Journal of Cleaner Production 250 (2020) 119476

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

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

Yoonhwa Jung [a], Jaehoon Jung [a], Byungil Kim [b], SangUk Han [a, *]

[a] Department of Civil and Environmental Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, 04763, Republic of Korea
[b] Department of Civil Engineering, Andong National University, 1375 Gyeongdong-ro, Andong-si, Gyeongsangbuk-do, 36729, Republic of Korea

ARTICLE INFO

Article history:
Received 27 August 2019
Received in revised form
1 November 2019
Accepted 26 November 2019
Available online 28 November 2019

Handling Editor: Bin Chen

Keywords:
Solar energy
Photovoltaic power
Long-term forecasting
Long short-term memory
Recurrent neural network

ABSTRACT

The sites selected for solar PV facilities significantly affect the amount of electric power that can be generated over the long term. Therefore, predicting the power output of a specific PV plant is important when evaluating potential PV sites. However, whether prediction models built with data from existing PV plants can be applied to other plants for long-term power forecasting remains poorly understood. In this case, topographical and meteorological conditions, which differ among sites and change over time, make it challenging to accurately estimate the potential for energy generation at a new site. This study proposes a monthly PV power forecasting model to predict the amount of PV solar power that could be generated at a new site. The forecasting model is trained with time series datasets collected over 63 months from 164 PV sites with data such as the power plant capacity and electricity trading data, weather conditions, and estimated solar irradiation. Specifically, a recurrent neural network (RNN) model with long short-term memory was built to recognize the temporal patterns in the time series data and tested to evaluate the forecasting performance for PV facilities not used in the training process. The results show that the proposed model achieves the normalized root-mean-square-error of 7.416% and the mean absolute-percentage-error (MAPE) of 10.805% for the testing data (i.e., new plants). Furthermore, when the previous 10 months' data were used, the temporal patterns were well captured for forecasting, with a MAPE of 11.535%. Thus, the proposed RNN approach successfully captures the temporal patterns in monthly data and can estimate the potential for power generation at any new site for which weather information and terrain data are available. Consequently, this work will allow planning officials to search for and evaluate suitable locations for PV plants in a wide area.

1. Introduction

Among other renewable energy sources (e.g., wind, tides, geothermal heat), photovoltaic (PV) solar energy is one of the most promising renewable energies available all over the world (International Energy Agency, 2018). However, solar energy generation is affected by geographical location, and thus accurately predicting the potential PV power available at candidate sites is critical to the success of solar PV projects (International Finance Corporation (IFC), 2019). For example, estimated power generation commonly serves as a crucial input to assess the feasibility of PV projects and select a suitable installation location for PV panels (Liu et al., 2017). Long-term forecasting of PV power is also important in balancing electricity supply and demand, improving energy performance (Lin and Pai, 2016), and financial planning (International Finance Corporation, 2019).

However, estimating the potential for solar PV power generation is challenging because of topographical and meteorological conditions at the site, which differ from region to region and vary over time (Das et al., 2018). The estimated amount of solar energy depends on the aspect and slope of the specific location, which can be extracted from terrain datasets (Gastli and Charabi, 2010). The amount of power generated is also strongly affected by weather conditions because the amount of solar energy that reaches the earth varies by season and even within days. Hontoria et al. (2019) pointed out that the use of solar irradiation data at high temporal

* Corresponding author.
E-mail addresses: joonv2@hanyang.ac.kr (Y. Jung), jhoonj1216@hanyang.ac.kr (J. Jung), bkim@anu.ac.kr (B. Kim), sanguk@hanyang.ac.kr (S. Han).

https://doi.org/10.1016/j.jclepro.2019.119476
0959-6526/

Table 1
Summary of previous studies on PV power prediction.

Study	Forecasting approach	Forecasting model	Inputs for PV power prediction	Target plant	Spatial resolution (m)	Prediction horizon	Temporal resolution
Yeo and Yee (2014)	Solar radiation → PV power	- ANN to estimate solar radiation - Solar radiation → PV power (conversion model)	Estimated solar radiation/total potential installment area, fraction of the covered area, efficiency of system, total potential area, area covered by solar panels, and efficiency	108	30 × 30	Monthly	1 month
Mellit and Pavan (2010)	Solar irradiance → PV power	- ANN to estimate solar irradiance - Solar irradiance → PV power (conversion model)	Estimated solar irradiance/PV array area, PV module efficiency, and system balance	1	–	24 h	1 h
Yona et al. (2008)	Solar insolation → PV power	- ANN for solar insolation - Solar insolation → PV power (conversion model)	Estimated insolation/air temperature, conversion efficiency of cell, array area, temperature	1	–	24 h	1 h
Almonacid et al. (2014)	Solar irradiance → PV module power → PV power	- ANNs - PV module power→ PV power (conversion model)	Estimated global solar irradiance using clarity index/air and cell temperature, I–V curve output of PV module/number of modules, power output of a PV module, number of modules, and system losses	1	–	1 day	1 h
Chen et al. (2011)	PV power	- ANNs for weather types: sunny, cloudy, rainy	Solar irradiation, wind speed, temperature, and relative humidity	1	–	24 h	1 h
Izgi et al. (2012)	PV power	- ANN	Past values of output power from a 750 W power capacity solar PV panel	1	–	0–60 min	1 min
Pedro and Coimbra (2012)	PV power	- ANNs	Hourly average of power output	1	–	1–2 h	1 h
Shi et al. (2012)	PV power	- SVMs for weather types: clear, cloudy, foggy, rainy	Historical PV power output	1	–	1 day	15 min
Wolff et al. (2016)	PV power	- SVR	PV power measurement scaled to installed capacity, solar irradiance and temperature from NWP (numerical weather prediction), cloud motion vector–based irradiance	921	12500 × 12500 and 7000 × 7000	5 h	15 min
Lin and Pai (2016)	PV power	- Least-squares SVR	Monthly solar power output	16	–	Monthly	1 month
Hossain et al. (2017)	PV power	- SVR - ANN - Extreme learning machine	Average solar irradiance, ambient temperature, module temperature, wind speed	3	–	1 day & 1 h	1 day & 1 h
Gensler et al. (2016)	PV power	- ANNs - LSTM - Auto-LSTM	Temperature from NWP, clear-sky filter, direct and diffuse solar radiation from NWP	21	–	1 day	3 h
Abdel-Nasser and Mahmoud (2017)	PV power	- LSTM	Past hourly PV power	2	–	1 h	1 h
Han et al. (2019)	PV power	- LSTM	Short-wave radiation, humidity, surface pressure, and temperature from historical data, PV power output	2	–	Monthly	15 min
Gao et al. (2019)	PV power	- LSTM	Solar irradiation, air temperature, relative humidity, and wind speed from NWP	1	–	1 h	1 h

resolutions (e.g., less than 1 h) is important when designing PV systems for smart grids. The combined uncertainties of topographical and meteorological conditions lead to difficulties in estimating the potential of power generation at new candidate sites. In addition, it was pointed out that the potential energy generation was not often verified with the actual PV power data in the previous studies (Liu et al., 2017; Al-Soud and Hrayshat, 2009).

To address those issues, machine learning–based approaches have been adopted (Table 1) to model the complex patterns in massive datasets (e.g., weather conditions) and predict energy potential through computational experimentation (Kalogirou, 2001). Table 1 summarizes the approaches presented in prior studies, which can be classified as direct and indirect forecasting models (Das et al., 2018). The direct methods (Chen et al., 2011; Izgi et al., 2012; Pedro and Coimbra, 2012; Shi et al., 2012; Wolff et al., 2016; Gensler et al., 2016; Abdel-Nasser and Mahmoud, 2017; Han et al., 2019; Gao et al., 2019) estimate the amount of PV power generation by directly learning the complex relationships among the variables (e.g., solar irradiation, temperature) and the power output (Antonanzas et al., 2016). Previously reported direct models mostly focus on operation planning for a specific PV system by forecasting the potential electric power in the short term. In contrast, the indirect methods (Yeo and Yee, 2014; Mellit and Pavan, 2010; Yona et al., 2008; Almonacid et al., 2014) begin by predicting solar irradiation and use a numerical or analytical conversion model with those predictions to estimate PV power generation. The indirect models can be site-specific (e.g., one target plant in Table 1) if the technical details of the PV equipment are not fully known (Antonanzas et al., 2016), and the effect of weather conditions (e.g., temperature, humidity) that directly affect the efficiency of PV cells (Mekhilef et al., 2012) can be considered differently depending on the selected conversion model.

The use of machine learning techniques has significantly improved the forecasting accuracy of solar energy models that use historical data. Because a machine learning model is generally data-dependent, various types (e.g., input variables) and resolutions (e.g., time horizons) of data have been tested to assess their performance in predicting the output of interest. Nevertheless, the effects of topographical variations at PV plant sites and temporal variations in meteorological conditions on PV power generation remain poorly understood, particularly when investigating wide areas to find potential sites for solar PV systems.

