Short-Term Forecasting of Photovoltaic Power Integrating Multi-Temporal Meteorological Satellite Imagery in Deep Neural Network

Hunsoo Song
Dept. of Civil & Env. Eng.
Seoul National University
Seoul, South Korea
songhunsoo92@gmail.com

Gwangjoong Kim
Dept. of Civil & Env. Eng.
Seoul National University
Seoul, South Korea
todayfirst@snu.ac.kr

Minho Kim
Dept. of Civil & Env. Eng.
Seoul National University
Seoul, South Korea
mhk93@snu.ac.kr

Yongil Kim
Dept. of Civil & Env. Eng.
Seoul National University
Seoul, South Korea
yik@snu.ac.kr

***Abstract*—Remotely-sensed satellite imagery offers crucial information on the atmosphere and the local environment, providing a broader perspective for more accurate photovoltaic (PV) power prediction. This study proposes a Deep Neural Network (DNN) framework which integrates meteorological satellite images with historical PV power output data to conduct short-term PV power prediction (2-hour ahead). For this study, Communication, Ocean, and Meteorological Satellite (COMS) was used, and the proposed model was evaluated on test sites in Yeongam and Jindo, South Korea. The proposed DNN model was able to consider the variations of atmospheric condition and successfully learn the complex meteorological patterns by using multi-temporal COMS satellite images stacked with historical PV data. The experiment on historical PV power output, compiled over three years from 2015 to 2017, confirms that the integration of multi-temporal satellite images is more accurate than using single mono-temporal satellite image in short-term PV power prediction.**

***Index Terms*—Photovoltaic (PV) Power Prediction, Deep Regression Network, Meteorological Satellite Imagery, Deep Learning, Solar Energy**

I. INTRODUCTION

Solar energy is a promising renewable energy which has become a significant resource for sustainable development in the future [1], [2]. The problem, however, is that the output of solar photovoltaic (PV) power fluctuates strongly over time, making it difficult to supply energy stably on demand. The variability of PV output is mainly attributed to the changeable nature of meteorological phenomena. In other words, PV generation is influenced by numerous meteorological factors such as solar irradiance, cloud movement, cloud coverage, and other weather conditions which continuously differ over time [3]-[6]. Understanding the atmospheric conditions is thus essential to predict the PV output effectively.

To interpret the meteorological factors, ground-based weather measurements are commonly utilized. However, ground-based weather measurements alone cannot fully capture the unpredictable nature of meteorological conditions because the measurements only represent present time and local environment conditions. Alternatively, remote sensing methods can offer crucial information of the atmosphere in local and remote environments by utilizing a broader coverage for more accurate PV power prediction.

Previous studies have predicted PV power using ground-based sky imager and meteorological satellite imagery. Ground-based sky imager has been used to predict cloud motion efficiently, thereby generating solar irradiance and PV estimation [5]. However, these imagers are restricted by small coverage and only RGB channels (bands), which ultimately limit the imagers' ability to acquire essential atmospheric features and to conduct only intra-hour measurements. Conversely, meteorological satellites have exhibited the potential to capture information on the atmosphere and the local environment given their wider coverage of observation and greater number of spectral bands. Although meteorological satellites can compensate for the shortcomings of ground-based sky imagers, these satellites are limited by the coarse spatial resolution [1], rendering small-scale, detailed variations of atmospheric conditions difficult [7]-[11].

Moreover, the power plant capacity and PV cell characteristics can also impact PV generation [2]. The former factor refers to the output encompassing the conversion from solar energy harnessed at the PV cell to the electricity generated at the power plant. With regards to the latter factor, PV generation is influenced by the PV cell's total size, panel's material type, panel's standing angle with respect to the sun, and solar altitude. Due to the diversity of these factors, including meteorological conditions and PV generation capacity, reliable datasets accumulated over time are required, but previous studies struggled to process and integrate these extensive datasets for PV power prediction.

