

Article

Multi-Site Photovoltaic Forecasting Exploiting Space-Time Convolutional Neural Network

Jaeik Jeong and Hongseok Kim * 

Department of Electronic Engineering, Sogang University, Baekbeom-ro 35, Mapo-gu, Seoul 04107, Korea; jaeik1213@sogang.ac.kr

* Correspondence: hongseok@sogang.ac.kr; Tel.: +82-2-705-7989

Received: 8 October 2019; Accepted: 21 November 2019; Published: 25 November 2019



Abstract: The accurate forecasting of photovoltaic (PV) power generation is critical for smart grids and the renewable energy market. In this paper, we propose a novel short-term PV forecasting technique called the space-time convolutional neural network (STCNN) that exploits the location information of multiple PV sites and historical PV generation data. The proposed structure is simple but effective for multi-site PV forecasting. In doing this, we propose a greedy adjoining algorithm to preprocess PV data into a space-time matrix that captures spatio-temporal correlation, which is learned by a convolutional neural network. Extensive experiments with multi-site PV generation from three typical states in the US (California, New York, and Alabama) show that the proposed STCNN outperforms the conventional methods by up to 33% and achieves fairly accurate PV forecasting, e.g., 4.6–5.3% of the mean absolute percentage error for a 6 h forecasting horizon. We also investigate the effect of PV sites aggregation for virtual power plants where errors from some sites can be compensated by other sites. The proposed STCNN shows substantial error reduction by up to 40% when multiple PV sites are aggregated.

Keywords: multi-site photovoltaic forecasting; spatio-temporal correlation; space-time matrix; CNN

1. Introduction

Recently, various renewable energy sources have been integrated into power grids to resolve fossil fuel depletion and environmental problems such as global warming. Among various renewable energy sources, photovoltaic (PV) power generation has received considerable attention because of its abundance and cleanness. The Paris Agreement has recently stressed the necessity of using renewable energy, and global penetration of PV generation rapidly increases [1]. However, PV power outputs heavily depend on weather conditions such as clouds, temperature, and humidity, and this causes substantial uncertainties, which brings adverse effects on economic benefit and the stability of power grids. Thus, accurate PV forecasting techniques are needed to exploit economical advantages and balance the power grid.

In the literature, many forecasting schemes have been proposed to improve forecasting accuracy. Since non-stationary characteristics of solar irradiance mainly come from cloud movement and its stochastic blocking of sunlight [1], the works in [2,3] used cloud motion vector schemes using sky imagers. However, sky imagers are too costly to be deployed with all PV sites, and they are also only capable of predicting in a very short-term horizon, i.e., less than an hour. The works in [4,5] use satellite images for hourly PV output forecasting. However, wide-area satellite images are not capable of capturing site-specific information, and thus not good for site-specific PV forecasting. The works in [6,7] use *forecasted* cloud information to improve forecasting accuracy. However, these schemes heavily depend on meteorological administration or the national weather service, and thus may not be adequate when weather forecasting is not accurate.

In this regard, several studies have recently used multi-site spatio-temporal historical data for PV forecasting without requiring exogenous cloud data [8–16]. For example, simple and fast multi-site forecasting techniques using autoregressive (AR) are used in [8,9], and the work in [10] further develops the AR model into the local vector AR model considering local climate change. However, nonlinear methods outperform the linear techniques for forecasting horizons more than an hour [17], and recent studies have used machine learning techniques to predict multi-site PV output for hourly horizons. For example, feedforward neural network (FNN) [11] and long short-term memory (LSTM) [12–14] are proposed for multi-site spatio-temporal PV forecasting. Artificial neural networks are also widely used in many other fields such as mechanical engineering [18], industrial engineering [19], and economy [20] as well as PV forecasting. Since LSTM is good for time-series forecasting by processing sequence information using internal memory, forecasting accuracy has been improved. In addition, spatial information can trace the indirect showing of cloud image, and recent studies exploit convolutional neural network (CNN) [15,16] to capture the complex spatial dependencies. As passing clouds certainly influence neighboring PV sites sequentially, they can capture cloud cover and cloud movement by considering spatial relations with improved forecasting performance.

Previous spatio-temporal solar forecasting studies had some limitations. First, many studies did not fully exploit the spatial information of multiple PV sites [8–14]. The PV values are simply concatenated as one vector without considering the locational relations, where tracing cloud movements cannot be exploited. Second, recent studies exploiting spatial information have very high model complexity as locational relations of the PV sites are represented by a large number of matrices [15] or complex graphs [16], which in turn require very long training data (e.g., at least more than five years) and thus is not suitable for newly-built PV sites. One might think that transfer learning can be applied by transferring the information obtained from data-rich sites to a newly-built sites. However, newly-built PV sites have their own locational information, which cannot be learned from spatial information of other sites. In this regard, a simple and low complex forecasting model capturing spatial and temporal features is required.

In this regard, we propose a simple space-time convolutional neural network (STCNN) for multi-sites PV forecasting. The STCNN represents spatial and temporal information in the rows and the columns of the input matrix. This technique is generally used in traffic forecasting [21–23] as traffic can be represented in one-dimension. Thus, spatio-temporal information of the traffic can be easily learned without a complex model structure or massive training data. We leverage this method in the area of PV forecasting to learn spatial and temporal features simultaneously with low computational requirements. In order to capture the spatial relations of the multi-site PV generation into one-dimension, we apply an ordering method called a greedy adjoining algorithm (GAA) to serialize two-dimensional PV site information while keeping the local PV sites adjacent to each other. This enables indirect capturing of cloud cover and cloud movements with simple matrix structure. The proposed method is particularly effective when the periods of historical PV data are not sufficiently long, e.g., just one year; this can be a typical case for many recently built PV sites. However, when a longer period of historical data is available (e.g., more than 5 years), methods with higher complexity such as [15,16] might be more accurate.

We summarize the contributions of this paper as follows. We develop a framework for multi-site PV forecasting technique using STCNN in the area of PV forecasting. Our model is based on CNN and takes a space-time matrix as an input, which is derived from geographic location and the historical datasets of multi-site PV generation to learn cloud cover and movement indirectly. Two-dimensional geographical PV site information is serialized into one dimension by using the GAA, which enables using a space-time matrix in the PV forecasting area. To verify the proposed forecasting technique, we perform extensive experiments based on diverse multi-site PV generation data. Specifically, to reflect various geographic conditions, we use 238 sites in California, 67 sites in New York, and 103 sites in Alabama, i.e., PV generation from west, east, and southeast parts of the US. The experimental

results confirm that the proposed STCNN outperforms other existing models of AR, FNN, and LSTM for short-term PV forecasting. For example, for 6 h forecasting horizon, the proposed algorithm achieves 4.6%, 4.8%, and 5.3% of the mean absolute percentage errors for California, New York and Alabama, respectively. These results have up to a 33% improvement compared to the popular existing methods. In addition, we find that our model is particularly effective for an aggregated forecasting for multiple PV sites, e.g., for virtual power plant (VPP) application [24]. When tracking the cloud movements is somewhat incorrect, it predicts some regions to have more clouds, which induces under-forecasting, while some other regions to have fewer clouds, which induces over-forecasting. This results in the balanced forecasting when aggregated, and errors from over-forecasted sites can be compensated by other under-forecasted sites. When the generations of multiple PV sites are aggregated, we show that the proposed STCNN has the highest error reduction by up to 40% compared to the existing methods of [10–12].

