jung, Y., et al., *Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea.* Journal of Cleaner Production, 2020. **250**: p. 119476.

[9] Konstantinou, M., S. Peratikou, and A.G. Charalambides, *Solar photovoltaic forecasting of power output using lstm networks.* Atmosphere, 2021. **12**(1): p. 124.

[10] LEE, S. and J. KIM, *SPV dominant parameter estimation exploration and visualization from weather and prediction model.* The Korean Institute of Electrical Engineers, 2021: p. 464-465.

[11] Mellit, A., S. Sağlam, and S.A. Kalogirou, *Artificial neural network-based model for estimating the produced power of a photovoltaic module.* Renewable Energy, 2013. **60**: p. 71-78.

[12] Su, D., E. Batzelis, and B. Pal. *Machine learning algorithms in forecasting of photovoltaic power generation*. in *2019 International Conference on Smart Energy Systems and Technologies (SEST)*. 2019. IEEE.

[13] Liu, L., et al., *Prediction of short-term PV power output and uncertainty analysis.* Applied energy, 2018. **228**: p. 700-711.

[14] Kanchana, W. and S. Sirisukprasert. *PV Power Forecasting with Holt-Winters Method*. in *2020 8th International Electrical Engineering Congress (iEECON)*. 2020. IEEE.

[15] Huang, R., et al. *Solar generation prediction using the ARMA model in a laboratory-level micro-grid*. in *2012 IEEE third international conference on smart grid communications (SmartGridComm)*. 2012. IEEE.

[16] Hassanzadeh, M., M. Etezadi-Amoli, and M. Fadali. *Practical approach for sub-hourly and hourly prediction of PV power output*. in *North American Power Symposium 2010*. 2010. IEEE.

[17] Chu, Y., et al., *Short-term reforecasting of power output from a 48 MWe solar PV plant.* Solar Energy, 2015. **112**: p. 68-77.

[18] Raza, M.Q., M. Nadarajah, and C. Ekanayake, *On recent advances in PV output power forecast.* Solar Energy, 2016. **136**: p. 125-144.

[19] Vagropoulos, S.I., et al. *Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting*. in *2016 IEEE International Energy Conference (ENERGYCON)*. 2016. IEEE.

[20] Ahmed, R., et al., *A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization.* Renewable and Sustainable Energy Reviews, 2020. **124**: p. 109792.

[21] Li, Y., Y. Su, and L. Shu, *An ARMAX model for forecasting the power output of a grid connected photovoltaic system.* Renewable Energy, 2014. **66**: p. 78-89.

[22] Dubey, A.K., et al., *Study and analysis of SARIMA and LSTM in forecasting time series data.* Sustainable Energy Technologies and Assessments, 2021. **47**: p. 101474.

[23] Lorenz, E., et al., *Qualified forecast of ensemble power production by spatially dispersed grid-connected PV systems.* Measurement, 2007: p. 1-7.

[24] Mirzapour, F., et al., *A new prediction model of battery and wind-solar output in hybrid power system.* Journal of Ambient Intelligence and Humanized Computing, 2019. **10**(1): p. 77-87.

[25] Chen, S., H. Gooi, and M. Wang, *Solar radiation forecast based on fuzzy logic and neural networks.* Renewable energy, 2013. **60**: p. 195-201.

[26] Shang, C. and P. Wei, *Enhanced support vector regression based forecast engine to predict solar*

power output. Renewable energy, 2018. **127**: p. 269-283.
[27] de Jesús, D.A.R., et al. *Solar pv power prediction using a new approach based on hybrid deep neural network*. in *2019 IEEE Power & Energy Society General Meeting (PESGM)*. 2019. IEEE.
[28] Gao, M., et al., *Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM.* applied sciences, 2019. **9**(15): p. 3192.
[29] Zhou, Y., et al., *Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine.* Energy, 2020. **204**: p. 117894.
[30] Islam, S.Z., et al., *Photovoltaic modules evaluation and dry-season energy yield prediction model for NEM in Malaysia.* Plos one, 2020. **15**(11): p. e0241927.
[31] Mandal, P., et al., *Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques.* Procedia Computer Science, 2012. **12**: p. 332-337.
[32] Mishra, M., et al., *Deep learning in electrical utility industry: A comprehensive review of a decade of research.* Engineering Applications of Artificial Intelligence, 2020. **96**: p. 104000.
[33] Yu, Y., J. Cao, and J. Zhu, *An LSTM short-term solar irradiance forecasting under complicated weather conditions.* IEEE Access, 2019. **7**: p. 145651-145666.
[34] Harvey, A.C. and N. Shephard, *10 Structural time series models.* 1993.
[35] Shin, K.-H., et al., *Estimation method of predicted time series data based on absolute maximum value.* Journal of Energy Engineering, 2018. **27**(4): p. 103-110.
[36] Cho, C., *Forecasting the Cement Traffic Volume at the Port of Donghae.* Korea Logistics Review, 2008. **18**(1): p. 33-53.
[37] Billah, B., et al., *Exponential smoothing model selection for forecasting.* International journal of forecasting, 2006. **22**(2): p. 239-247.
[38] Kim, J.-T., *Forecasting number of student by Holt-Winters additive model.* Journal of the Korean Data and Information Science Society, 2009. **20**(4): p. 685-694.
[39] Nielsen, A., *Practical time series analysis: Prediction with statistics and machine learning*. 2019: O'Reilly Media.
[40] Sharadga, H., S. Hajimirza, and R.S. Balog, *Time series forecasting of solar power generation for large-scale photovoltaic plants.* Renewable Energy, 2020. **150**: p. 797-807.
[41] Maitanova, N., et al., A machine learning approach to low-cost photovoltaic power prediction based on publicly available weather reports. energies, 2020. **13**(3): p. 735.
[42] Wang, K., X. Qi, and H. Liu, A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Applied Energy, 2019. **251**: p. 113315.
[43] Wang, F., et al., A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Conversion and Management, 2020. **212**: p. 112766.
[44] Zang, H., et al., Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. International Journal of Electrical Power & Energy Systems, 2020. **118**: p. 105790.