Recent studies have conducted PV power prediction by implementing a machine learning based methods. For instance, Alfadda (2017) attempted to predict short-term PV output using

[1]Meteorological satellite images usually have a spatial resolution of more than 1 km.

a support vector machine (SVM) based method [8]. Gandelli (2014) proposed an artificial neural network (ANN) based method while using weather temperature, wind speed, humidity, pressure, and cloud cover [15], while Massidda (2018) used gradient-boosted regression trees using numerical weather forecast data, astronomical data, and global horizontal irradiation [9]. However, in the majority of these studies, the used factors rely on measurements taken locally which may not be in the vicinity of the studied solar power plant location.

To address these aforementioned limitations, we propose a Deep Neural Network (DNN) framework which can successfully learn the features of atmospheric variations with stacked meteorological satellite images acquired over different time stamps and historical 3-year, PV power output data. Multiple meteorological satellite images were acquired over different time stamps (4-hour prior, 2-hour prior, and present time) in order to interpret cloud coverage and their presumed motion inherent in the images over time. In addition, previously measured PV output data were included to reinforce the model's prediction accuracy by considering the power plant capacity and PV cell's specifications. Finally, the hyper-parameters of the proposed model were fine-tuned to enhance the model's training and learning capabilities for optimal performance, and the model was evaluated with a three-year PV output data on test sites in Yeongam and Jindo, South Korea.

The present paper is structured as follows: Section II introduces the two study sites used to assess PV power prediction. Specifications of the meteorological satellite data and historical PV output data are also provided. Next, Section III describes the input datasets and outlines the methodology behind the DNN model used in our approach. Then, Section IV displays the results from PV estimates acquired from each test sites with relevant analysis. Lastly, the conclusion provides a summative review of our research.

II. Study Area and Datasets

A. Study Area

This study was conducted on two solar energy farms, both located in South Jeolla Province, South Korea. The first test site is situated in Yeongam-gun (34°50'14.9"N 126°40'49.5"E) with a capacity of 1,500 KW and the second test site is in Jindo-gun (34°32'33.5"N 126°20'14.9"E) with a capacity of 1,000 KW. The precise locations of the two test sites are labelled in the highlighted area shown in Fig. 1.

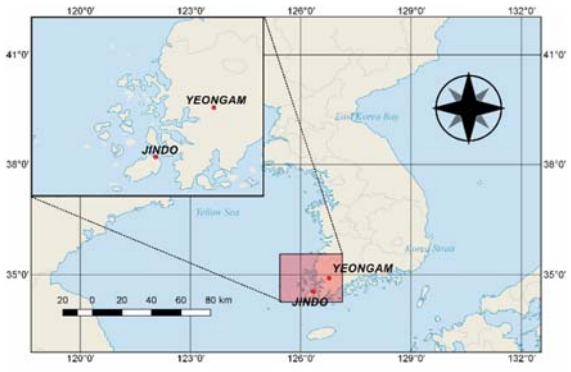

Figure 1. Location of study area showing Yeongam and Jindo test sites.

B. Meteorological Satellite Data

We used Communication, Ocean, and Meteorological Satellite (COMS) imagery as a meteorological satellite image source. COMS images were supplied by National Meteorological Satellite Center (NMSC). The satellite was launched as Korea's first geostationary, multi-purpose satellite in 2010 at 36,000 km above the Earth's equator. Images are acquired over the Korean peninsula at 15-minute intervals using the Meteorological Imager (MI) sensor module aboard COMS. The sensor's detailed specifications are organized in Table I.