The rest of this paper is organized as follows. In Section 2, we describe the methodologies and propose the STCNN based forecasting technique. In Section 3, we describe the model selection of STCNN. In Section 4, performance evaluations are provided, followed by the conclusion in Section 5.

2. Proposed Methodology

2.1. Space-Time Matrix

The proposed GAA serializes the PV generating sites so that the local PV sites stay adjacent to each other in ordering, which transfers two-dimensional spatial information of the PV sites into one-dimension. Let $D(p, q)$ denote the distance between two PV sites denoted by p and q . The distance can be defined as geographic distance, measured by latitude and longitude information such as

$$D(p, q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}, \quad (1)$$

where p_x, q_x are the latitude positions and p_y, q_y are the longitude positions of the sites, respectively.

To serialize PV sites, we need to select the first PV site. Since the PV output values of the first site are located at the edge of the space-time matrix, the first PV site experiences low-level features [25]. Thus, we select the most remote PV site as the first PV site, denoted by $s(1)$, as follows:

$$s(1) = \operatorname{argmax}_{q \in \mathcal{P}_1} \sum_{p \in \mathcal{P}_1} D(p, q), \quad (2)$$

where $\mathcal{P}_1$ denotes a set of all PV sites. Let $s(n)$ denote the n -th selected PV site. Then, $s(n)$ is determined by

$$s(n) = \operatorname{argmin}_{q \in \mathcal{P}_n} D(s(n-1), q), \quad 2 \leq n \leq N, \quad (3)$$

where $\mathcal{P}_n = \mathcal{P}_{n-1} \setminus \{s(n-1)\}$ and $N = |\mathcal{P}_1|$. In this way, we repeatedly select $s(n)$, which is the closest to $s(n-1)$ in a set of unselected sites $\mathcal{P}_n$. This process results in a vector $[s(1) \cdots s(N)]^T$ that implicitly contains the spatial relationship of PV sites.

We now construct a space-time matrix $\mathbf{X}_t$ at time slot t , which is the input of CNN. Spatial and temporal information is represented in the rows and the columns of the input matrix. Let $x_{s(n),t}$ denote the PV output value at time slot t on the site $s(n)$. When we use the past M values at time slot t , the n -th row consists of $[x_{s(n),t-M+1}, x_{s(n),t-M+2}, \dots, x_{s(n),t}]$, and thus we have

$$\mathbf{X}_t = \begin{bmatrix} x_{s(1),t-M+1} & x_{s(1),t-M+2} & \cdots & x_{s(1),t} \\ x_{s(2),t-M+1} & x_{s(2),t-M+2} & \cdots & x_{s(2),t} \\ \vdots & \vdots & \ddots & \vdots \\ x_{s(N),t-M+1} & x_{s(N),t-M+2} & \cdots & x_{s(N),t} \end{bmatrix}. \quad (4)$$

2.2. Convolutional Neural Network

CNNs have a great ability of extracting features from the input matrix and have been applied to many prediction applications with feature extractions [15,16,21–23]. The CNN for multi-site PV forecasting has three types of layers: convolutional layer, max pooling layer, and fully connected layer. The convolutional layer has the locally-connected property, which makes output neurons connected to their local nearby PV generation. Kernels are used for convolution operation, and the number of kernels determines how many features are extracted. The max pooling layer combines the cluster of output neurons into a single neuron to select the maximum value of each cluster. After the features are extracted, the output features are concatenated into a dense vector. This vector is fully connected with the output.

To predict the next H hours of PV generation, the dimension of the output matrix is given by $N \times H$. Let the predicted output space-time matrix at time slot t be $\hat{\mathbf{Y}}_t$ and the target output space-time matrix at time slot t be $\mathbf{Y}_t$ such as

$$\hat{\mathbf{Y}}_t = \begin{bmatrix} \hat{x}_{s(1),t+1} & \hat{x}_{s(1),t+2} & \cdots & \hat{x}_{s(1),t+H} \\ \hat{x}_{s(2),t+1} & \hat{x}_{s(2),t+2} & \cdots & \hat{x}_{s(2),t+H} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{x}_{s(N),t+1} & \hat{x}_{s(N),t+2} & \cdots & \hat{x}_{s(N),t+H} \end{bmatrix}, \quad (5)$$

$$\mathbf{Y}_t = \begin{bmatrix} x_{s(1),t+1} & x_{s(1),t+2} & \cdots & x_{s(1),t+H} \\ x_{s(2),t+1} & x_{s(2),t+2} & \cdots & x_{s(2),t+H} \\ \vdots & \vdots & \cdots & \vdots \\ x_{s(N),t+1} & x_{s(N),t+2} & \cdots & x_{s(N),t+H} \end{bmatrix}, \quad (6)$$

where $\hat{x}_{s(n),t+h}$ is the predicted PV output value at time slot $t+h$ on the site $s(n)$. Our objective is to find the model parameters that minimize the error between $\hat{\mathbf{Y}}$ and $\mathbf{Y}$. The mean squared error (MSE) is employed, and our objective function is given by

$$\operatorname{argmin}_{\theta} \|\hat{\mathbf{Y}}_t - \mathbf{Y}_t\|_F^2, \quad (7)$$

where $\|\cdot\|_F$ is the Frobenius norm, and θ represents the model parameters. Using the back-propagation with gradient descent during training, weights and biases are adjusted with their gradients. The feature extraction, forecasting, and back-propagation process are repeated until convergence. After the training process is completed, the CNN model is used for PV power forecasting. The overall architecture of the proposed STCNN is summarized in Figure 1. Algorithm 1 summarizes the STCNN training process.

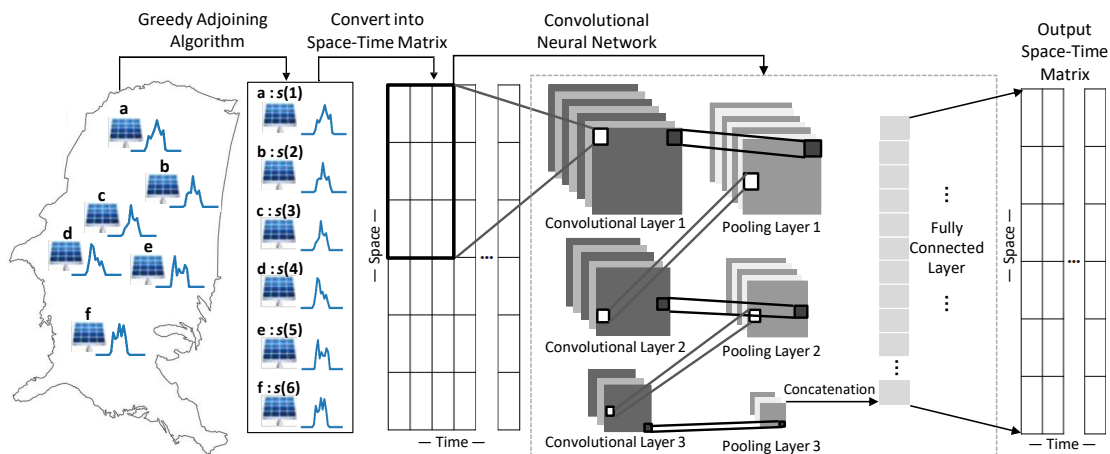


Figure 1. A framework of the proposed STCNN for multi-site PV forecasting.