Figures and Tables

Table 1. Summary of Pearson Correlation Coefficients (PCC) for Case 1, Case 2, and Case 3.

Case	Seasons			
	Spring	Summer	Fall	Winter
1	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling, Rain	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling, Rain	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling, Dust PM10
2	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling Rain	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Rain	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling Dust PM10	Temperature, Insolation, Sunshine, Cloud amount Cloud level, Cloud ceiling,
3	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling, Rain	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Rain	Temperature, Insolation, Sunshine, Cloud amount, Cloud level, Cloud ceiling, Dust PM10	Temperature, Insolation, Sunshine, Cloud level, Cloud ceiling, Cloud amount, Snow

Table 2. p, d, q values from ARIMA model.

Case	Spring (p,d,q)	Summer (p,d,q)	Fall (p,d,q)	Winter (p,d,q)
1	2,1,5	4,1,4	3,0,5	5,1,1
2	4,1,3	5,1,3	3,0,5	4,0,5
3	2,1,5	4,1,5	3,0,3	4,0,5

Table 3. p, d, q and P, D, Q, S values from SARIMA model

Case	Spring (p,d,q) (P,D,Q,S)	Summer (p,d,q) (P,D,Q,S)	Fall (p,d,q) (P,D,Q,S)	Winter (p,d,q) (P,D,Q,S)
1	(2,1,5) (2,0,0,24)	(4,1,4) (2,0,1,24)	(3,0,5) (1,0,0,24)	(5,1,1) (0,0,1,24)
2	(4,1,3) (1,0,0,24)	(5,1,3) (1,0,2,24)	(3,0,5) (1,0,0,24)	(4,0,5) (0,0,1,24)
3	(2,1,5) (2,0,1,24)	(4,1,5) (1,0,0,24)	(3,0,3) (2,0,2,24)	(4,0,5) (0,0,0,24)

Table 4. Window average values of daytime for Case 1, Case 2 and Case 3.

Case	Spring	Summer	Fall	Winter
1	13 hours	14 hours	11 hours	10 hours
2	12 hours	14 hours	12 hours	10 hours
3	12 hours	14 hours	12 hours	9 hours

Table 5. Comparison of the error rate results Forecast with the actual data using 7 models (MLP, Holt-winters, ARIAM, SARIMA, ARIMAX, SARIMAX, LSTM)

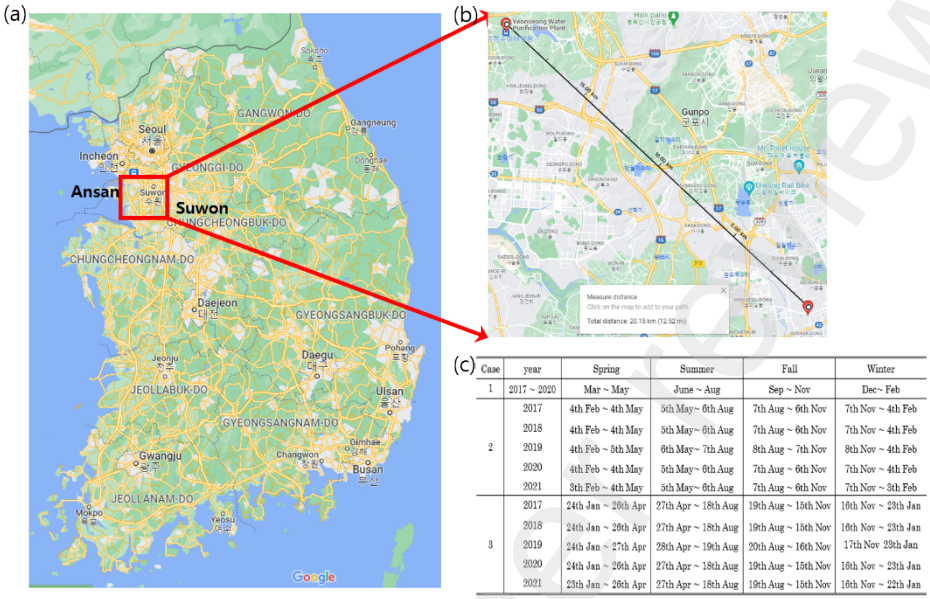
Seasonal Case	Parameter Case	**Model**	**Spring**	**Summer**	**Fall**	**Winter**
1	A	Holt winters	12.53	10.96	18.18	13.30
		ARIMA	7.23	7.89	8.93	8.15
		SARIMA	7.23	7.89	8.93	8.16
	B	Multivariate	15.02	14.95	16.68	15.41
		ARIMAX	7.22	8.06	9.15	8.00
		SARIMAX	7.35	8.12	9.11	8.00
		LSTM	5.61	7.21	6.73	5.90
	C	Multivariate	15.11	15.56	16.89	14.91
		ARIMAX	6.74	15.95	11.23	9.57
		SARIMAX	7.29	9.85	9.19	8.20
		LSTM	5.45	7.00	6.83	5.46
2	A	Holt winters	12.45	11.12	13.47	15.05
		ARIMA	7.46	7.82	8.96	8.67
		SARIMA	7.54	7.83	8.95	8.68