Many of the previous studies listed in Table 1 focused on generating power predictions for a single PV plant to analyze energy supply and demand. When data-driven approaches are applied for new plants, however, the data from one site might not adequately capture the circumstances at other sites with different geographical and topographical features. For instance, the topographical conditions, which can be extracted from terrain datasets, determine the spatial variations of solar irradiation (Šúri et al., 2005). Particularly in mountainous regions with complex terrains, this issue can be significant when estimating the amount of solar radiation available because of shading effects (Jung et al., 2019). On the other hand, the direct methods presented in Table 1 directly predict PV power at multiple sites distributed across a large area, but the purpose of those studies was mainly to model the temporal relationship between available solar energy (e.g., solar irradiation) and power output, rather than comparing spatial features among PV facilities by relying on available solar irradiation datasets. For example, Hossain et al. (2017) presented forecasting models for three PV plants that were built and individually tested with weather datasets. Datasets with spatial resolutions of 12.5×12.5 km^2 and 7×7 km^2 were used in Wolff et al. (2016) to extract precise topographical features; spatial resolution is important because the necessary features cannot be extracted if the PV system is smaller than a single raster cell on the map.

Although short-term forecasting (e.g., 1 h or one day ahead) is suitable for balancing variations between power supply and demand, long-term forecasting is needed to assess the economic feasibility (e.g., annual revenues) of new PV plants when investigating potential PV sites (Antonanzas et al., 2016; Craig et al., 2002). For example, annual data are often adopted to evaluate suitable PV sites because of the long payback time and maintenance period of a PV facility (Dalton et al., 2008; Jain et al., 2011). However, the performances of the machine learning models in most of the studies in Table 1 were evaluated with short time horizons. Thus, it is questionable whether data with short temporal resolutions can capture monthly or seasonal temporal patterns that can vary nonlinearly by month or season. For instance, Gao et al. (2019) built four individual models for the four seasons, reporting root mean square errors (RMSEs) in predicting PV power 1 h ahead of 5.34%, 9.57%, 13.86%, and 9.26% for spring, summer, fall, and winter, respectively. Abdel-Nasser and Mahmoud (2017) showed that temporal patterns of PV power were more apparent in a monthly resolution than in hourly, daily, or weekly resolutions. Whether temporal patterns of meteorological conditions can be preserved in monthly data and successfully modeled for long-term power prediction remains undiscovered, and it's an important question when using a single forecasting model with data from multiple PV sites to make predictions about other sites.

Thererfore, the purpose of this study is to propose and evaluate a machine learning–based forecasting model trained with data from multiple PV plants to predict the monthly PV power generation at new PV sites. Specifically, the network model is built with time series data collected from 164 spatially distributed actual PV plants on a monthly temporal resolution. To acquire the temporal patterns observed in the data from multiple sites, a recurrent neural network (RNN) with long short-term memory (LSTM) layers is adopted to process sequences of monthly data (e.g., monthly solar irradiation) and learn the temporal and topographical variations in solar irradiation and weather conditions. The LSTM unit in the forecasting model is expected to recognize and remember seasonal patterns hidden in the monthly weather and PV capacity datasets. Unlike in previous studies—in which multiple LSTM-based models were individually built for each facility (e.g., 21 models in Han et al., 2019) or season (e.g., 4 seasonal models in Gao et al., 2019)—this study uses a single integrated forecasting model that can be applied to PV sites whose data were not used during training. Therefore, the proposed method is intended to improve prediction performance by incorporating seasonal temporal patterns into the modeling to accurately estimate the potential for solar PV power generation at any candidate site.

2. Method

This study proposes a LSTM-based forecasting model for predicting the potential for solar PV power generation by learning the temporal patterns of solar irradiation and weather conditions from monthly datasets collected at multiple PV facilities. Fig. 1 illustrates the overall procedures. The monthly temporal resolution was selected to consider seasonal effects on PV power systems because seasonal weather forecasting has often been based on the monthly time horizon (e.g., Alonzo et al., 2017; De Felice et al., 2015). For the topographical conditions at specific PV plants, the computerized estimation method using digital elevation models (DEMs) from Jung et al. (2019) was applied to calculate the hourly potential of solar irradiation under clear-sky conditions. The obtained potential solar irradiation, summed to a monthly basis, is used as model input along with the monthly weather data (mean monthly temperature, relative humidity, wind speed, cloud amount,

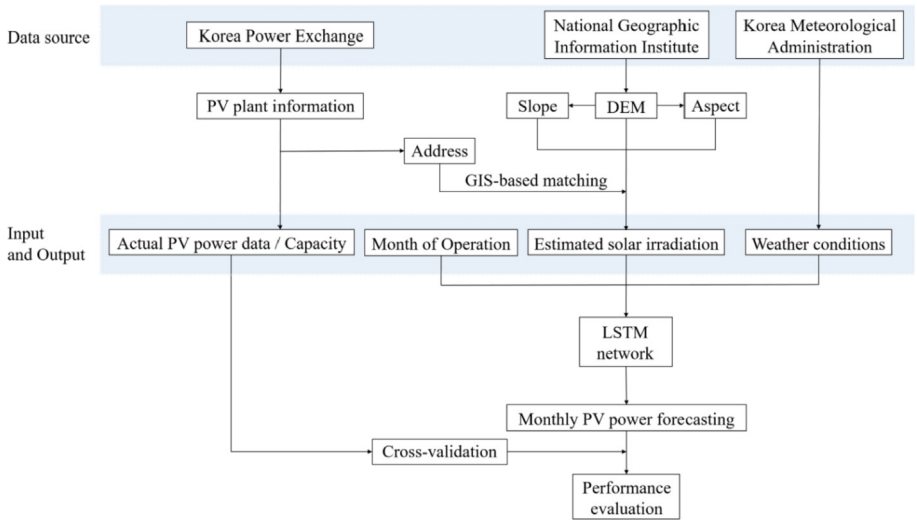

Fig. 1. Research overview: data flow and computational procedure.

precipitation, and duration of sunshine) to predict PV power output. Thus, the comprehensive effects of weather conditions and solar irradiation on the efficiency of PV cells can be incorporated into a training process that iteratively updates the model parameters of a neural network. All the time series datasets were randomly divided by PV plants into training and testing datasets: 80% of the PV plants (134 plants) were used to train the model, and the other 20% (30 plants) were used to verify that the model could predict PV power at a completely new plant. In the forecasting model, which consists of two stacked LSTM layers in an RNN, previous time series information is computationally connected to new input using memory cells. This approach can recognize and model temporal dynamic behavior in sequential data with seasonal trends because it allows the RNN to learn the patterns within the sequence itself and the temporal causality over a long-term horizon. To validate the model, the predicted amounts of power generation were compared with the actual PV power outputs of the test plants through cross-validation.

2.1. Description of the LSTM-RNN model

The RNN approach is adopted to use the temporal relationships among the inputs in the learning process, with inherent dynamic memory provided by the units in the recurrent connections (Coulibaly et al., 2011). The traditional approaches used in Table 1, such as artificial neural networks (ANNs), might not properly recognize temporal patterns in time-series data (Coulibaly et al., 2011; Giles et al., 1997), although seasonal trends in PV power generation are observed in practice. Instead, ANN models mainly learn complex relations among variables. The RNN model used in this study includes LSTM layers (Fig. 2) that compose memory blocks. In a memory block, each memory cell contains a self-connected linear unit that enforces a constant error flow by tracking errors as flowing back in time, which allows it to link huge time lags between events (Hochreiter and Schmidhuber, 1997). In addition to the memory cells, a memory block implements parameter updating in the form of a gradient descent to adjust the weights through the input gate i_t, forget gate f_t, and output gate o_t (Gers et al., 2000). These three gates store and reset the information in each memory cell to control the information flow during each activation function of the neural network layers. In turn, the output of a memory block is recurrently connected back to its input. Because the memory blocks store information for long periods by recirculating activation, a LSTM can recall information longer than conventional RNNs. An RNN model with two stacked LSTM layers was tested in this study to obtain a more complex hierarchical representation of the monthly input datasets for multiple variables (e.g., solar irradiation, temperature).