TABLE I. Specifications of COMS-MI

Band	Wavelength (μm)	Observations and Applications
Visible[a]	0.55 - 0.80	Cloud coverage, atmospheric motion vector, fog, Asian dust
Shortwave Infrared	3.5 - 4.0	Land surface temperature, night fog, low-level clouds
Water Vapor	6.5 - 7.0	Mid-upper atmospheric humidity, upper atmospheric motion
Infrared-1	10.3 - 11.3	Cloud information, sea surface temperature, Asian dust
Infrared-2	11.5 - 12.5	Cloud information, sea surface temperature, Asian dust

a. The visible band has a spatial resolution of 1 km while the other bands are 4 km

C. Historical PV Output Data

Historical PV power output, complied over three years from 2015 to 2017, were used in hourly intervals. The PV power datasets of the two test sites were supplied by Open Data Portal operated by the Korea Minister of Government Administration and Home Affairs.

III. Methodology

A. Three Models: MULTI, MONO, and TWO-PV

To evaluate the efficiency of using multi-temporal satellite imagery, we prepared three DNN models (i.e., *MULTI*, *MONO*, and *TWO-PV*) which used three different types of input datasets. First, the *MULTI* model used 5-band COMS images acquired across three periods (4-hour prior, 2-hour prior, and present time) and two-hourly PV output data (1-hour prior and present time). Second, the *MONO* model used 5-band COMS images taken only in the present time and the two-hourly PV output data. Last, the *TWO-PV* model only used the two-hourly PV output data. The three models and their input datasets are shown in Table II. The time of data acquisition in hourly intervals with respect to the present time (0H) is henceforth denoted as "nH", where "-nH" refers to the time before present time and "H" symbolizes hours. We used the pixel of the corresponding COMS images at the point where the power plant is located. Therefore, the *MULTI* model has a total of 17 variables consisting of 15 variables from multi-temporal COMS images and 2 PV variables, the *MONO* model has a total of 7 variables consisting of 5 variables from mono-temporal satellite imagery and 2 PV variables, and the *TWO-PV* model has only 2 PV variables. Lastly, the PV data corresponding to each of the two solar farms were randomly divided into a 30% training set and a 70% testing set. Then, 30% of the training set was used to create a validation set of 9%.

TABLE II. THREE MODELS AND INPUT DATASETS

Model	PV Output		5-Band COMS Image(s)		
MULTI	0H	-1H	0H	-2H	-4H
MONO	0H	-1H	0H		
TWO-PV	0H	-1H			

B. Deep Regression Network

In our experiment, we used a DNN with two hidden layers, each consisting of 64 neurons. We adopted a relatively simple network, robust to avoid overfitting, but enough to learn the features given the small size of the input data. The network takes each of three types of inputs, then passes it through two hidden layers, which results in a single output estimating the 2-hour ahead PV value. We used a rectified linear unit as an activation function. The parameters in the network are updated by full-batch gradient descent, and we adopted the RMSProp [12] with a learning rate of 0.001. The mean squared error, the most commonly used loss function, was selected, and an early stopping strategy with patience of 15 was used to avoid overfitting. The three DNN models and their different combinations of input data are visualized in Fig. 2.

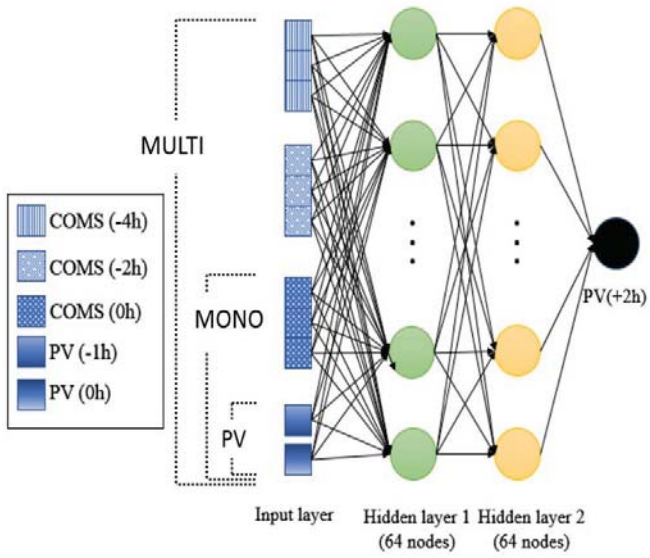

Figure 2. Architectures of three DNN models.