	B	Multivariate	16.03	15.71	16.15	14.99
		ARIMAX	7.74	8.00	9.32	9.41
		SARIMAX	7.73	8.04	9.30	9.34
		LSTM	5.99	6.17	7.06	5.88
	C	Multivariate	15.93	15.61	16.06	14.50
		ARIMAX	7.62	8.07	9.40	9.50
		SARIMAX	7.62	8.07	9.40	9.50
		LSTM	7.16	6.61	8.07	6.62
3	A	Holt winters	11.86	11.48	14.53	13.01
		ARIMA	7.17	8.25	8.42	7.95
		SARIMA	7.32	8.22	8.46	7.95
	B	Multivariate	15.07	15.37	15.32	15.39
		ARIMAX	7.13	8.22	8.17	8.47
		SARIMAX	7.20	8.24	8.17	8.47
		LSTM	5.34	6.89	5.70	5.24
	C	Multivariate	15.20	15.36	15.54	14.87
		ARIMAX	7.13	8.24	8.22	8.40
		SARIMAX	7.12	8.24	8.22	8.40
		LSTM	5.55	6.70	5.90	5.57

Table 6.

Year	Location	Horizon	Model	Error rate	Ref.
2021	northern region India	15 mins 30 mins	ARIMA	9.5% 12.6%	[4]
2020	South China	3 hours	BI-LSTM	6.1% ~ 10.2%	[40]
2020	Germany	1 day	LSTM	12%	[41]
2019	Alice Springs (Australia)	1 day	LSTM, CNN, C-LSTM	8% ~ 11.2%	[42]
2020	Nevada (USA)	1 day	LSTM-RNN	6.29%	[43]
2020	Alice Springs (Australia)	1 day	CNN (ResNet and DenseNet)	18%, 15%	[44]
2022	Korea	1 hour	LSTM	5.24 ~ 6.70%	This work

Figures



Case	year	Spring	Summer	Fall	Winter
1	2017 ~ 2020	Mar ~ May	June ~ Aug	Sep ~ Nov	Dec~ Feb
2	2017	4th Feb ~ 4th May	5th May~ 6th Aug	7th Aug ~ 6th Nov	7th Nov ~ 4th Feb
	2018	4th Feb ~ 4th May	5th May~ 6th Aug	7th Aug ~ 6th Nov	7th Nov ~ 4th Feb
	2019	4th Feb ~ 8th May	6th May~ 7th Aug	8th Aug ~ 7th Nov	8th Nov ~ 4th Feb
	2020	4th Feb ~ 4th May	5th May~ 6th Aug	7th Aug ~ 6th Nov	7th Nov ~ 4th Feb
	2021	3th Feb ~ 4th May	5th May~ 6th Aug	7th Aug ~ 6th Nov	7th Nov ~ 3th Feb
3	2017	24th Jan ~ 26th Apr	27th Apr ~ 18th Aug	19th Aug ~ 15th Nov	16th Nov ~ 23th Jan
	2018	24th Jan ~ 26th Apr	27th Apr ~ 18th Aug	19th Aug ~ 15th Nov	16th Nov ~ 23th Jan
	2019	24th Jan ~ 27th Apr	28th Apr ~ 19th Aug	20th Aug ~ 16th Nov	17th Nov 23th Jan
	2020	24th Jan ~ 26th Apr	27th Apr ~ 18th Aug	19th Aug ~ 15th Nov	16th Nov ~ 23th Jan
	2021	23th Jan ~ 26th Apr	27th Apr ~ 18th Aug	19th Aug ~ 15th Nov	16th Nov ~ 22th Jan

Fig. 1. (a) photograph of South Korea map and (b) selected sites where PV power generation data collected. (c) Summary of data selection in different seasons.