The stacked RNN model with LSTM units presented in this study was built to predict the amount of PV power generated per unit capacity at a monthly time resolution by handling temporal correlations in time series data and the nonlinear relationships among the eight input variables (i.e., weather conditions and solar irradiation, see Section 3) and the power output. In this process, various time steps (1–12 months) were also tested to assess how the temporal patterns in previous months' datasets affected the forecasting. Eventually, the network model trained with the entire training dataset was applied to the testing dataset to evaluate its

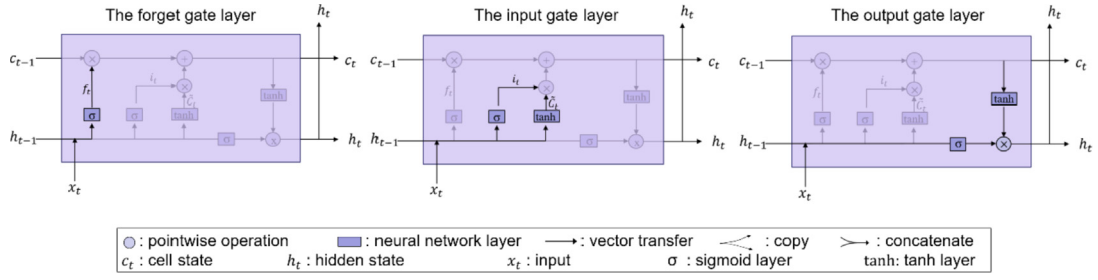

Fig. 2. A basic LSTM-RNN architecture.

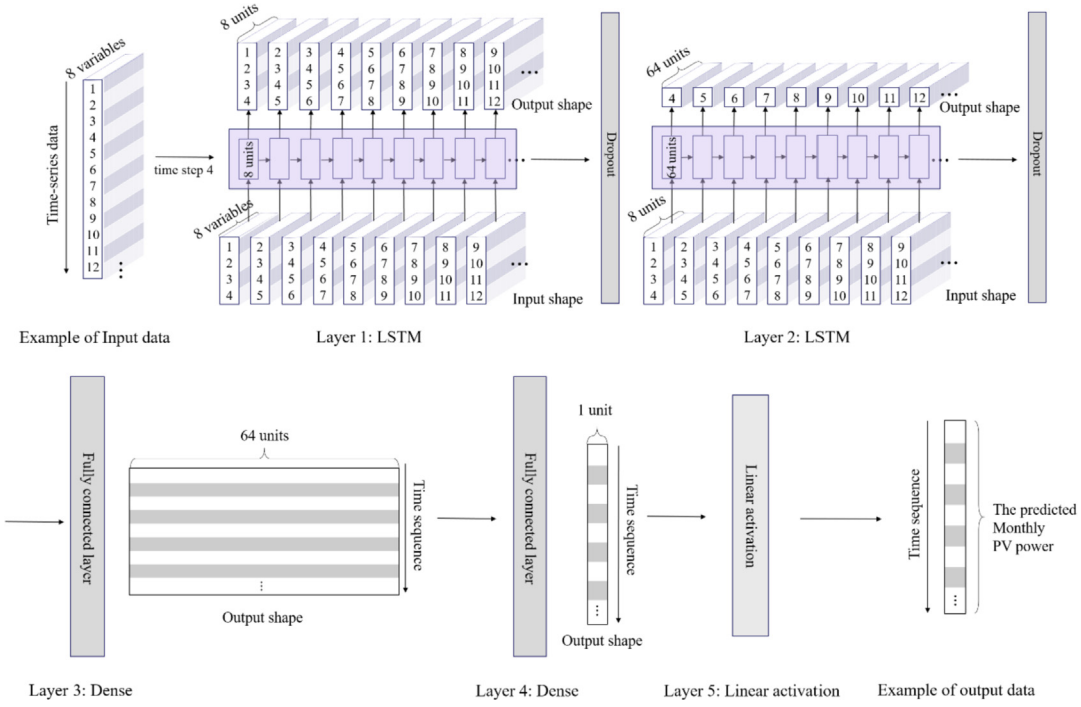

Fig. 3. An example of the proposed model process with a time step of 4.

forecasting accuracy for new PV facilities in a wide geographic region, its variations of errors for each month, and its performance with PV systems with different capacities.

The proposed network model is composed of five layers: two LSTM layers with dropouts, two dense layers, and a linear layer (Fig. 3). For the LSTM layers, the input data are reshaped into a 3D array of samples, time steps, and features. In this case, the samples depict the number of rows (i.e., data points) in the input dataset, the time step depicts the amount of time (e.g., how many months each data point includes) is being input into the model, and the feature is the number of variables in each sample in the input dataset. This process thus represents how many months of information are used to consider the temporal patterns in the time-series data. For instance, by setting the time step as four (Fig. 3), the weight of the first LSTM layer is recurrently updated with 4 months of time series data through the LSTM memory blocks that determine how long the layer remembers that information. Specifically, the first LSTM layer includes 8 units (conceptually, 8 nodes in a conventional neural network), as experimentally determined. In this study, eight input variables are used as features, and time steps between one to twelve were tested to determine the best one for predicting power output. Subsequently, the second LSTM layer controls and adjusts the output shape to compute the information (i.e., signals) of the current month (i.e., at time t). Thus, this 3D array of input data is directly input into the first LSTM layer, which passes it through the second LSTM layer to get the information for the target month, as seen in the upper row of Fig. 3. The dense layers, i.e., fully connected layers, are then used to output the prediction while adjusting the dimensions of the data. For example, the first dense layer (Layer 3) receives inputs (i.e., signals) from 64 LSTM units and outputs 64 dense units (i.e., 64 nodes) connecting the outputs from those 64 dense units to one dense unit (i.e., one node) in Layer 4, which converts the signal dimension into a one-dimensional prediction of PV power. The following linear layer is added to deal with the sign (e.g., positive and negative) of the output values from the dense layer.

Layer 1. The first LSTM layer is set to include eight hidden units to input eight initial variables (i.e., the month of operation, estimated solar irradiation, mean monthly temperature, relative humidity, wind speed, precipitation, cloud amount, and duration of sunshine). The memory cells and gates receive a 3D array of the input data, and the long-term memory is stored in vector form in a memory cell (c_t). As activation functions, the hyperbolic tangent in the gates can reset or access the memory in the cell, and a LSTM unit can control the information at the cell state for the next time step. Eventually, the weights and bias vectors are learned through back-propagation in the training process. In addition, the dimension of the input data for this layer is maintained as input, so that the 3D array data can be entered into the second LSTM layer. In this way, the LSTM layers can be stacked and connected. At the end of this layer, a dropout operator is added as a non-recurrent connection to prevent the overfitting that can result from using wide-range data (Zaremba et al., 2014). The dropout rate is set to be 0.2 in this experiment, as a range from 0.2 to 0.5 is commonly used (Srivastava et al., 2014).

Layer 2. The second layer contains 64 LSTM units, called hidden states, to further deal with the complexity and nonlinearity of the input data. In this layer, the output dimension is reduced to a 2D array to predict the following layers. A dropout operator is also added at the end of this layer, and the dropout rate is set to be 0.2 as in Layer 1.

Layers 3 and 4. In these dense layers, each dense unit is fully connected to every unit in the previous layer. Specifically, in Layer 3, 64 dense units are set to connect with the previous 64 LSTM units and control the magnitude of the signals, producing output signals in a range of $[-1, 1]$. Herein, the hyperbolic tangent function is used as an activation function for the output values. In Layer 4, one dense unit with a rectified linear unit is used as an activation function that returns the original input for positive values and zero for negative values.

The output of this dense layer can be calculated using Eq. (1).

$$y_{output} = \text{activation function}\left((Weight \cdot x_{input}) + bias\right) \quad (1)$$

Layer 5. In the output layer, a linear activation function is configured to output a positive value for the predicted PV power for a regression, such as outputting a single numerical value (i.e., the amount of PV power generation).

During training, the weight and bias parameters in the stacked LSTM-RNN model are learned and updated to minimize the prediction errors through back-propagation. To minimize the difference between the actual and predicted values, the mean-sqaured-error (MSE) is selected as a loss function for the regression model. In addition, to reduce the effects of the different scales in the input datasets on the parameter updating, all datasets are normalized between 0 and 1 using min-max normalization, as shown in Eq. (2).

$$\acute{y_i} = \frac{y_i - y_{min}}{y_{max} - y_{min}} \quad (2)$$

where y_i is the original data value, $\acute{y_i}$ is the normalized variable, y_{min} is the minimum value in $\{y_i\}$, and y_{max} is the maximum value in $\{y_i\}$.