Algorithm 1 : STCNN Training Algorithm**input**: a set of all PV sites $\mathcal{P}_1$; historical PV output dataset $\mathcal{D}$ **output**: Learned STCNN model

```

//greedy adjoining algorithm
1:  $s(1) \leftarrow \operatorname{argmax}_{q \in \mathcal{P}_1} \sum_{p \in \mathcal{P}_1} D(p, q)$ 
2: for  $n \leftarrow 2$  to  $|\mathcal{P}_1|$  do
3:    $\mathcal{P}_n \leftarrow \mathcal{P}_{n-1} \setminus \{s(n-1)\}$ 
4:    $s(n) = \operatorname{argmin}_{q \in \mathcal{P}_n} D(s(n-1), q)$ 
//space-time matrix construction
5:  $\mathcal{B} \leftarrow \phi$ 
6: for all available time slot  $t$  do
7:   construct  $\mathbf{X}_t$  according to (4) using  $\mathcal{D}$ 
8:   construct  $\mathbf{Y}_t$  according to (6) using  $\mathcal{D}$ 
9:   put an training instance  $(\mathbf{X}_t, \mathbf{Y}_t)$  into  $\mathcal{B}$ 
//model training
10: initialize all learnable model parameters  $\theta$  in STCNN
11: repeat
12:   select a batch of instances from  $\mathcal{B}$ 
13:   find  $\theta$  by minimizing (7) with the selected batch of instances
14: until convergence

```

3. Model Selection

There are three critical factors that need to be considered when designing the structure of STCNN: the size of input and output space-time matrices, hyperparameters related with convolutional layers and max pooling layers, and the depth of the STCNN. First, we describe PV generation data used for training and testing. Then, we describe how the hyperparameters of the proposed STCNN are chosen.

3.1. Data Description

We use the PV generation data of California, New York, and Alabama, i.e., west, east, and southeast parts of the US, released by the National Renewable Energy Laboratory (NREL) [26]. These one year long datasets are also used by other studies to evaluate the forecasting algorithms [13,27,28]. We consider the PV sites having fixed tilt PV modules equal to the latitude. Consequently, 238 sites in California, 67 sites in New York, and 103 sites in Alabama are selected. The data are normalized between 0 and 1 using the installed PV generation capacity, and sampled every 1 h. The dataset is split into training set (60%), validation set (20%), and test set (20%).

3.2. Hyperparameter Selection of the Proposed STCNN

The size of the space-time matrix is determined by the number of PV sites and time steps; the number of rows N is simply equal to the number of PV sites, but the number of time steps (i.e., M for the input matrix $\mathbf{X}$ and H for the output matrix $\mathbf{Y}$) need be chosen accordingly. Since a reasonable time horizon is 6 h when exogenous weather data are not used [17], we determine $H = 6$. Based on this, we determine $M = 18$ to reflect the entire day (24 h).

Then, we determine the kernel size of convolutional layers and the cluster size of max pooling layer. Since there is no general rule for these, we refer to the well-known CNN architectures such as AlexNet [29] and VGGNet [30]. As the size of the proposed space-time matrix is relatively smaller than conventional images, we select the smallest kernel size and cluster size of the referred networks. In this regard, the kernel size of convolutional layers is set to (3,3), and the cluster size of pooling layers is set to (2, 2).

Next, we choose the number of layers (depth) using the validation set. Batch normalization is applied to accelerate the training by reducing internal covariate shift and also to prevent overfitting [31]. For an activation function, rectified linear unit (ReLU) is applied to prevent the vanishing gradient problem [32]. All networks are initialized with He initialization [33] and trained based on the Adam optimizer [34] using the mini-batch size of 128. Our framework is built using Tensorflow [35].

Figure 2 shows the validation errors in the number of epochs for all three states. The number of kernels is set to the multiples of 32 as in the referred network. As can be seen, depth-1 models show the worst performance for all three states, but depth-2, depth-3, and depth-4 show similar performances. Table 1 summarizes the MSE of the validation set according to the depth of CNN. Model complexity is measured by the number of learnable model parameters in the CNN (i.e., weights and biases). Note that the depth-1 model has the highest model complexity as it has only one max pooling layer. Depth-3 shows the smallest MSE with the reasonable model complexity. In this regard, we select the depth-3 model with the number of kernels in each layer, 128, 64, and 32, respectively. In addition, the epoch is set to 150 that shows the smallest MSE. The final structure of the proposed STCNN is shown in Table 2.

Table 1. MSE of the validation set according to the depth of CNN.

	California			New York			Alabama		
	MSE	Epoch	Model Complexity	MSE	Epoch	Model Complexity	MSE	Epoch	Model Complexity
Depth-1	0.310	30	48,942,164	0.162	20	3,821,330	0.166	10	9,078,122
Depth-2	0.113	150	10,804,788	0.101	150	842,802	0.094	130	1,997,322
Depth-3	0.074	150	2,745,332	0.094	150	299,762	0.081	150	568,778
Depth-4	0.076	150	1,031,028	0.095	150	441,714	0.083	150	509,130

Table 2. Structure of the proposed STCNN.