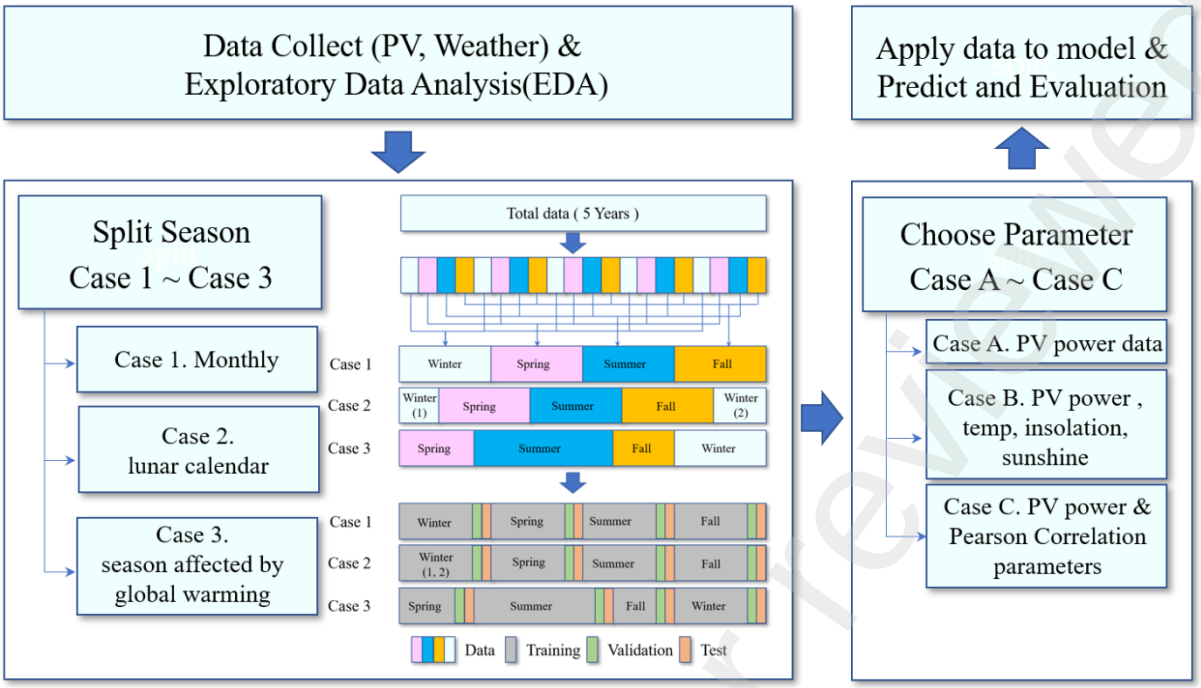


Fig. 2. Schematic of the time series procedure. [All collected data were taken from government open portal systems managed by Ministry of the Interior and safety, Korea meteorological administration (KMA), and Korea Environment Corporation]