2.2. Performance evaluation

The evaluation metrics used to examine the performance of the proposed model are the MSE, RMSE, mean absolute percent error (MAPE), and correlation coefficient of determination (R^2). In particular, the RMSE indicates the extent of concentration of the predicted values around the actual values, and the MAPE represents the relative error. R^2 represents how well the dependent variables describe the variation in an independent variable in a regression model. The formulas are given as follows:

$$MSE = \frac{1}{m}\sum_{t=1}^{m}(P_t - \widehat{P_t})^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(P_t - \widehat{P_t})^2} \quad (4)$$

$$nRMSE\ (\%) = \frac{RMSE}{P_{max} - P_{min}} * 100 \quad (5)$$

$$MAPE\ (\%) = \frac{1}{m}\sum_{t=1}^{m}\frac{P_t - \widehat{P_t}}{P_t} * 100 \quad (6)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{m}(P_t - \widehat{P_t})^2}{\sum_{t=1}^{m}(P_t - \overline{P})^2} \quad (7)$$

where P_t is the monthly measured power output, $\widehat{P_t}$ is the monthly predicted power, m is the size of the evaluation dataset, and $\overline{P} = \frac{1}{m}\sum_{t=1}^{m} P_t$. These error measures can be used to assess the forecasting model by penalizing large errors and quantifying the correlation between predicted and actual values. In addition, the 5-fold cross-validation method was used to test the ability of the forecasting model. For 5-fold cross-validation, the dataset is randomly divided into 5 subsets, and 4 subsets are used for training, with the remaining subset used for testing. This process is repeated iteratively 5 times, fairly using every subset once as the testing dataset to minimize overfitting and bias associated with the training data selection. A total of five trials with different combinations are implemented, and the average error among the five trials is calculated to evaluate the model performance. Furthermore, the errors for each month were analyzed and compared to assess the seasonal influence on forecasting accuracy, and the performance was compared between PV plants with low and high capacities to assess whether the data distribution was biased toward low capacity plants.

3. Implementation: data collection and model setting

The proposed LSTM-RNN model was implemented and tested with actual datasets. The details of data collection and variable selection, as well as the results of model setting with the collected data, are described in this section.

3.1. Data collection and processing

The datasets collected to build the forecasting model contain three types of data: (1) power generation information from existing PV facilities, (2) available solar irradiation estimated for each facility, and (3) weather conditions in the region of each facility. An overview of data organization with samples is provided in Fig. 4. The country is mostly mountainous, so publicly available solar energy maps (e.g., at spatial resolution of 1×1 km^2) were not adopted because the resolution was inadequate. Instead, a DEM with 30×30 m^2 resolution was used to estimate the potential solar irradiation at the PV facilities.

The PV plant datasets contain information about 164 solar PV facilities collected from the Korea Power Exchange. The datasets include the PV capacity and monthly power trade from January 2013 to March 2018 for each facility, as well as its specific location. The power capacity of a PV plant is a determinant factor affecting the electricity generated. To reflect variations in the operational efficiency of different plants, the power generation per unit capacity per hour, obtained by dividing the monthly power output (MWh) by the capacity (MW), was used as the output data of the forecasting model. As shown in Table 2 and Fig. 5, the model outputs range widely from 14.995 h to 203.809 h, exhibiting a normal distribution with a mean of ~107.94 h. The first month is excluded from the dataset in the modeling because the date of initial operation is unknown, leading to uncertainties in the operational efficiency in the first month. In addition, a DEM of the specific location of each PV facility was used to estimate the solar irradiation at its location. The address information, initially in the form of a street address, was manually converted into latitude and longitude coordinates through Google Earth.

Because the available solar energy is the primary resource for PV power, the amount of solar irradiation is used as input to predict the PV power output, as done in previous studies (Table 1). The potential solar irradiation at existing PV facilities at a specific time is estimated using the DEM-based solar energy model presented in a prior work (Jung et al., 2019). The solar energy estimation includes an assessment of the diffuse, ground-reflected, and direct nominal radiation under clear-sky conditions that results from calculating the adjusted solar constant with respect to the Earth–sun distance at a particular time of year and the use of solar radiation models (Gueymard and Thevenard, 2009) for computing each component (e.g., ground reflection, sky diffusion, beam radiation). To adjust the irradiation amount for shading caused by adjacent topographies, the elevation, slope, and aspect are extracted from the DEM. The use of the topographic datasets enables the calculation of shading effects that can occur where topographic obstacles exist between the sun and the surface (e.g., in a mountainous region). The data processing for solar irradiation estimation was conducted on DEMs with a grid cell size of 30×30 m^2, collected from the National

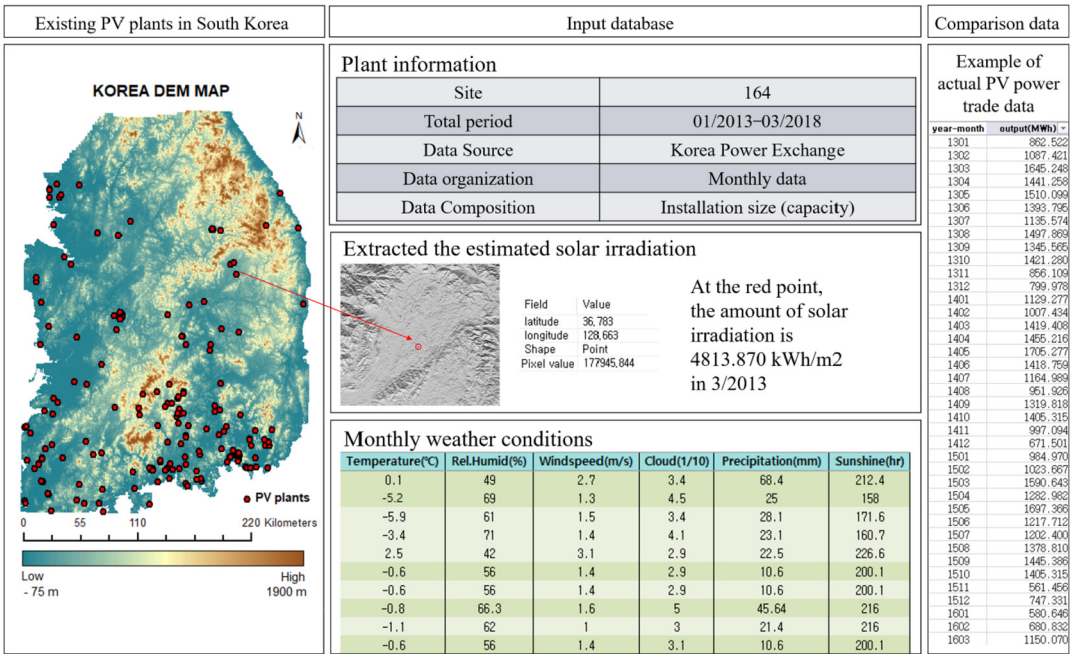

Fig. 4. Overview of the organization of the sample data.

Table 2
Descriptive statistics of data collection on solar PV plants.

Number of PV plants	Period	PV power generation per unit capacity (h)				
		Mean	Median	Max	Min	Standard deviation
164	January 2013 to March 2018	107.935	108.490	203.809	14.995	27.083

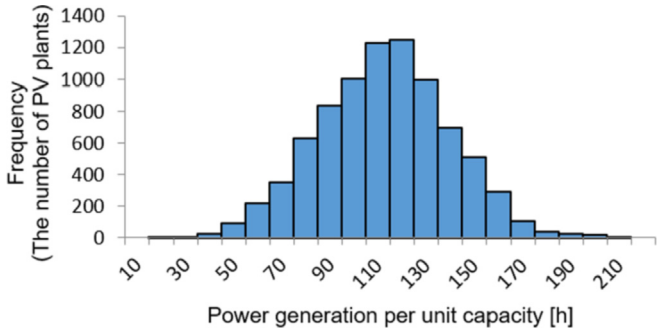

Fig. 5. Histogram of the forecasting model output.

Geographic Information Institute and performed at an hourly temporal resolution (Fig. 6). Eventually, the amount of solar irradiation estimated on an hourly basis is summed to obtain the monthly solar irradiation map, whose time horizon corresponds with that of the PV output data. In addition, for the solar irradiation amounts at existing PV facilities, the latitude and longitude coordinates of each PV site in the WGS84 reference system (Fig. 7a) were transformed into UTM coordinates (Zone 52) to match the facility site with a corresponding point within the solar map (Fig. 7b).

The monthly weather data comprise the mean monthly temperature, relative humidity, wind speed, cloud amount, precipitation, and duration of sunshine, which are all factors that can affect PV power generation. The weather conditions at each PV facility are estimated based on the observation station nearest to the site using datasets collected by the Korea Meteorological Administration. The selection of these weather-related variables was based on previous studies, as follows: temperature affects the performance of the PV structure by affecting the voltage, which is determined by the speed of electrons travelling through an electrical circuit (Fesharaki et al., 2011). Humidity decreases the reception of solar radiation because the sunlight can be refracted, reflected, or diffracted through water droplets in the air (Mekhilef et al., 2012). Chen et al. (2011) also reported that relative humidity negatively affects the amount of PV power. Wind speed influences the PV power generator by cooling the PV cell, which affects the electricity efficiency in the PV system (Chen et al., 2011). In addition to the temperature, a PV generator is also affected by precipitation (Izgi et al., 2012; Long et al., 2014). PV panels can be damaged and decompose upon extended exposure to rain, and thus the efficiency of PV cells can decrease (Hailegnaw et al., 2015). Meanwhile, snow can cover a PV module, causing energy production to be overestimated (Antonanzas et al., 2016). Furthermore, the available solar energy generally depends on the coverage of clouds and fog and the duration of sunshine (Sobri et al., 2018; Shivashankar et al., 2016).