Layer	Name	B	Size
0	Inputs	1	$(N, 18)$
1	Convolution	128	$(3, 3)$
	Pooling	128	$(2, 2)$
2	Convolution	64	$(3, 3)$
	Pooling	64	$(2, 2)$
3	Convolution	32	$(3, 3)$
	Pooling	32	$(2, 2)$
4	Fully-Connected	1	.
5	Output	1	$(N, 6)$

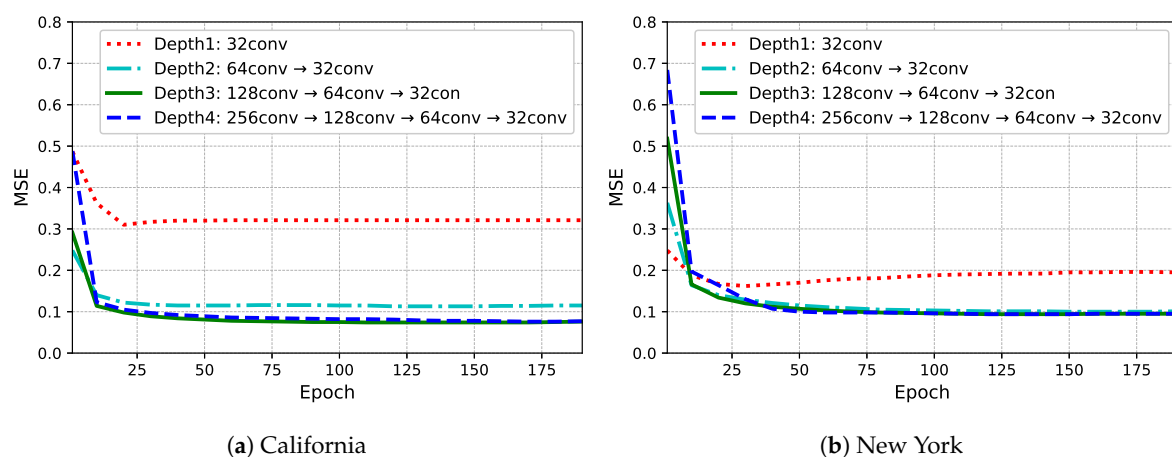
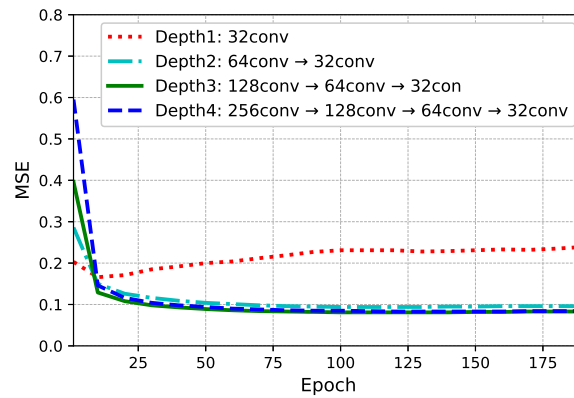


Figure 2. Cont.



(c) Alabama

Figure 2. Validation errors according to the epoch.

3.3. Hyperparameter Selection of the Compared Models

We compare the proposed STCNN model with three well-known methods: AR [10], FNN [11], and LSTM [12]. The AR based model uses a single-layer perceptron so that forecasting output values are linearly dependent on the previous values of all sites [10]. The FNN based model uses a two-layer perceptron network as in [11]. When training greater number of layers, it spends additional time with few differences in MSE of the validation set. The number of hidden nodes is determined to be twice the number of input nodes based on the validation set. The LSTM based model uses a two-layer LSTM network as in [12], where the outputs of the first layer become the inputs of the second layer, and the last LSTM cell is used to generate the predicted output values. The two-layer LSTM can make the second layer capture longer-term dependencies of the input sequence [36]. The number of hidden nodes is determined to be twice the number of output nodes based on the validation set.

4. Performance Evaluation

In order to evaluate the performance of each prediction scheme, three performance metrics are used, the normalized root mean square error (NRMSE), the mean absolute percentage error (MAPE), and the mean absolute scaled error (MASE) defined in [37,38], which are given by

$$\text{NRMSE} = \frac{1}{P_0} \sqrt{\frac{1}{|\mathcal{T}|} \sum_{t=1}^{|\mathcal{T}|} \frac{\|\hat{\mathbf{Y}}_t^{\mathcal{T}} - \mathbf{Y}_t^{\mathcal{T}}\|_F^2}{N \times H}}, \quad (8)$$

$$\text{MAPE} = \frac{1}{P_0} \cdot \frac{1}{|\mathcal{T}|} \sum_{t=1}^{|\mathcal{T}|} \frac{\|\hat{\mathbf{Y}}_t^{\mathcal{T}} - \mathbf{Y}_t^{\mathcal{T}}\|_{1,1}}{N \times H}, \quad (9)$$

$$\text{MASE} = \frac{1}{|\mathcal{T}|} \cdot \frac{\sum_{t=1}^{|\mathcal{T}|} \|\hat{\mathbf{Y}}_t^{\mathcal{T}} - \mathbf{Y}_t^{\mathcal{T}}\|_{1,1}}{\frac{1}{|\mathcal{T}|-1} \sum_{t=2}^{|\mathcal{T}|} \|\mathbf{Y}_t^{\mathcal{T}} - \mathbf{Y}_{t-1}^{\mathcal{T}}\|_{1,1}}, \quad (10)$$

where $\mathcal{T}$ is a test dataset, $\hat{\mathbf{Y}}_t^{\mathcal{T}}$ and $\mathbf{Y}_t^{\mathcal{T}}$ are the t -th predicted and target output space-time matrices in the test dataset, respectively, P_0 is the installed PV generation capacity, and $\|\cdot\|_{1,1}$ is the L_1 norm of the matrix. Note that, if $\mathbf{Y}_t^{\mathcal{T}}$ is already normalized by the installed capacity for better training, which is the case of this paper, P_0 is simply 1.

4.1. PV Generation Profiles

Figure 3 shows the space-time matrix (238 sites and 7 days) of California. Figure 3a,b are for the cases before and after the GAA is applied, respectively. As can be seen in Figure 3b, PV generations become highly correlated after the GAA. Figure 4 shows the normalized PV generation profiles of

238 PV sites in California for seven days; the dashed line shows the mean PV generation, and each colored line represents individual PV generation. The peak of total PV generation is 5897 MW, which roughly corresponds to six nuclear power plants. We see that 238 sites have diverse generation profiles because of cloud cover and cloud movement even in California.

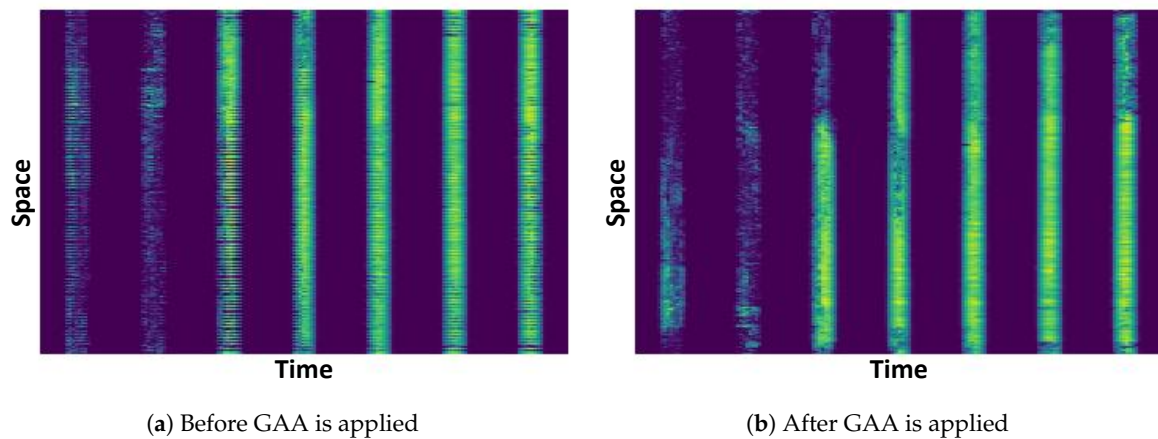


Figure 3. Images of the space-time matrix for seven days (California).