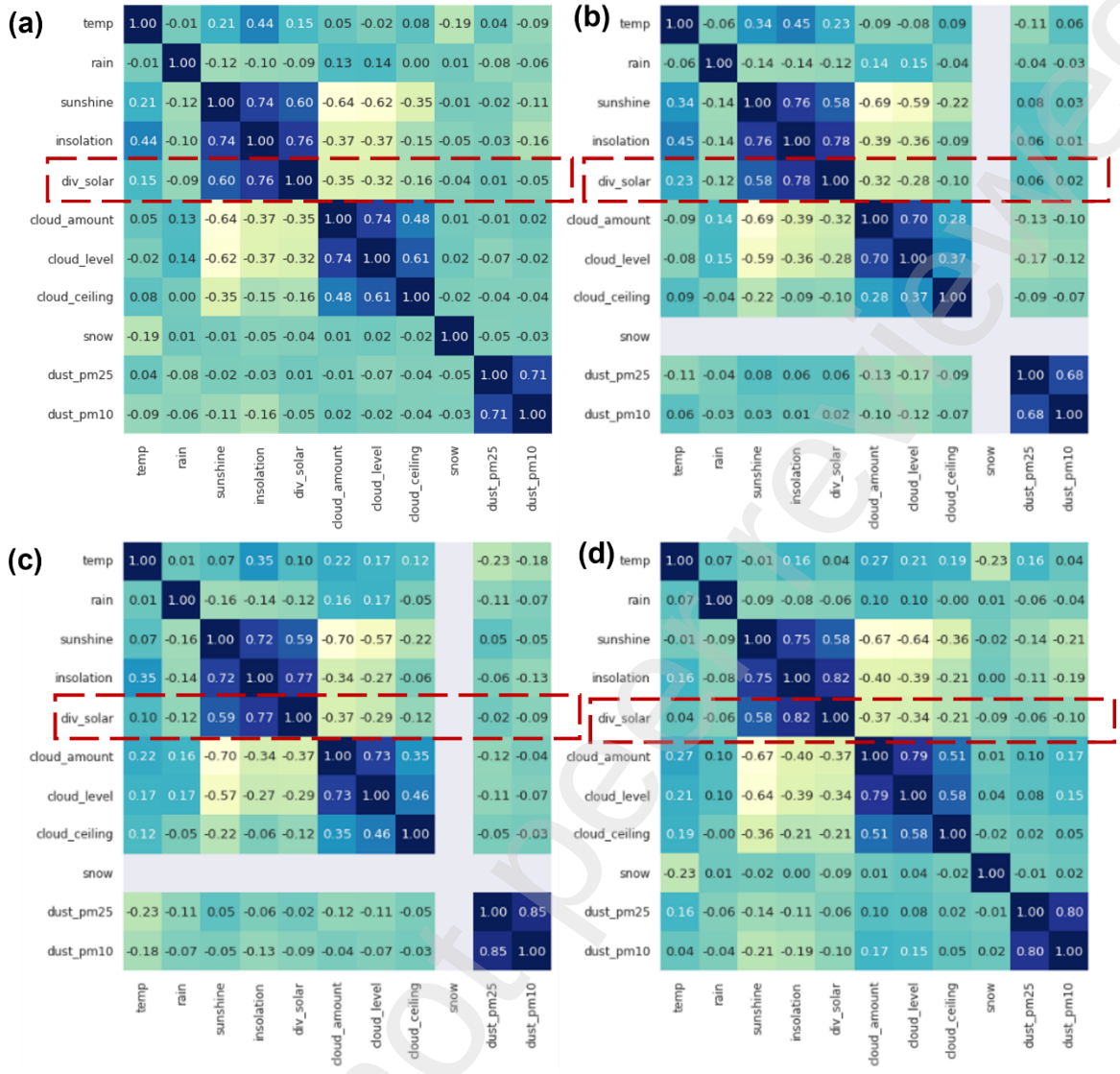


Fig. 3. Pearson correlation coefficients for Case 1 in different seasons as (a) spring, (b) summer, (c) fall and (d) winter.