In summary, eight variables—the month of operation, estimated solar irradiation, mean monthly temperature, relative humidity, wind speed, precipitation, cloud amount, and duration of sunshine—were used as inputs to predict the output, which is the amount of PV power generation per unit capacity, as summarized in Table 3. All the inputs and outputs were used together to train the network, which learned the relationships and patterns among them. Once the network was fully trained, only the inputs were used to predict the PV power output.

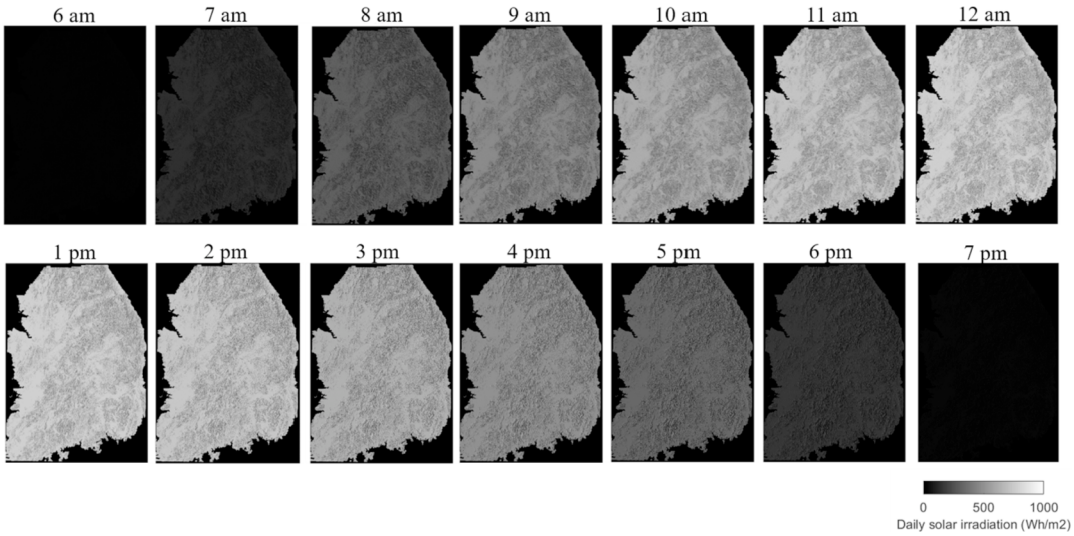

Fig. 6. Changes in solar irradiation potential at an hourly interval in South Korea on August 1, 2017.

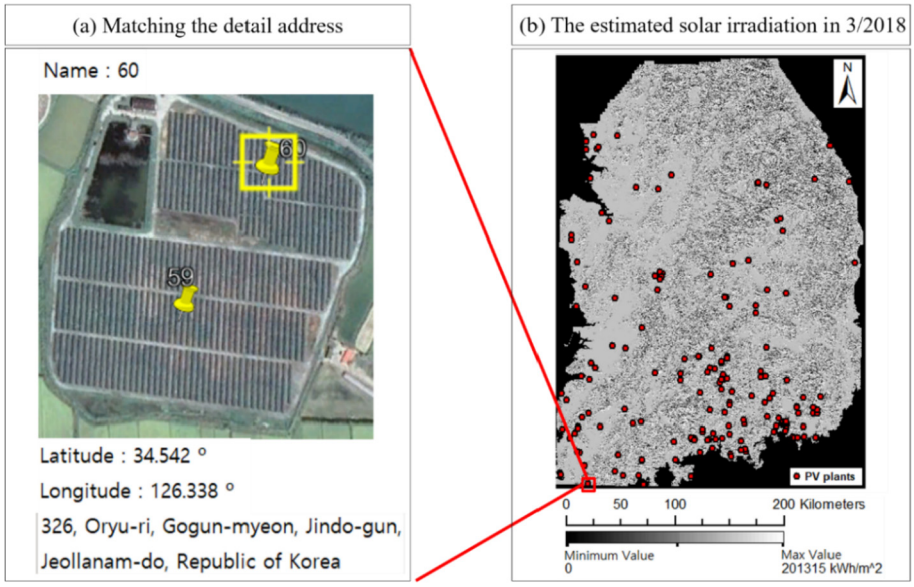

Fig. 7. Extraction of solar irradiation amount at a specific PV facility from solar irradiation maps.

Table 3
Description of input and output data used for forecasting model.

Input variable	Source	Unit	Temporal resolution	Reference
Month of operation	Korea Power Exchange	Month	Monthly	Kalogirou (2001)
Solar irradiation	Estimated using [40]	kWh/m^2		Pedro and Coimbra (2012)
Temperature	Korea Meteorological Administration	°C		Chen et al. (2011)
Relative humidity		%		Mekhilef et al. (2012)
Wind speed		m/s		Chen et al. (2011)
Precipitation		mm		Izgi et al. (2012)
Cloud amount		1/10		Sobri et al. (2018)
Duration of sunshine		Hour		Shivashankar et al. (2016)
Output variable	**Source**	**Unit**	**Temporal resolution**	**Reference**
PV power generation per unit capacity	Korea Power Exchange	Hour	Monthly	Antonanzas et al. (2016)

3.2. Model setting: hyper-parameter optimization

Hyper-parameters (e.g., epochs, batch sizes, initial weights, and learning rates) have a huge effect on the results produced by most machine learning algorithms because they configure the level of model complexity during the learning process (Claesen and De

Table 4
Hyper-parameters selected for the experiment.

Hyper-parameter	Value/method
Epoch	150
Batch size	32
Learning rate	0.001
Network weight initialization	Xavier initialization
Weight optimization algorithm	Adam

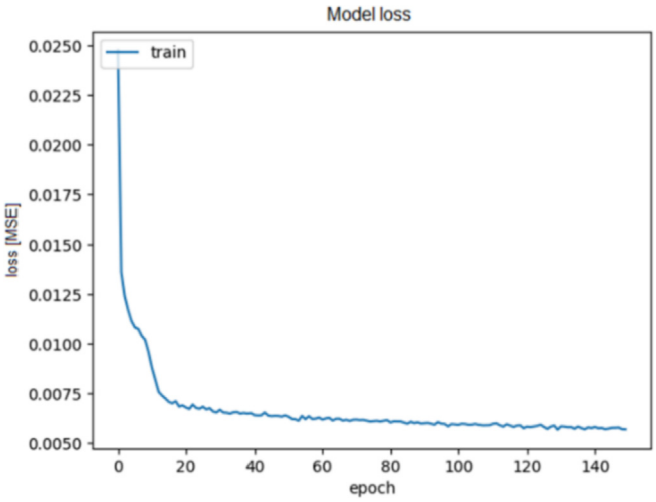

Fig. 8. Loss improvement over the training epochs.

Moor, 2015). To maximally generalize the model performance, hyper-parameters such as epoch and batch size (i.e., decision variables) were optimized through the grid search method, which divides the hyper-parameter space into grids and selects optimal hyper-parameters by calculating and comparing an error (i.e., objective variable) at each grid point. Through this optimization process (e.g., epochs of 50, 100, and 150 and batch sizes of 8, 16, 24, and 32), 150 epochs with a batch size of 32 were obtained, as shown in Table 4. For the network weights, Xavier initialization was adopted for the normalized weight initialization of recurrent matrices (Glorot and Bengio, 2010), and adaptive moment estimation (Adam) with a learning rate of 0.001, which is a simple and efficient algorithm for large datasets (Kingma and Ba, 2014), was selected as the gradient-based optimization algorithm to optimize weight parameters through the learning process. Based on the selected hyper-parameters, the learning processes were visually assessed, as illustrated in Fig. 8, and they show a loss improvement over the epochs. In other words, the loss value decreases continuously as the number of epochs increases until the 150th epoch.

4. Results

The proposed LSTM-RNN model was built using training datasets composed of 80% of the PV facilities and tested with a testing dataset to represent new facilities (the other 20% of actual facilities). In the training, time steps from one to twelve were tested, and the time step of ten produced the smallest RMSE, so it was selected. Table 5 summarizes the results of the error analysis for both the training and testing datasets. Overall, normalized root-mean-square-errors (nRMSEs) of 7.705% and 7.416% and MAPEs of 11.535% and 10.805% were achieved for training and testing, respectively. In this experiment, the testing errors were slightly smaller than the training errors, which suggests that the forecasting model was adequately trained. Notably, these results also imply that the proposed approach using historical data from existing PV facilities could be applied to predict the power generation potential of new facilities. In addition, the nRMSEs are smaller than the MAPEs, which indicates that the forecasting model was iteratively trained to minimize the sum of MSEs, which is used as a loss function, rather than the MAPEs, which represent relative errors. For the same amount of error, a higher MAPE is obtained for a small output (i.e., power generation per unit capacity), and a lower MAPE is obtained for a large output. That is, a larger relative error (i.e., a larger MAPE) can occur for a smaller output. Fig. 9 visually illustrates the forecasting performance: most points are plotted around the red reference line. Moreover, the smaller outputs are slightly over-estimated, whereas larger outputs are slightly under-estimated in both the training and testing results.