C. Model Evaluation

The following metrics provided in Table III were used to assess the performance of the proposed model. y_i and $\hat{y}_i$ are the observed and predicted values of PV output, which are rescaled to the range of [0,1] by min–max normalization, respectively. Also, μ_{y_i} and $\mu_{\hat{y}_i}$ represent the mean values of y_i and $\hat{y}_i$, respectively. These metrics were selected based on their suitability for analyzing PV estimates and their common usage in previous studies [13], [14].

IV. RESULTS AND DISCUSSION

TABLE III. PERFORMANCE EVALUATION METRICS

Metric	Formula
Correlation Coefficient (*r*)	$r = \frac{\Sigma_{i=1}^{N}(y_i - \mu_{y_i})(\hat{y}_i - \mu_{\hat{y}_i})}{\sqrt{\Sigma_{i=1}^{N}(y_i - \mu_{y_i})^2}\sqrt{\Sigma_{i=1}^{N}(\hat{y}_i - \mu_{\hat{y}_i})^2}}$
Normalized Root Mean Square Error (*NRMSE*)	$NRMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(\hat{y}_i - y_i)^2}$
Normalized Mean Absolute Error (*NMAE*)	$NMAE = \frac{1}{N}\Sigma_{i=1}^{N}\lvert\hat{y}_i - y_i\rvert$

A. 2-Hour PV Forecast

The results of the 2-hour ahead PV power predictions using the three models (i.e., *MULTI*, *MONO*, and *TWO-PV*) in the Yeongam and Jindo test sites are displayed in Tables IV and V, respectively. The metric values for each case were produced by averaging the results produced from five iterations. The most optimal value pertaining to each metric is highlighted in bold-face for convenience. Using the input data specifications outlined in the previous section, the proposed model was able to learn and train the three datasets successfully. The predicted PV outputs were compared with previously measured PV output data to evaluate the prediction accuracy and the performance of the proposed model.

TABLE IV. EVALUATION OF 2-HOUR PV FORECAST IN YEONGAM

Metric \ Model	*MULTI*	*MONO*	*TWO-PV*
r	**0.8780**	0.8600	0.8322
NRMSE	**0.1348**	0.1430	0.1515
NMAE	**0.1037**	0.1073	0.1183

TABLE V. EVALUATION OF 2-HOUR PV FORECAST IN JINDO

Metric \ Model	*MULTI*	*MONO*	*TWO-PV*
r	**0.8668**	0.8463	0.8372
NRMSE	**0.1513**	0.1591	0.1548
NMAE	**0.1176**	0.1230	0.1221

B. Analysis of 2-Hour PV Estimates

The results from both test sites demonstrated that the proposed DNN model can predict PV output to a high degree of accuracy. In more detail, the results revealed that *MULTI* model showed the best performance for both test sites, and in turn, demonstrated that the use of multi-temporal satellite imagery improves the PV power prediction performance. With respect to *MONO* and *TWO-PV*, *MONO* returned better values for most metrics. This observation suggests that the inclusion of meteorological satellite imagery, albeit as present time (0H) acquisitions, is nonetheless effective for enhancing PV power

prediction accuracy. Overall, these results indicate that the integration of meteorological satellite images can improve the performance of PV power prediction when using our proposed DNN model approach.

As previously mentioned, the unpredictable nature of meteorological phenomena, especially cloud motion, limits reliable PV power prediction. However, this study shows that using multi-temporal satellite images can help to increase the interpretability and content of the data to better understand these limitations. The use of a wider time horizon in the satellite images enables the proposed model to take into account the temporal variations in the atmosphere. The integration of multi-temporal satellite images is, therefore, more effective than using single mono-temporal satellite image to further augment the accuracy of PV power prediction.