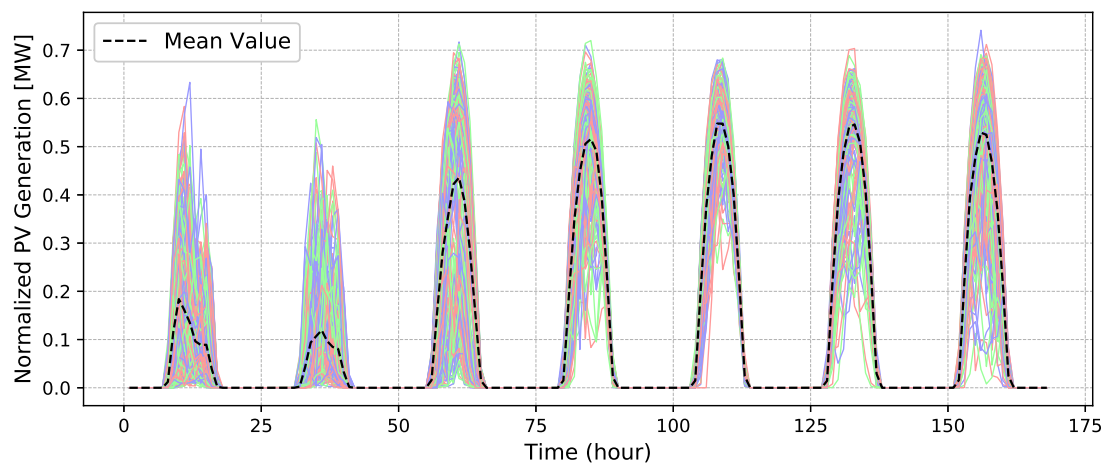


Figure 4. Multi-site PV generation profiles for seven days (238 sites in California).

4.2. Forecasting of Individual PV Sites

Table 3 shows the PV forecasting errors for AR, FNN, LSTM, and the proposed STCNN for all the sites in California, New York, and Alabama. AR shows the worst performance, and FNN improves the performance compared to AR. LSTM is substantially better than AR and FNN, and STCNN outperforms all three models. For example, in the case of MAPE, STCNN achieves 3.98%, 4.02%, and 4.39% in California, New York, and Alabama, respectively. This confirms that the intuition of sorting PV sites using the GAA and extracting spatial features using convolutional network works well.

Table 3. Forecasting error of individual PV site.

	California				New York				Alabama			
	AR	FNN	LSTM	STCNN	AR	FNN	LSTM	STCNN	AR	FNN	LSTM	STCNN
NRMSE [%]	10.59	9.80	8.74	7.98	11.13	10.80	9.91	8.53	11.64	10.02	9.03	8.84
MAPE [%]	5.53	5.15	4.25	3.98	6.38	5.76	5.06	4.02	6.31	5.43	4.98	4.39
MASE	1.16	1.08	0.89	0.84	2.00	1.80	1.58	1.26	1.49	1.29	1.18	1.04

We then evaluate the performance at each forecasting horizon of individual PV sites. Figure 5 shows NRMSE, MAPE, and MASE for every forecasting horizon. All the models show similar performances at 1 h forecasting horizon. However, the performance of AR model starts to degrade sharply after 2 h horizon because forecasting heavily depends on nonlinearity. After 2 h horizon, the accuracy of FNN is better than AR, and LSTM shows better performance than FNN. The proposed STCNN outperforms all the other algorithms after 2 h horizon, and at 6 h horizon, the STCNN shows 4.6%, 4.8%, and 5.3% of MAPE for California, New York and Alabama states, respectively, which are 4–33% improvement compared to the other models. The performance gap between LSTM and STCNN tends to increase in forecasting horizon. Although STCNN sometimes has similar performance with LSTM, e.g., 5 h horizon in California, or 6 h horizon in Alabama, STCNN steadily shows better performance than all the other algorithms.