To investigate temporal patterns in the time-series datasets, times steps from one month to twelve months were tested, for which twelve forecasting models were built and compared in terms of errors, as presented in Fig. 10 and Table 6. With time steps of four and ten months, which use temporal patterns in the datasets from the previous four and ten months, respectively, smaller errors were obtained. In this experiment, the amount of information was held to be the identical (i.e., the same dataset was used for all time steps), which results in a relatively smaller number of data points in training with larger time steps. For example, with twelve data points, only one data point (twelve months) was used for a time step of twelve, whereas twelve data points (one per month) were used for a time step of one. Accordingly, more data points were used to train the forecasting model with a time step of four than with a time step of ten. Because the model trained with a time step of ten slightly outperformed that with a time step of four, despite the tradeoff between the amount of information and the number of training data points, the time step of ten months was used in further experiments.

The errors for each month listed in Table 7 were computed to investigate the effects of seasonal patterns on the prediction. Additionally, the boxplot in Fig. 11 shows the normalized residual errors in the testing data for each month. Here, the boxplot includes five elements: the lower extreme, 25th percentile error, median, 75th percentile error, and upper extreme (Williamson et al., 1989). Overall, all the median residuals are near 0, ranging from −0.010 to 0.025, and the monthly nRMSEs, except for spring (March, April, and May), are lower than 7.705% (i.e., the mean of the cross-validation results in Table 5). Specifically, as shown in Fig. 11,

Table 5
Results of performance evaluation with a time step of 10.

Time step	Evaluation metric	Training dataset: K-fold cross-validation (K = 5)						Testing dataset
		Run 1	Run 2	Run 3	Run 4	Run 5	Mean	
10	RMSE (h)	17.825	11.960	14.578	14.048	14.325	14.547	14.003
	nRMSE (%)	9.441	6.334	7.721	7.440	7.587	7.705	7.416
	MAPE (%)	13.832	9.119	11.396	11.428	11.898	11.535	10.805
	R^2	0.611	0.725	0.763	0.757	0.758	0.723	0.724

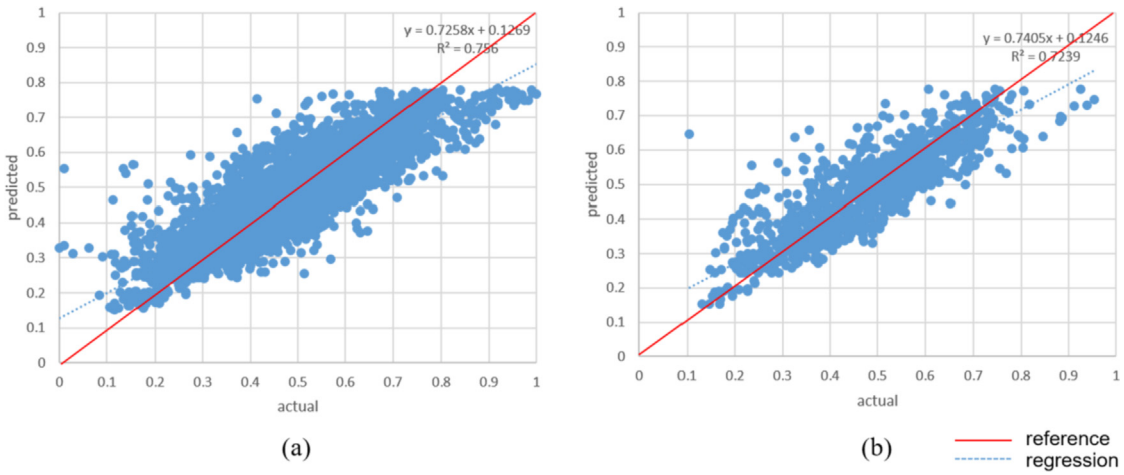

Fig. 9. Predicted vs. actual outputs with regression analysis (time step of 10): (a) training data, and (b) testing data.

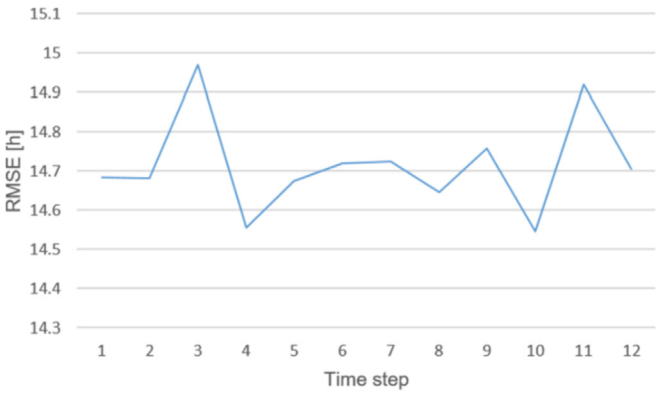

Fig. 10. RMSE change by time step.

relatively large median residual errors were observed in March, September, and October. Nevertheless, the nRMSEs and standard deviations in September and October were relatively small, with the highest values in both measures observed in March. This phenomenon becomes explicit when the errors are compared with respect to seasons. Here, each season is classified by its three-month period (Table 7). The nRMSE for spring (8.516%) is higher than the others, and that of fall (6.089%) is the lowest. Spring and summer have particularly high numbers of outliers, suggesting that unknown factors not included in this experiment might be affecting power generation. The correlation between the seasonal amount of particulate matter (PM 10) (collected from Korean Statistical Information Service , 2019) and the seasonal errors in Table 7 is strong (correlation coefficient of 0.761), implying that dust might have influenced the amount of solar radiation arriving at the PV panel. Further research is required to identify other factors that affect PV power generation.

As an example of the forecasting results, the predicted power outputs are plotted in a time domain in Fig. 12, showing that the temporal trends are efficiently captured over time for both large and small capacity PV samples. Particularly, two groups of PV facilities, divided based on the median of PV capacities (1.373 MW) were compared because the PV plants considered in this experiment are biased toward small capacities (e.g., the PV capacities plotted on the right y-axis in Fig. 13). Fig. 13 illustrates the residual errors in the increasing order of PV capacity size, and no significant difference was observed in the errors between the large and small groups. When a Mann-Whitney test, which is a nonparametric test to compare differences between two independent groups

Table 6
Time step testing results in performance evaluation using five-fold cross-validation.

Time step (month)	1	2	3	4	5	6	7	8	9	10	11	12
RMSE (h)	14.684	14.681	14.969	14.556	14.674	14.721	14.724	14.647	14.758	14.547	14.919	14.705
nRMSE (%)	7.777	7.775	7.928	7.709	7.772	7.797	7.749	7.757	7.816	7.705	7.901	7.788
MAPE (%)	11.599	11.579	11.796	11.398	11.612	11.684	11.808	11.626	11.810	11.535	12.086	11.693
R^2	0.714	0.715	0.711	0.717	0.714	0.714	0.715	0.722	0.722	0.723	0.718	0.715

Table 7
Statistical summary of monthly residuals in testing results.

Measure	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Max residual	0.158	0.110	0.208	0.237	0.209	0.209	0.140	0.163	0.185	0.132	0.138	0.137
Min residual	−0.244	−0.233	−0.541	−0.270	−0.302	−0.196	−0.319	−0.311	−0.214	−0.141	−0.176	−0.208
Median residual	−0.002	−0.010	0.017	−0.006	0.000	0.005	−0.005	0.001	0.019	0.025	0.009	0.003
nRMSE (%)	7.789	7.674	9.020	7.698	8.830	7.124	7.466	7.412	5.866	6.496	5.903	7.030
Standard deviation	0.078	0.076	0.091	0.077	0.089	0.071	0.074	0.074	0.057	0.063	0.059	0.071

Measure	Spring(Mar., Apr., May)	Summer(Jun., Jul., Aug.)	Fall(Sep., Oct., Nov.)	Winter(Dec., Jan., Feb.)
nRMSE (%)	8.516	7.334	6.089	7.498

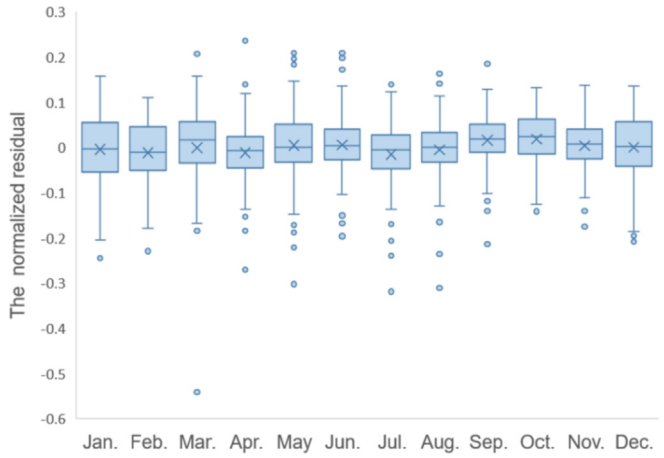

Fig. 11. Monthly residuals in testing results.