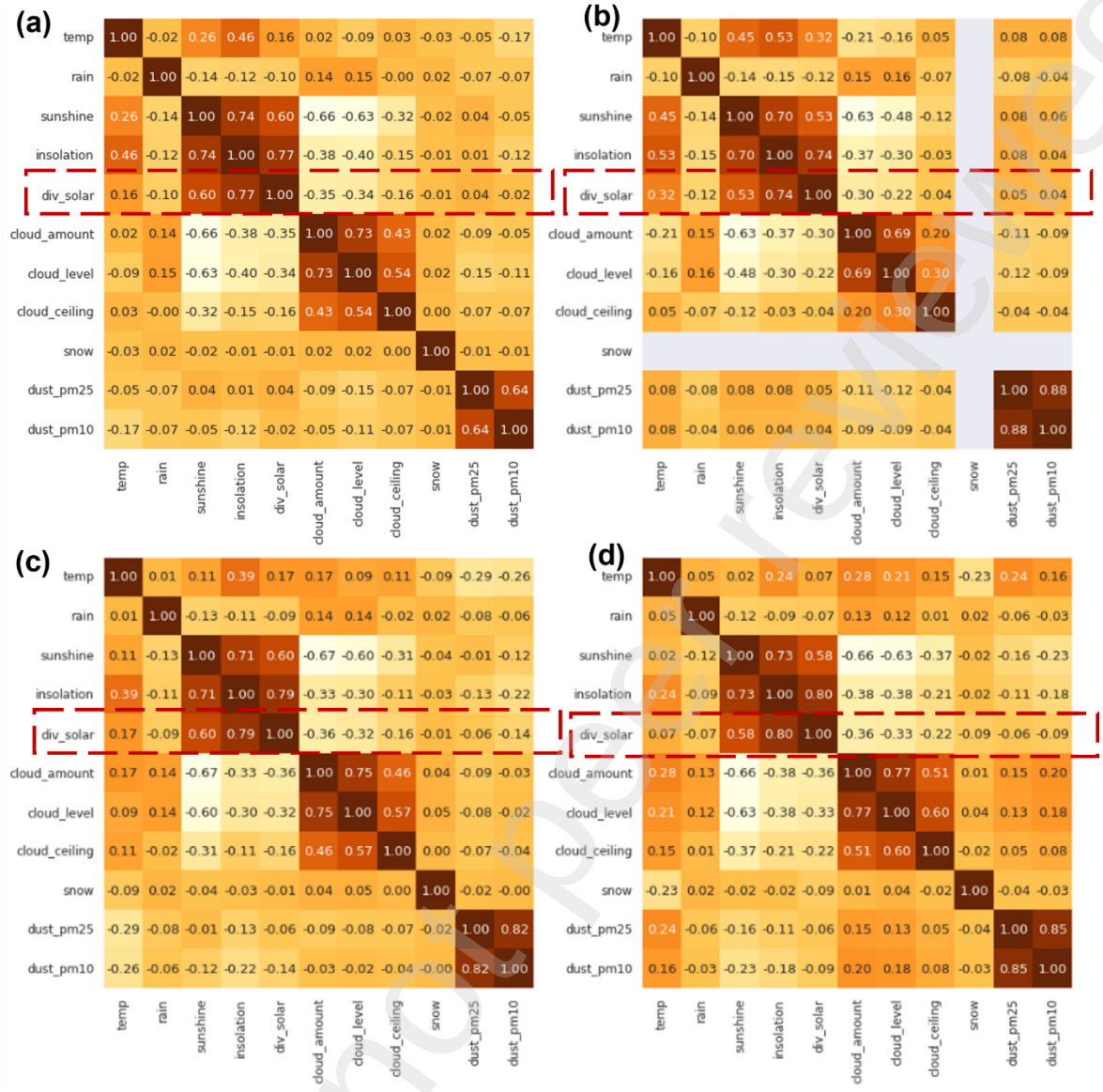


Fig. 4. Pearson correlation coefficients for Case 2 in different seasons as (a) spring, (b) summer, (c) fall and (d) winter.

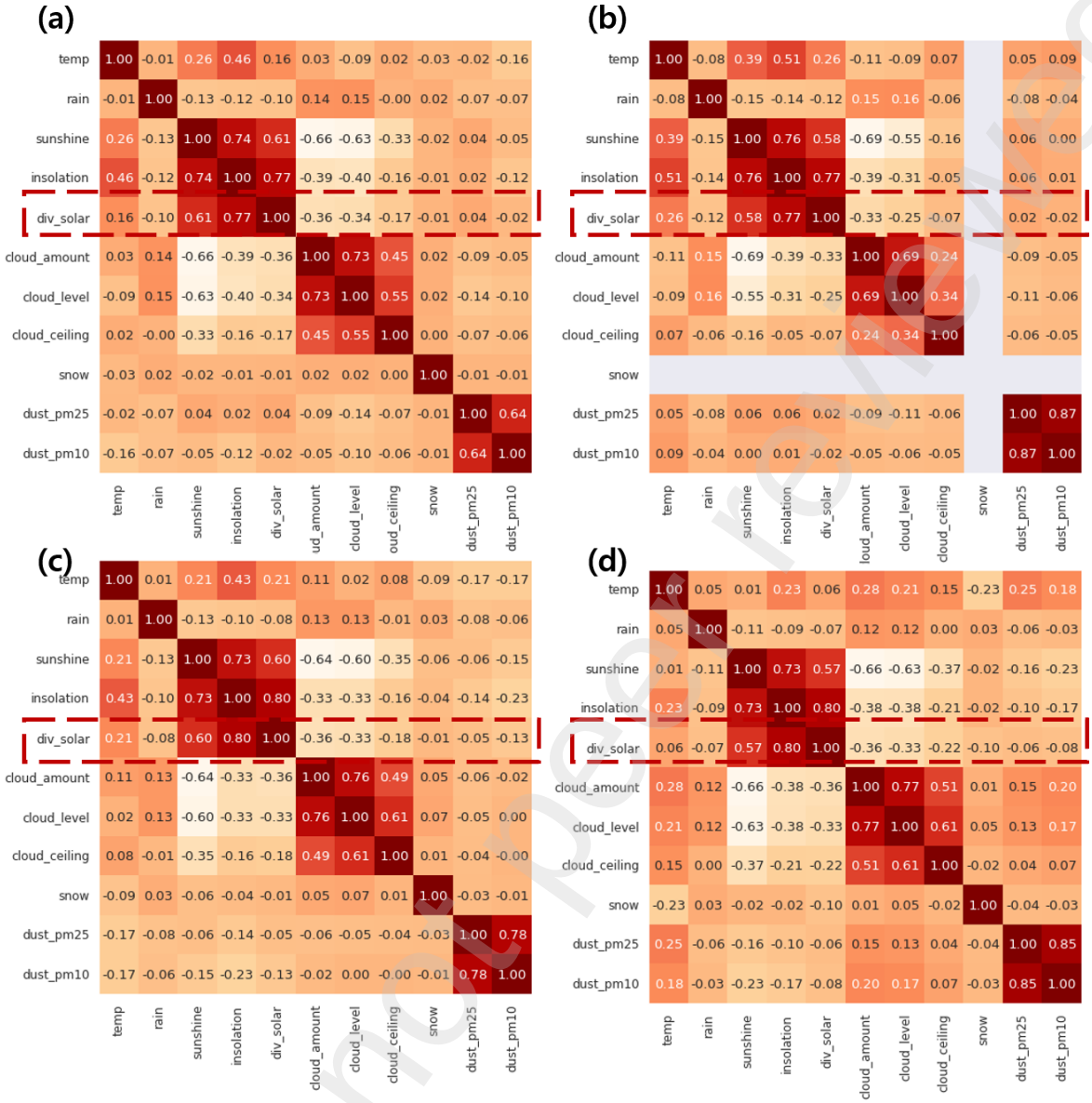


Fig. 5. Pearson correlation coefficients for Case 3 in different seasons as (a) spring, (b) summer, (c) fall and (d) winter.