C. Limitations of Input data Using MULTI and MONO

We conducted additional experiments on the Yeongam and Jindo test sites to explore the limitations of using meteorological satellite imagery as an input source and to validate the performance of models in longer time horizons. The present time and multiple forecast time horizons (2-hour, 3-hour, 4-hour ahead) were tested using the *MULTI* and *MONO* models, and the results are presented in Table VI. The *TWO*-PV model was excluded due to its poor performance. Assuming present time PV output is unknown, the present PV (0-hour) can be estimated by using meteorological satellite imagery of the present time rather than 2-hour prior imagery. The 0-hour predictions are considered as ideal estimations of the PV output since the input datasets were taken at the same 0-hour time stamp. As a result, PV estimates using 0-hour satellite imagery demonstrated higher accuracy than when using satellite imagery acquired at later times. The difference in accuracy indicates the limitation of deriving future atmospheric conditions from the 0-hour image through the proposed DNN model. To note however, error was still prevalent in the 0-hour prediction, in spite of using present time satellite image acquisitions. This result therefore suggests that the addition of other factors uninterpretable from using only previous PV outputs and satellite images can help to improve the accuracy of PV output estimation. However, these additional factors are outside the scope of this research, and were not analyzed in further detail.

The 3 and 4 hour-ahead PV forecast were conducted using the same training dataset used in the 2-hour ahead PV forecast. The prediction accuracy decreased according as the time horizon increased (from 2 to 4 hour-ahead). This result was consistent for both study areas. Although the metric values, *NMRSE* and *NMAE* can vary based on the geographical location and local PV capacity conditions of the test site, the *MULTI* model outperforms the *MONO* model in every case, confirming that the use of multi-temporal satellite images produces superior performance in PV prediction.

V. Conclusion

The present study confirmed that the integration of multi-temporal meteorological satellite imagery and historical PV data in a DNN framework can successfully predict short-term (2-hour ahead) PV output. To the best of author's knowledge, this study is the first to utilize multi-temporal meteorological satellite images and historical PV data for solar energy forecast in a DNN model framework. Results from the study showed that the combination of multi-temporal (4-hour prior, 2-hour prior, and present time) 5-band COMS images in conjunction with historical (1-hour prior and present time) PV output data outperformed the other two combinations of mono-temporal satellite imagery and only historical PV output data for all metric evaluations. This study showed that using 30% of the three-year dataset as a training set produced acceptable prediction results. Based on this study, this suggests that using a one-year dataset can be sufficient for effective PV forecasting. While meteorological phenomena can complicate the prediction of reliable PV output, this result demonstrated that using meteorological satellite images acquired using a wider time horizon can help the model to better understand and interpret these variations. Therefore, the composition of multi-temporal imagery used as input data for the model should be varied in accordance to the target time.

Despite these promising results, the study was not without limitations. The performance of the proposed model was expected to be largely dependent on the geographical location. Future works should therefore test the proposed model on different locations, especially for test sites where PV generation facilities are more widely dispersed to validate the versatility of the model to geographic variations. In addition, although the sizes of the solar farms in Yeongam and Jindo are within an area 2 km by 2 km, the spatial resolution of meteorological satellite imagery may not be able to fully resolve features for larger test sites. In this case, future research can utilize satellite images using patches (array of pixels) instead of analyzing per

TABLE VI. Comparison of *MULTI* and *MONO* in Diverse Time Horizons

	Forecast Time	Present Time (0-hour)		2-hour Ahead		3-hour Ahead		4-hour Ahead	
Test Site	Model / Metric	*MULTI*	*MONO*	*MULTI*	*MONO*	*MULTI*	*MONO*	*MULTI*	*MONO*
Yeongam	*R*	**0.9138**	0.9033	**0.8780**	0.8600	**0.8068**	0.7688	**0.7015**	0.6727
	NRMSE	**0.1150**	0.1183	**0.1348**	0.1430	**0.1646**	0.1806	**0.2027**	0.2048
	NMAE	**0.0876**	0.0895	**0.1037**	0.1073	**0.1302**	0.1417	**0.1620**	0.1644
Jindo	*R*	**0.8933**	0.8749	**0.8668**	0.8463	**0.7899**	0.7636	**0.6928**	0.6637
	NRMSE	**0.1301**	0.1368	**0.1513**	0.1591	**0.1760**	0.1845	**0.2054**	0.2119
	NMAE	**0.0971**	0.0999	**0.1176**	0.1230	**0.1377**	0.1440	**0.1635**	0.1683