(Fagerland and Sandvik, 2009), was applied to the residual errors computed with the training data used to build the model, that result was also not statistically significant (p-value > 0.05). However, when the Mann-Whitney test was applied to the testing data, the difference was statistically significant, possibly caused by the random sampling of the testing data in terms of PV capacity. In addition, the Mann-Whitney test results show a statistical difference between the MAPEs of the two groups (e.g., p-value < 0.05); specifically, the MAPEs in the large capacity group (9.571%) are slightly smaller than those in the small capacity group (11.777%). This result suggests that the estimated amount of PV power generation is more sensitive to errors for small PV plants than for large ones.

5. Discussion

The amount of available solar energy, which directly affects PV power generation, varies with location and time. To incorporate topographical and temporal patterns into PV power forecasting, this study presents and evaluates a LSTM-RNN model that predicts the monthly power output of PV facilities using historical data. Moreover, terrain datasets, such as DEMs corresponding to the region in which a PV facility is located, were used to estimate the potential solar irradiation at each PV facility. In that way, geographical variations in the locations of PV systems could be better captured in the model, compared to using the longitude and latitude as input variables. The datasets collected to build the network model include power generation information for 63 months, and a time horizon of one month (larger than most papers in Table 1) was selected to predict the monthly power output. The temporal patterns in those monthly datasets were learned by a stacked LSTM-RNN network with memory cells that recall the temporal effect during the learning process. The results imply that temporal patterns exist in the monthly datasets because similar patterns were found in the data in a time resolution of 3 h (Gensler et al., 2016) and 1 h (Abdel-Nasser and Mahmoud, 2017). Overall, the proposed model predicted the power output of new PV plants (the testing dataset) on a long-term, wide regional scale, achieving an nRMSE of 7.416% and a MAPE of 10.805%. This is a notable result in comparison to previous machine learning–based studies. For example, Gensler et al. (2016) achieved the lowest previous nRMSE of 7.13% for 21 plants in a short-term prediction horizon (one day

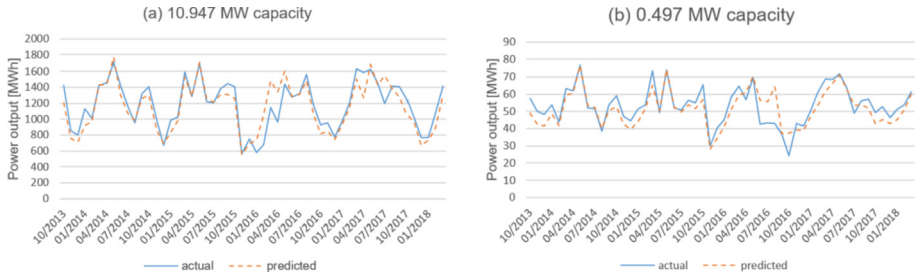

Fig. 12. Comparison between actual data and predicted PV power in a time domain: PV plant samples (a) with 10.947 MW capacity and (b) with 0.497 MW capacity.

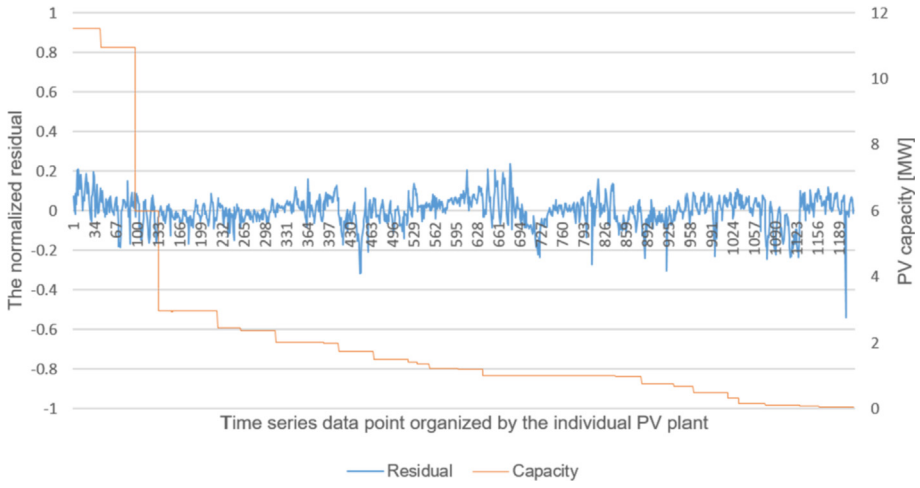

Fig. 13. Comparison between actual and predicted PV power for testing data points in increasing order of PV plant capacity.

ahead), although direct comparisons cannot be made due to the different datasets used to build the models. The results thus show that the monthly or yearly potential for PV power generation can be estimated using relevant historical data from existing plants and then applied to select suitable locations for new solar PV systems.

In this study, the one-month time horizon was selected to investigate the seasonal effects of meteorological conditions on power generation. The results show that forecasting errors in summer are similar to those in winter (nRMSEs of 7.334% in summer and 7.498% in winter). This similarity could be caused by similar precipitation patterns in the two seasons (e.g., rain and snow). In contrast, a difference of 2.427% in RMSEs was observed between spring and fall (nRMSEs of 8.516% in spring and 6.089% in fall). The maximum nRMSE difference between months was 3.153% between March (9.020%) and September (5.866%), which could be affected by seasonal conditions such as yellow dust. This result implies that although other variables (e.g., dust) might further be considered to reduce the error, the seasonal patterns can be discerned when using monthly data.

The spatial resolution of the terrain data should be selected carefully to reflect the purpose and target of the power forecasting. In this study, a resolution of 30×30 m^2 was selected based on the common sizes of PV facilities in the study area. However, when forecasting for residential PV systems, which are significantly smaller than commercial and utility-scale systems, higher resolutions could be required to extract precise topographical features from the terrain data and estimate a PV facility's area, which is strongly related to the amount of solar energy available.

Further research is recommended to address the following issues. First, the PV power data collected in this study have a bias toward plants with a low capacity (e.g., a mean of 2.175 MW and a standard deviation of 2.737 MW). Consequently, the difference in the MAPEs of the two groups with different capacities was statistically significant. That implies that the forecasting model could be further trained to improve its accuracy by training with more high-capacity plants. Second, the scope of this work is limited to predicting PV power generation without weather forecasting. However, because weather conditions are used as inputs for the proposed forecasting model, historical weather data (e.g., the Korean standard weather data in Passive House Institute Korea, 2017) or numerical weather prediction data (e.g., Gensler et al., 2016; Gao et al., 2019) could be used to predict future power output. Third, the presented model was tested only for PV facilities in South Korea, which has complex and mountainous terrain. Further research is thus required to generalize the findings to other regions.

6. Conclusion

This study has presented a stacked LSTM-RNN model for predicting the monthly PV power output of potential solar PV systems at new sites. For the power prediction, the proposed model learns the complex and nonlinear patterns between the power output and various influencing factors, such as weather conditions (i.e., temperature, relative humidity, wind speed, cloud amount, precipitation, and duration of sunshine), electricity generation per unit capacity of widely spread PV plants, and the estimated amount of solar irradiation based on the topographical conditions of each individual PV facility. Through experiments using historical training data from 134 PV facilities during approximately five years, the proposed approach was found to perform well, recognizing and predicting the varying topographical and changing meteorological conditions in the time-series testing datasets of 30 other PV facilities not used in the training process. The major findings and significance of the proposed approach to PV power forecasting are summarized as follows: (1) the proposed model can predict the power output of completely different PV plants using only one model, achieving an nRMSE of 7.416%, equivalent to an RMSE of 14.003 h and a MAPE of 10.805%; (2) by estimating the potential solar irradiation using terrain data, topographical features from 164 PV plants widely spread over the country were integrated into a single forecasting model; (3) for temporal patterns, LSTM-RNN models built and tested with time-series data at various time steps (1–12 months) had the smallest errors (RMSE of 14.547 h and a MAPE of 11.535%) at 10 months; (4) as validated with actual PV plant data, the proposed approach is suitable for PV system planning at a regional level for plants with large or small capacity (MAPE of 9.571% for the large-capacity group and 11.777% for the small-capacity group). The proposed model can thus be used to scrutinize potential areas for the PV plant installation, not only within certain known regions but any place for which terrain data and historical weather data are available. The long-term resolution forecasting available with the LSTM-RNN model could help support the energy planning and feasibility analysis of any plant by providing the key input data for decision making.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was supported by a grant from the Technology Advancement Research Program funded by the Ministry of Land, Infrastructure and Transport (MLIT) of Korea (19CTAP-C141728-02). Any opinions in this paper are those of the authors and do not necessarily represent those of the MLIT.