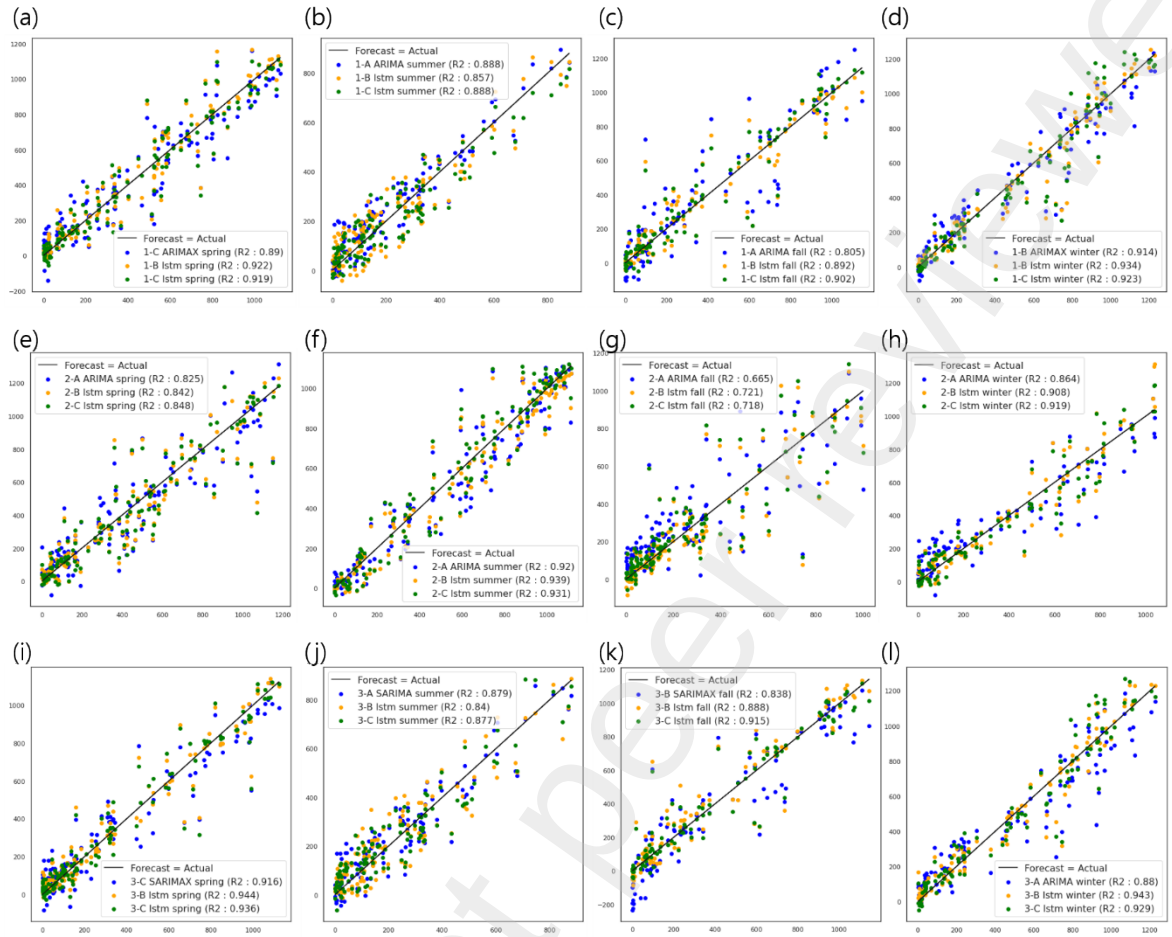


Fig. 6. R^2 graph between actual PV power output data and predict data. (a-d) are for Case 1, (e-h) are for Case 2 and (i-l) are for Case 3.