pixel. In this case, since the number of inputs will increase, a more complex network may be required to fully facilitate the input data. To elaborate, in order to learn features from larger images, we plan on developing a convolutional neural network framework integrating a wide range of satellite imagery and various meteorological elements to enhance the performance of PV power prediction and the robustness of the model.

ACKNOWLEDGMENT

The authors gratefully acknowledge SK Telecom for sponsoring this research and the National Meteorological Satellite Center for supplying COMS images.

REFERENCES

S. Ghosh and S. Rahman, "Global deployment of solar photovoltaics: Its opportunities and challenges," In *Proc. 2016 IEEE PES Innovative Smart Grid Technologies Conf.,* pp. 1-6.

C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE J POWER ENERGY*, vol. 1(4), pp. 38-46, Dec. 2015.

C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *RENEW ENERG*, vol. 105, pp. 569-582, May. 2017.

A. Alfadda, S. Rahman, and M, Pipattanasomporn, "Solar irradiance forecast using aerosols measurements: A data driven approach," *SOL ENERGY*, vol. 170, pp. 924-939, Aug. 2018.

C. L. Fu, C. L and H. Y. Cheng, "Predicting solar irradiance with all-sky image features via regression," *SOL ENERGY*, vol. 97, pp. 537-550, Nov. 2013.

M. Hosenuzzaman, N. A. Rahim, J. Selvaraj, M. Hasanuzzaman, A. A. Malek, and A. Nahar, "Global prospects, progress, policies, and environmental impact of solar photovoltaic power generation," *RENEW SUST ENERG REV*, vol. 41, pp. 284-297, Jan. 2015.

B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, and D. Heinemann, "Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data," *SOL ENERGY*, vol. 135, pp. 197-208, Oct. 2016.

A. Alfadda, R. Adhikari, M. Kuzlu, and S. Rahman, "Hour-ahead solar PV power forecasting using SVR based approach*." IEEE Power & Energy Society Innovative Smart Grid Technologies Conf. (ISGT),* pp. 1-5, Apr. 2017.

L. Massidda, and M. Marrocu, "Quantile Regression Post-Processing of Weather Forecast for Short-Term Solar Power Probabilistic Forecasting." *ENERGIES*, vol.11(7), 1763. Jun. 2018.

M. Almeida, M. Muñoz, I. de la Parra, and O. Perpiñán, "Comparative study of PV power forecast using parametric and nonparametric PV models." *SOL ENERGY*, vol.155, pp. 854-866, Jul. 2017.

C. W. Chow, S. Belongie, and J. Kleissl, "Cloud motion and stability estimation for intra-hour solar forecasting," *SOL ENERGY*, vol. 115, pp. 645-655, May. 2016.

T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning," vol. 4(2), 2012, pp. 26-31.

J. Zeng, and W. Qiao, "Short-term solar power prediction using a support vector machine." *RENEW ENERG*, vol. 52, pp. 118-127, Apr. 2013.

J. Zhang, A. Florita, B.M. Hodge, S. Lu, H.F. Hamann, V. Banunarayanan, and A.M. Brockway, "A suite of metrics for assessing the performance of solar power forecasting." *SOL ENERGY*, vol. 111, pp. 157-175, Jan. 2015.

A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari. "Hybrid model analysis and validation for PV energy production forecasting." *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 1957-1962, 2014.