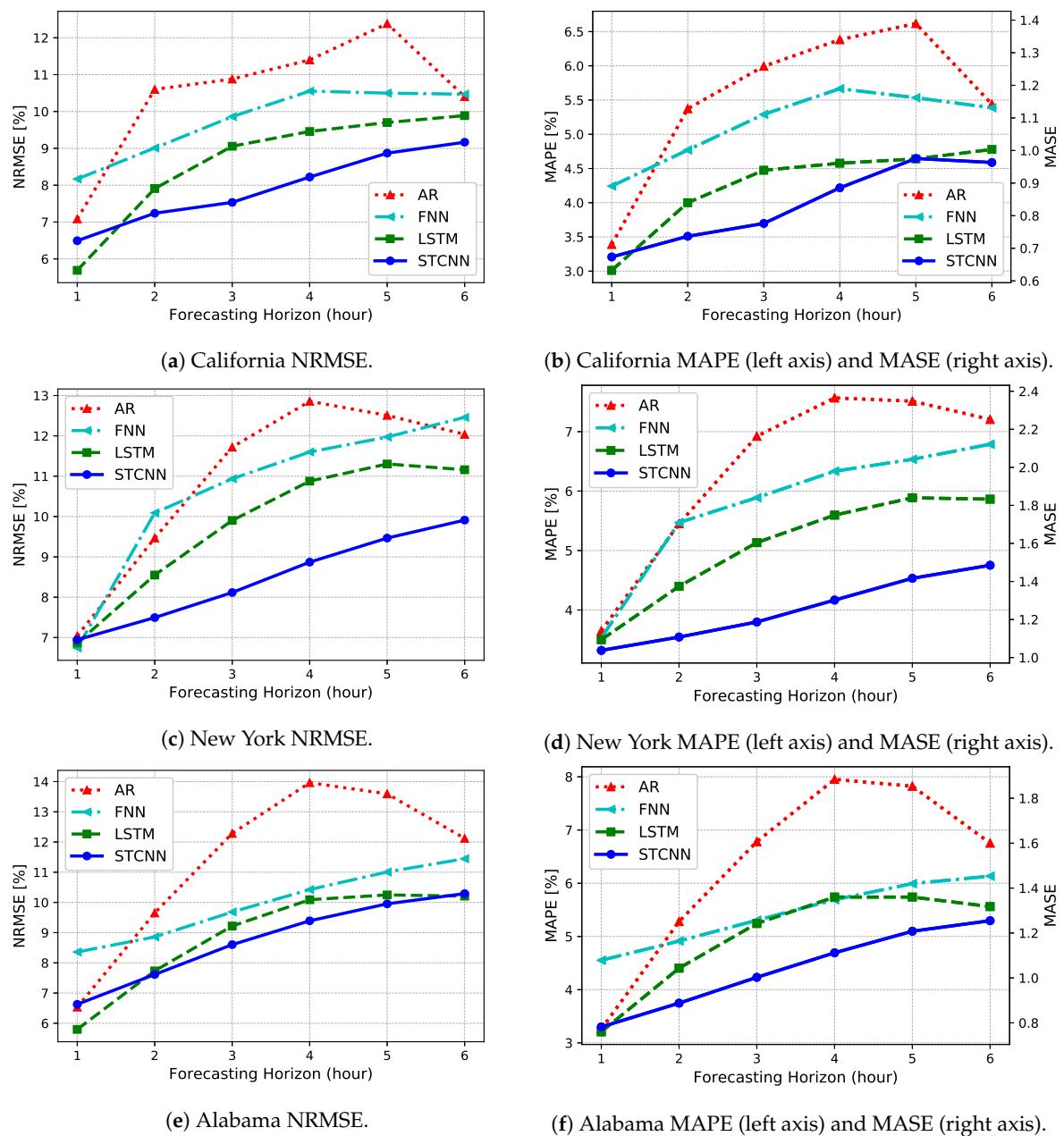


Figure 5. Performance evaluation of individual PV sites.

4.3. Forecasting of Aggregated PV Generation

Next, we investigate the effect of PV sites aggregation, which is the concept of VPP [24]. In this case, the results can be different from the previous cases because errors from one site can be compensated by other sites. Thus, balancing over-forecasting errors and under-forecasting errors is more important in addition to reducing individual PV sites errors. First, we demonstrate the locational dependency of forecasting accuracy to see the ratio of over-forecasting and under-forecasting. Figure 6 shows the mean errors of each PV site for all forecasting horizons, where over-forecasting and under-forecasting errors are set to positive and negative, respectively. The proposed STCNN model shows that the ratio of over-forecasting and under-forecasting is best balanced as well as the forecasting errors is the lowest. By contrast, the other models show that the number of over-forecasted sites is much higher than the number of under-forecasted sites. This can be intuitively interpreted as the effect of exploiting the spatial information, which enables indirect capturing of cloud cover and cloud movements. When tracking the cloud movements is somewhat incorrect, it induces that some regions have more clouds while some other regions have fewer clouds. This results in the balanced over-forecasting and under-forecasting, which can have a high potential of reducing errors when PV sites are aggregated.

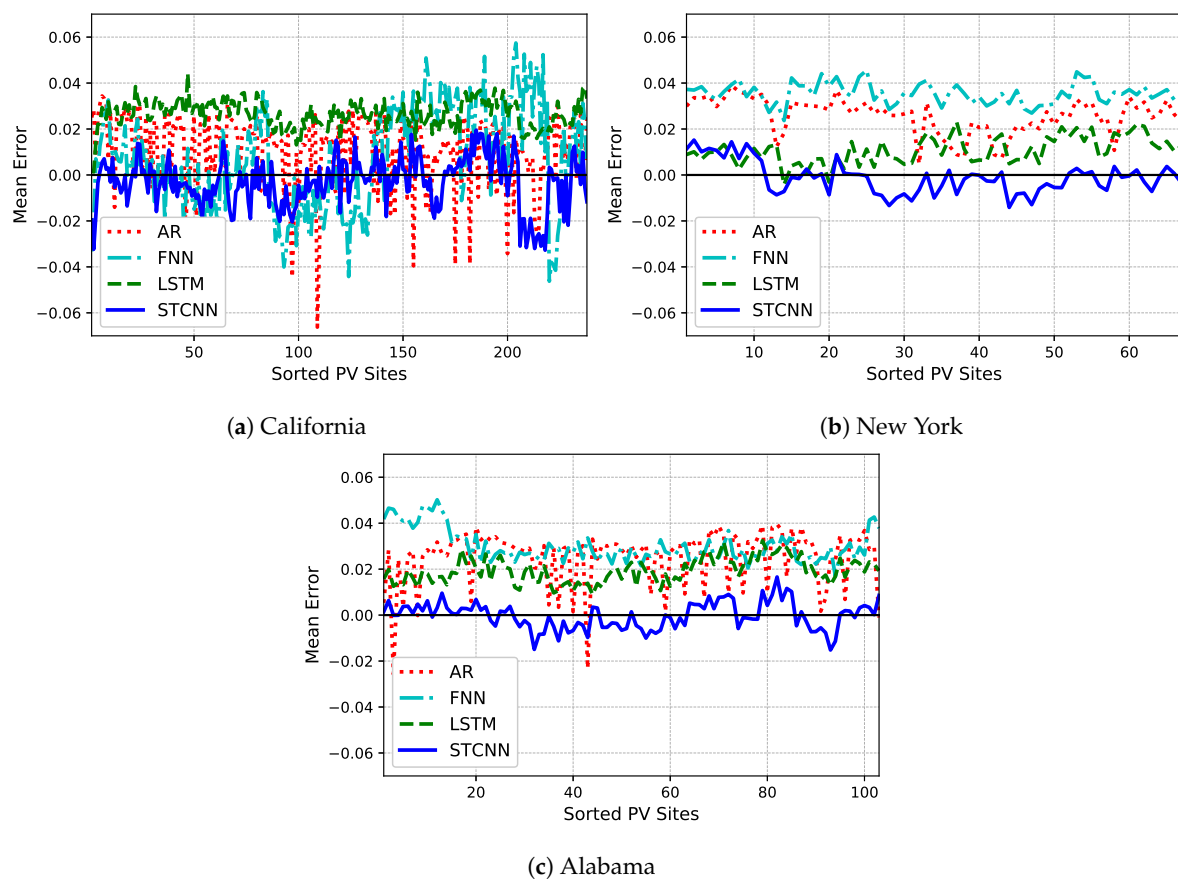


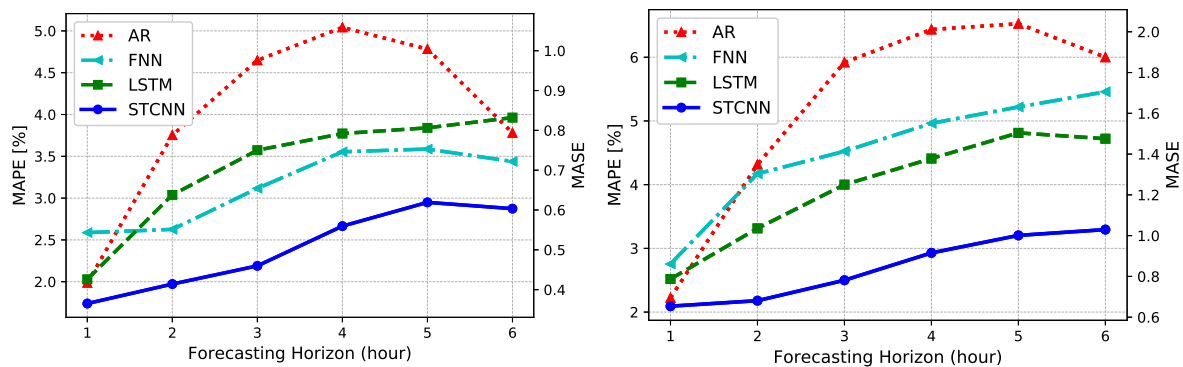
Figure 6. Site-specific evaluation of mean errors.