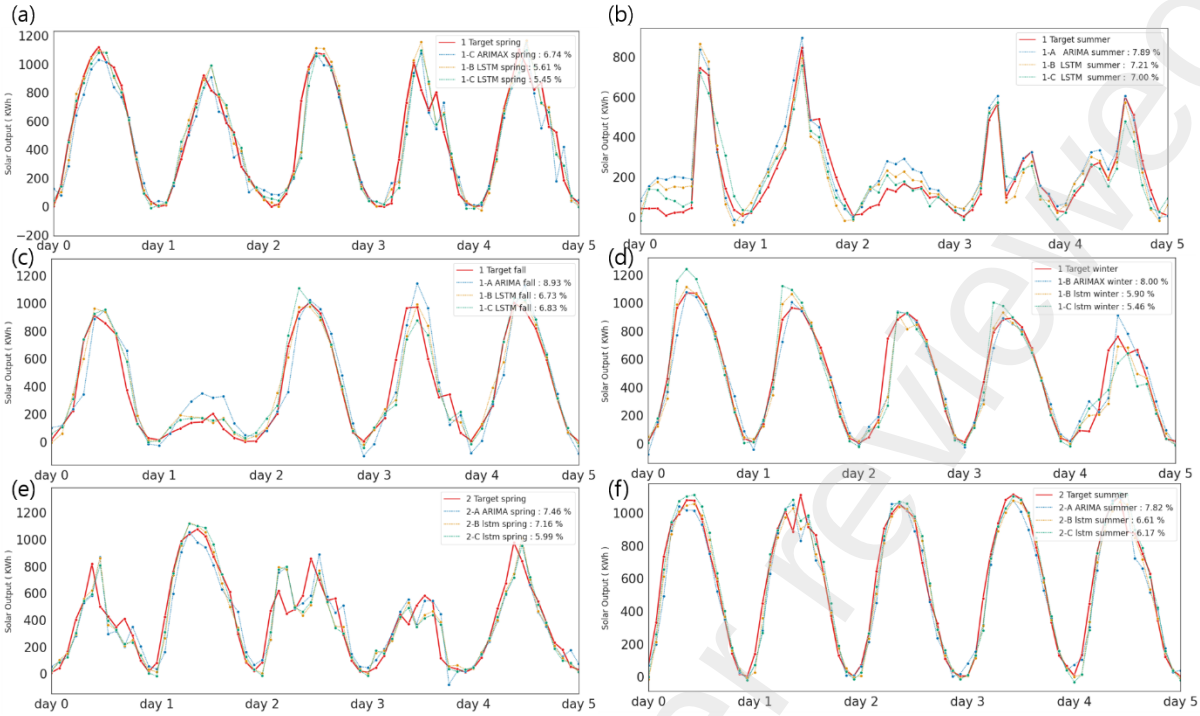


Fig 7. Regression plot extracted from results of top three models for PV power prediction, 5 days ahead for the validation. (a-d) for Case 1, and (e-f) for Case 2.

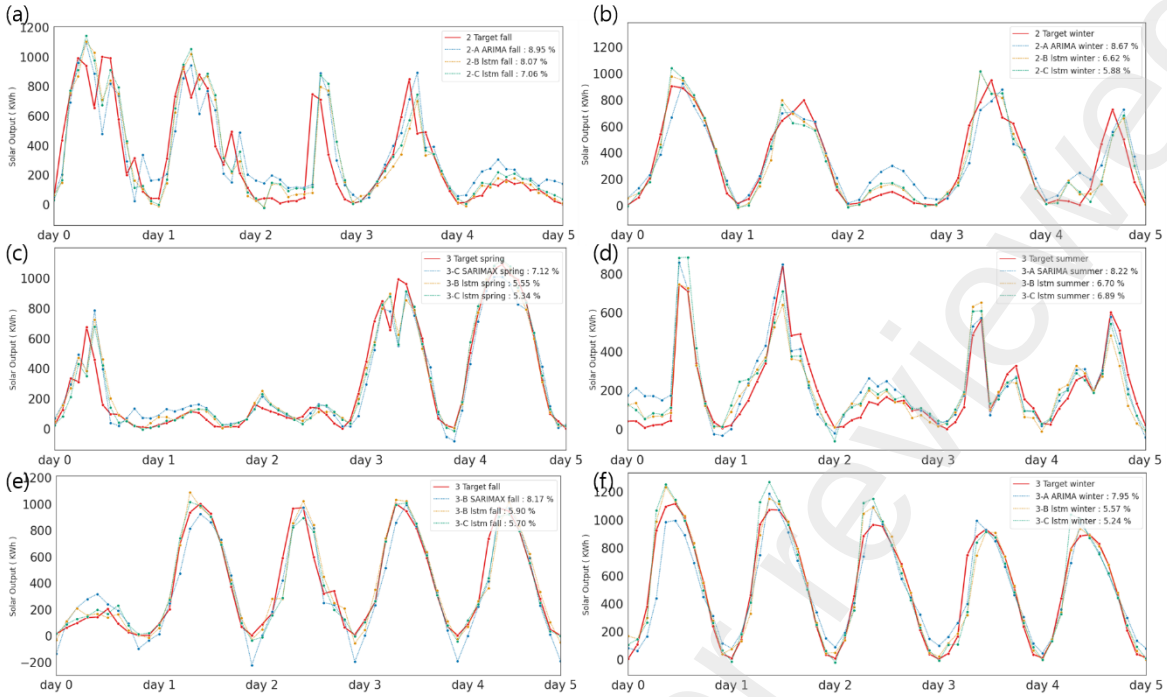


Fig 8. Regression plot extracted from results of top three models for PV power prediction, 5 days ahead for the validation. (a-b) for Case 2, and (c-f) for Case 3.