References

Abdel-Nasser, M., Mahmoud, K., 2017. Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput. Appl. 1, 14.
Al-Soud, M.S., Hrayshat, E.S., 2009. A 50 MW concentrating solar power plant for Jordan. J. Clean. Prod. 17 (6), 625–635.
Almonacid, F., Pérez-Higueras, P.J., Fernández, E.F., Hontoria, L., 2014. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. Energy Convers. Manag. 85, 389–398.
Alonzo, B., Ringkjob, H.K., Jourdier, B., Drobinski, P., Plougonven, R., Tankov, P., 2017. Modelling the variability of the wind energy resource on monthly and seasonal timescales. Renew. Energy 113, 1434–1446.
Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. Sol. Energy 136, 78–111.
Chen, C., Duan, S., Cai, T., Liu, B., 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol. Energy 85 (11), 2856–2870.
Claesen, M., De Moor, B., 2015. Hyperparameter Search in Machine Learning arXiv preprint. arXiv:1502.02127.
Coulibaly, P., Anctil, F., Bobee, B., 2011. Multivariate reservoir inflow forecasting using temporal neural networks. J. Hydrol. Eng. 6 (5), 367–376.
Craig, P.P., Gadgil, A., Koomey, J.G., 2002. What can history teach us? A retrospective examination of long-term energy forecasts for the United States. Annu. Rev. Energy Environ. 27 (1), 83–118.
Dalton, G.J., Lockington, D.A., Baldock, T.E., 2008. Feasibility analysis of stand-alone renewable energy supply options for a large hotel. Renew. Energy 33 (7), 1475–1490.
Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Van Deventer, W., et al., 2018. Forecasting of photovoltaic power generation and model optimization: a review. Renew. Sustain. Energy Rev. 81, 912–928.
De Felice, M., Alessandri, A., Catalano, F., 2015. Seasonal climate forecasts for medium-term electricity demand forecasting. Appl. Energy 137, 435–444.
Fagerland, M.W., Sandvik, L., 2009. The wilcoxon–mann–whitney test under scrutiny. Stat. Med. 28 (10), 1487–1497.
Fesharaki, V.J., Dehghani, M., Fesharaki, J.J., Tavasoli, H., 2011. The effect of temperature on photovoltaic cell efficiency. In: Proceedings of the 1stInternational

Conference on Emerging Trends in Energy Conservation—ETEC, pp. 20–21. Tehran, Iran.
Gao, M., Li, J., Hong, F., Long, D., 2019. Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM. Appl. Sci. 9 (15), 3192.
Gastli, A., Charabi, Y., 2010. Solar electricity prospects in Oman using GIS-based solar radiation maps. Renew. Sustain. Energy Rev. 14 (2), 790–797.
Gensler, A., Henze, J., Sick, B., Raabe, N., 2016. Deep Learning for solar power forecasting—an approach using AutoEncoder and LSTM Neural Networks. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp. 2858–2865.
Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: continual prediction with LSTM. Neural Comput. 12 (10), 2451–2471.
Giles, C.L., Lawrence, S., Tsoi, A.C., 1997. Rule inference for financial prediction using recurrent neural networks. In: Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr). IEEE, pp. 253–259.
Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256.
Gueymard, C.A., Thevenard, D., 2009. Monthly average clear-sky broadband irradiance database for worldwide solar heat gain and building cooling load calculations. Sol. Energy 83 (11), 1998–2018.
Hailegnaw, B., Kirmayer, S., Edri, E., Hodes, G., Cahen, D., 2015. Rain on methylammonium lead iodide based perovskites: possible environmental effects of perovskite solar cells. J. Phys. Chem. Lett. 6 (9), 1543–1547.
Han, S., Qiao, Y.H., Yan, J., Liu, Y.Q., Li, L., Wang, Z., 2019. Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. Appl. Energy 239, 181–191.
Hochreiter, S., Schmidhuber, J., 1997. LSTM can solve hard long time lag problems. In: Advances in Neural Information Processing Systems, vol. 9. NIPS, pp. 473–479.
Hontoria, L., Rus-Casas, C., Aguilar, J.D., Hernandez, J.C., 2019. An improved method for obtaining solar irradiation data at temporal high-resolution. Sustainability 11 (19), 5233.
Hossain, M., Mekhilef, S., Danesh, M., Olatomiwa, L., Shamshirband, S., 2017. Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems. J. Clean. Prod. 167, 395–405.
International Energy Agency (IEA), 2018. Renewables information 2018: overview. https://webstore.iea.org/renewables-information-2018-overview accessed 19 July 2018.
International Finance Corporation (IFC), 2019. Utility-scale solar photovoltaic power plants: a project developer's guide. https://www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/sustainability-at-ifc/publications/publications_utility-scale+solar+photovoltaic+power+plants accessed 30 March 2019.
Izgi, E., Öztopal, A., Yerli, B., Kaymak, M.K., Şahin, A.D., 2012. Short–mid-term solar power prediction by using artificial neural networks. Sol. Energy 86 (2), 725–733.
Jain, A., Mehta, R., Mittal, S.K., 2011. Modeling impact of solar radiation on site selection for solar PV power plants in India. Int. J. Green Energy 8 (4), 486–498.
Jung, J., Han, S., Kim, B., 2019. Digital numerical map-oriented estimation of solar energy potential for site selection of photovoltaic solar panels on national highway slopes. Appl. Energy 242, 57–68.
Kalogirou, S.A., 2001. Artificial neural networks in renewable energy systems applications: a review. Renew. Sustain. Energy Rev. 5 (4), 373–401.
Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization arXiv preprint. arXiv:1412.6980.
Korean Statistical Information Service (KOSIS), 2019. Fine dust (PM10) monthly air pollution by city, extracted from air pollution status by Ministry of Environment, Korea.]. https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_03_0200045. accessed June 2019.
Liu, J., Xu, F., Lin, S., 2017. Site selection of photovoltaic power plants in a value chain based on grey cumulative prospect theory for sustainability: a case study in Northwest China. J. Clean. Prod. 148, 386–397.
Lin, K.P., Pai, P.F., 2016. Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression. J. Clean. Prod. 134, 456–462.
Long, H., Zhang, Z., Su, Y., 2014. Analysis of daily solar power prediction with data-driven approaches. Appl. Energy 126, 29–37.
Mekhilef, S., Saidur, R., Kamalisarvestani, M., 2012. Effect of dust, humidity and air velocity on efficiency of photovoltaic cells. Renew. Sustain. Energy Rev. 16 (5), 2920–2925.
Mellit, A., Pavan, A.M., 2010. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy. Sol. Energy 84 (5), 807–821.
Passive House Institute Korea, 2017. Standard weather data from 70 locations in Korea (2017 Ver.). http://www.phiko.kr/bbs/board.php?bo_table=z3_01&wr_id=2479. accessed 1 September, 2017.
Pedro, H.T., Coimbra, C.F., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol. Energy 86 (7), 2017–2028.
Shi, J., Lee, W.J., Liu, Y., Yang, Y., Wang, P., 2012. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. IEEE Trans. Ind. Appl. 48 (3), 1064–1069.
Shivashankar, S., Mekhilef, S., Mokhlis, H., Karimi, M., 2016. Mitigating methods of power fluctuation of photovoltaic (PV) sources–A review. Renew. Sustain. Energy Rev. 59, 1170–1184.
Sobri, S., Koohi-Kamali, S., Rahim, N.A., 2018. Solar photovoltaic generation forecasting methods: a review. Energy Convers. Manag. 156, 459–497.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.
Šúri, M., Huld, T.A., Dunlop, E.D., 2005. PV-GIS: a web-based solar radiation database for the calculation of PV potential in Europe. Int. J. Sustain. Energy 24 (2), 55–67.
Williamson, D.F., Parker, R.A., Kendrick, J.S., 1989. The box plot: a simple visual method to interpret data. Ann. Intern. Med. 110 (11), 916–921.
Wolff, B., Kühnert, J., Lorenz, E., Kramer, O., Heinemann, D., 2016. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. Sol. Energy 135, 197–208.
Yeo, I.A., Yee, J.J., 2014. A proposal for a site location planning model of environmentally friendly urban energy supply plants using an environment and energy geographical information system (E-GIS) database (DB) and an artificial neural network (ANN). Appl. Energy 119, 99–117.
Yona, A., Senjyu, T., Saber, A.Y., Funabashi, T., Sekine, H., Kim, C.H., 2008. Application of neural network to 24-hour-ahead generating power forecasting for PV system. In: 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century. IEEE, pp. 1–6.
Zaremba, W., Sutskever, I., Vinyals, O., 2014. Recurrent Neural Network Regularization arXiv preprint. arXiv:1409.2329.