We then see the performance of aggregated PV sites in terms of MAPE to see the effect of PV aggregation. We normalize the aggregated data between 0 and 1 using the number of PV sites for each state. Table 4 shows the MAPE of PV forecasting for AR, FNN, LSTM, and the proposed STCNN for the aggregated PV sites in California, New York, and Alabama, and Figure 7 shows MAPE and MASE for every forecasting horizon. We see the different performance results from the individual performance. For example, LSTM, the second best model, shows the worst improvement of MAPE after aggregation, and the performance is sometimes even worse than FNN. Nevertheless, as can be seen in Table 4 and Figure 7, the proposed STCNN model outperforms all the other algorithms in

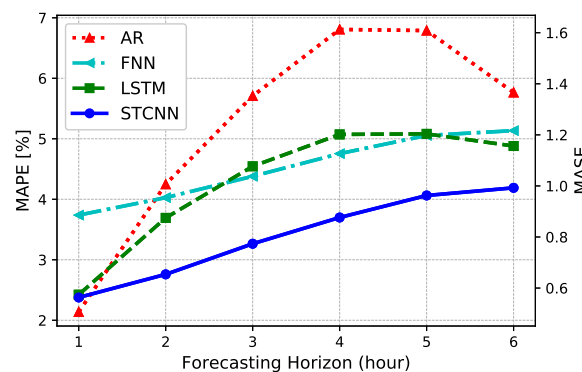
most cases. In addition, the noticeable result is that the proposed STCNN shows the highest MAPE improvement. The performance improvement over the other models becomes much larger than the case of individual PV sites forecasting, which shows the benefit of balanced over-forecasting and under-forecasting.

Table 4. MAPE comparison of individual PV sites and PV aggregation.

	California				New York				Alabama			
	AR	FNN	LSTM	STCNN	AR	FNN	LSTM	STCNN	AR	FNN	LSTM	STCNN
MAPE of individual PV sites [%]	5.53	5.15	4.25	3.98	6.38	5.76	5.06	4.02	6.31	5.43	4.98	4.39
MAPE of PV aggregation [%]	4.00	3.15	3.37	2.40	5.24	4.51	3.96	2.70	5.24	4.51	4.28	3.39
MAPE Improvement [%]	27.67	38.83	20.71	39.70	17.87	21.70	21.74	32.84	16.96	16.94	14.06	22.78



(a) California MAPE (left axis) and MASE (right axis). (b) New York MAPE (left axis) and MASE (right axis).



(c) Alabama MAPE (left axis) and MASE (right axis).

Figure 7. Performance evaluation of PV aggregation.

Figure 8 demonstrates the aggregated PV generation of all sites in New York and its forecasting. Figure 8a shows the result of 2 h horizon. The STCNN shows accurate forecasting by closely following the real PV generation while the forecasted profiles of the other methods deviate from the real generation. Figure 8b compares the predictions for 6 h horizon. Obviously, the forecasting errors become larger as forecasting horizon increases, but we still observe that the performance of STCNN is better than the other models. These results confirm that the proposed STCNN model effectively captures the cloud cover and cloud movement.

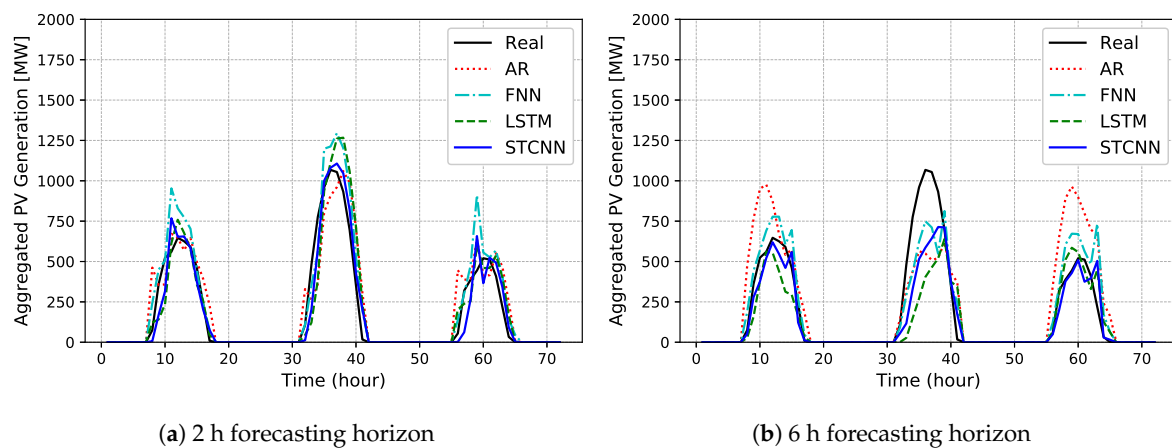


Figure 8. Aggregated PV generation and forecasting (New York).

5. Conclusions

In this paper, we proposed a novel short-term (up to 6 h) spatio-temporal PV forecasting framework. By extracting spatio-temporal features from the multi-site PV datasets, the proposed STCNN indirectly captured the cloud cover and cloud movement without using a complex model structure. In doing this, we proposed the GAA to construct the space-time matrix that retains the spatial and temporal relations of the multi-site PV generations. Then, CNN is applied to learn spatio-temporal relations of multi-site PV generation. Extensive simulations with multiple PV sites in California, New York, and Alabama showed that the proposed STCNN outperforms the other forecasting models based on AR, FNN, and LSTM in terms of NRMSE, MAPE, and MASE. In addition, the proposed STCNN shows the highest error reduction when multiple PV sites are aggregated. We expect that STCNN can be further extended by being combined with other algorithms. For example, the combination of STCNN and LSTM is possible. Since LSTM is generally better than CNN in time-series forecasting, the performance can be improved when the spatial features are extracted by CNN and the temporal features are extracted by LSTM.

Author Contributions: J.J. designed the algorithm, performed the simulations, and prepared the manuscript as the first author. H.K. led the project and research. Both of the authors discussed the simulation results and approved the publication.

Funding: This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT under Grant (NRF-2017R1A1A1A05001377), and by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, industry & Energy (MOTIE) of Korea (20192010107290).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. [\[CrossRef\]](#)
2. Chow, C.W.; Urquhart, B.; Lave, M.; Dominguez, A.; Kleissl, J.; Shields, J.; Washom, B. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol. Energy* **2011**, *85*, 2881–2893. [\[CrossRef\]](#)
3. Marquez, R.; Coimbra, C. Intra-hour DNI forecasting based on cloud tracking image analysis. *Sol. Energy* **2013**, *91*, 327–336. [\[CrossRef\]](#)
4. Jang, H.S.; Bae, K.Y.; Park, H.; Sung, D.K. Solar Power Prediction Based on Satellite Images and Support Vector Machine. *IEEE Trans. Sustain. Energy* **2016**, *7*, 1255–1263. [\[CrossRef\]](#)
5. Perez, R.; Kivalov, S.; Schlemmer, J.; Hemker, K., Jr.; Renné, D.; Hoff, T.E. Validation of short and medium term operational solar radiation forecasts in the US. *Sol. Energy* **2010**, *84*, 2161–2172. [\[CrossRef\]](#)
6. Bae, K.Y.; Jang, H.S.; Sung, D.K. Hourly Solar Irradiance Prediction Based on Support Vector Machine and Its Error Analysis. *IEEE Trans. Power Syst.* **2017**, *32*, 935–945. [\[CrossRef\]](#)

7. Sharma, N.; Sharma, P.; Irwin, D.; Shenoy, P. Predicting solar generation from weather forecasts using machine learning. In Proceedings of the 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium, 17–20 October 2011; pp. 528–533.
8. Yang, C.; Thatte, A.A.; Xie, L. Multitime-Scale Data-Driven Spatio-Temporal Forecast of Photovoltaic Generation. *IEEE Trans. Sustain. Energy* **2015**, *6*, 104–112. [[CrossRef](#)]
9. Agoua, X.G.; Girard, R.; Kariniotakis, G. Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production. *IEEE Trans. Sustain. Energy* **2018**, *9*, 538–546. [[CrossRef](#)]
10. Xu, J.; Yoo, S.; Heiser, J.; Kalb, P. Sensor network based solar forecasting using a local vector autoregressive ridge framework. In Proceedings of the 31st Annual ACM Symposium on Applied Computing, Pisa, Italy, 4–8 April 2016; pp. 2113–2118.
11. Kashyap, Y.; Bansal, A.; Sao, A.K. Spatial approach of artificial neural network for solar radiation forecasting: modeling issues. *J. Sol. Energy* **2015**, *2015*, 410684. [[CrossRef](#)]
12. Ghaderi, A.; Sanandaji, B.M.; Ghaderi, F. Deep Forecast: Deep Learning-based Spatio-Temporal Forecasting. *arXiv* **2017**, arXiv:1707.08110.
13. Lai, G.; Chang, W.C.; Yang, Y.; Liu, H. Modeling long-and short-term temporal patterns with deep neural networks. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 95–104.
14. Lee, W.; Kim, K.; Park, J.; Kim, J.; Kim, Y. Forecasting Solar Power Using Long-Short Term Memory and Convolutional Neural Networks. *IEEE Access* **2018**, *6*, 73068–73080. [[CrossRef](#)]
15. Zhu, Q.; Chen, J.; Zhu, L.; Duan, X.; Liu, Y. Wind speed prediction with spatio-temporal correlation: A deep learning approach. *Energies* **2018**, *11*, 705. [[CrossRef](#)]
16. Khodayar, M.; Mohammadi, S.; Khodayar, M.E.; Wang, J.; Liu, G. Convolutional Graph Autoencoder: A Generative Deep Neural Network for Probabilistic Spatio-temporal Solar Irradiance Forecasting. *IEEE Trans. Sustain. Energy* **2019**. [[CrossRef](#)]
17. Lauret, P.; Voyant, C.; Soubdhan, T.; David, M.; Poggi, P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Sol. Energy* **2015**, *112*, 446–457. [[CrossRef](#)]
18. Zagórski, I.; Kulisz, M. Effect of technological parameters on vibration acceleration in milling and vibration prediction with artificial neural networks. In *MATEC Web of Conferences*; EDP Sciences: Kazimierz Dolny, Poland, 2019; Volume 252, p. 03015.
19. Kluz, R.; Antosz, K.; Trzepieciński, T.; Gola, A. Predicting the Error of a Robot's Positioning Repeatability with Artificial Neural Networks. In *International Symposium on Distributed Computing and Artificial Intelligence*; Springer: Avila, Spain, 2019; pp. 41–48.
20. Parot, A.; Michell, K.; Kristjanpoller, W.D. Using Artificial Neural Networks to forecast Exchange Rate, including VAR-VECM residual analysis and prediction linear combination. *Intell. Syst. Acc. Finance Manag.* **2019**, *26*, 3–15. [[CrossRef](#)]
21. Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y.; Wang, Y. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* **2017**, *17*, 818. [[CrossRef](#)]
22. Yang, G.; Wang, Y.; Yu, H.; Ren, Y.; Xie, J. Short-term traffic state prediction based on the spatiotemporal features of critical road sections. *Sensors* **2018**, *18*, 2287. [[CrossRef](#)]
23. Ke, R.; Li, W.; Cui, Z.; Wang, Y. Two-Stream Multi-Channel Convolutional Neural Network (TM-CNN) for Multi-Lane Traffic Speed Prediction Considering Traffic Volume Impact. *arXiv* **2019**, arXiv:1903.01678.
24. Pudjianto, D.; Ramsay, C.; Strbac, G. Virtual power plant and system integration of distributed energy resources. *IET Renew. Power Gener.* **2007**, *1*, 10–16. [[CrossRef](#)]
25. Fu, W.; Johnston, M.; Zhang, M. Low-Level Feature Extraction for Edge Detection Using Genetic Programming. *IEEE Trans. Cybern.* **2014**, *44*, 1459–1472. [[CrossRef](#)]
26. NREL. Solar Power Data for Integration Studies. Available online: <http://www.nrel.gov/grid/solar-power-data.html> (accessed on 7 October 2019).
27. Ceci, M.; Corizzo, R.; Fumarola, F.; Malerba, D.; Rashkovska, A. Predictive Modeling of PV Energy Production: How to Set Up the Learning Task for a Better Prediction? *IEEE Trans. Ind. Inform.* **2017**, *13*, 956–966. [[CrossRef](#)]
28. Feng, C.; Cui, M.; Hodge, B.; Lu, S.; Hamann, H.; Zhang, J. Unsupervised Clustering-Based Short-Term Solar Forecasting. *IEEE Trans. Sustain. Energy* **2018**, *10*, 2174–2185. [[CrossRef](#)]

29. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Stateline, NV, USA, 3–8 December 2012; pp. 1097–1105.
30. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
31. Ioffe, X.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
32. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; Volume 16, pp. 265–283.
36. Tai, K.S.; Socher, R.; Manning, C.D. Improved semantic representations from tree-structured long short-term memory networks. *arXiv* **2015**, arXiv:1503.00075.
37. Zhang, J.; Florita, A.; Hodge, B.M.; Lu, S.; Hamann, H.F.; Banunarayanan, V.; Brockway, A.M. A suite of metrics for assessing the performance of solar power forecasting. *Sol. Energy* **2015**, *111*, 157–175. [[CrossRef](#)]
38. Sobri, S.; Koohi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).