

Very Short-term Forecasting of Distributed PV Power Using GSTANN

Tiechui Yao, *Student Member, IEEE*, Jue Wang, Yangang Wang, Pei Zhang, *Fellow, IEEE*,
Haizhou Cao, Xuebin Chi, and Min Shi

Abstract—Photovoltaic (PV) power forecasting is essential for secure operation of a power system. Effective prediction of PV power can improve new energy consumption capacity, help power system planning, promote development of smart grids, and ultimately support construction of smart energy cities. However, different from centralized PV power forecasts, three critical challenges are encountered in distributed PV power forecasting: 1) lack of on-site meteorological observation, 2) leveraging extraneous data to enhance forecasting performance, 3) spatial-temporal modelling methods of meteorological information around the distributed PV stations. To address these issues, we propose a Graph Spatial-Temporal Attention Neural Network (GSTANN) to predict the very short-term power of distributed PV. First, we use satellite remote sensing data covering a specific geographical area to supplement meteorological information for all PV stations. Then, we apply the graph convolution block to model the non-Euclidean local and global spatial dependence and design an attention mechanism to simultaneously derive temporal and spatial correlations. Subsequently, we propose a data fusion module to solve the time misalignment between satellite remote sensing data and surrounding measured on-site data and design a power approximation block to map the conversion from solar irradiance to PV power. Experiments conducted with real-world case study datasets demonstrate that the prediction performance of GSTANN outperforms five state-of-the-art baselines.

Index Terms—Distributed photovoltaic power forecasting, graph convolutional networks, satellite images, spatial-temporal attention.

NOMENCLATURE

A. Abbreviations

PV	Photovoltaic.
CPV	Centralized photovoltaic.
DPV	Distributed photovoltaic.
ARIMA	Auto-regressive integrated moving average.
SVR	Support vector regression.

LSTM	Long-short-term memory.
RNN	Recurrent neural network.
NWP	Numerical weather forecast.
SWR	Shortwave radiation.
GCN	Graph convolutional network.
GSTANN	Graph spatial-temporal attention neural network.
CTSA	Combined temporal and spatial attention.
PVOD	Photovoltaic output dataset.
MAE	Mean absolute error.
RMSE	Root mean square error.
MAPE	Mean absolute percentage error.
ST-GCN	Spatial-temporal graph convolutional network.
GMAN	Graph multi-attention network.

B. Variables

T_h	The historical time steps of the timeline.
T_l	The lag time steps of the timeline.
T_f	The future time steps of the timeline.
$X_{T_h}^{SWR}$	The theoretical inferred irradiance from short-wave radiation points.
$X_{T_*}^{CPV}$	The measurements from centralized photovoltaic stations.
$\hat{Y}_{T_f}$	The predicted power of distributed photovoltaic stations.
Y_{T_f}	The ground truth power of distributed photovoltaic stations.
N^S	The number of shortwave radiation points.
N^D	The number of distributed photovoltaic stations.
N^C	The number of centralized photovoltaic stations.
$I_{T_f}^{CPV}$	The ground truth total solar irradiance of centralized photovoltaic stations.
G	The unweighted graph covering shortwave radiation, centralized photovoltaic, and distributed photovoltaic stations.
V	The vertex number of the graph.
A	The adjacency matrix of the graph.
$\hat{I}_{T_f}$	The predicted theoretical inferred irradiance of DPV stations.
$HC^{(*)}$	The input or hidden output of combined temporal and spatial attention.
$H^{(*)}$	The input or hidden output of transform attention.

Manuscript received January 5, 2022; revised April 28, 2022; accepted May 27, 2022. Date of online publication October 12, 2022; date of current version October 18, 2023. This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences (No.XDA27000000).

T. C. Yao, J. Wang (corresponding author, email: wangjue@sccas.cn), Y. G. Wang, H. Z. Cao, and X. B. Chi are with Computer Network Information Center of the Chinese Academy of Sciences, Beijing 100190, China; and also with University of Chinese Academy of Sciences, Beijing 100190, China.

P. Zhang is with School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China.

M. Shi is with State Grid Hebei Electric Power Co., Ltd., Shijiazhuang 050000, China.

DOI: 10.17775/CSEEJPES.2022.00110

I. INTRODUCTION

GLOBAL economic development and population growth have led to an increasing demand for global energy and electricity consumption. Meanwhile, given current high carbon emissions, environmental damage, and limited fossil fuels, development of renewable power technologies and the ability to accurately predict photovoltaic (PV) power are imperative to development of smart energy cities [1]–[3]. Solar energy can be converted to electric energy through the PV effect, which is a promising renewable power technique and has broad space in development [4], [5]. However, the inherent volatility and variability of meteorological factors can threaten safe operation and economic viability of power systems. Since solar PV power is weather dependent and stochastic in nature, accurate PV power forecasting remains a challenge and has received much attention from researchers. Compared with centralized PV (CPV), several differences of distributed PV (DPV) are as follows: 1) DPV systems are mostly installed on the surface of buildings and are the mainstream of grid-connected PV power generation in developed countries; 2) The power fluctuation of DPV is stronger than of CPV, which makes the spatial-temporal correlation of power of DPV more difficult to capture; 3) DPV stations are not equipped with multi-sensor local weather monitoring systems like CPV stations [6]. In China, the PV industry is gradually developing from CPV stations to DPV stations [7], [8].

From the perspective of forecasting technology, there are two main approaches for PV forecasting, physical method and statistical method [9]. The first method uses complex analytical equations to model the PV system based on factors related to power production, but its application is limited to the complexities of real-world power stations [10]. Contrarily, the second focuses on directly predicting PV power by mining the inherent laws of different data sources by using statistical models, machine learning models, deep learning models, etc. With development of big data and artificial intelligence, the latter method is widely used in PV power forecasting [11], [12]. Despite this, few researchers focus on very short-term DPV power forecasting using novel deep learning methods.

Traditional statistical methods such as auto-regressive moving average (ARMA) [13] and auto-regressive integrated moving average (ARIMA) variant [14] are applied to PV power forecasting using historical power data. However, their prediction errors are larger than machine learning or deep learning models because traditional models are flawed in supporting multivariate forecasting. In terms of machine learning models, reference [15] adopts genetic algorithm-based support vector regression (SVR) for short-term PV power forecasting. Multivariate meteorological data and historical solar irradiance are taken as input to predict PV power using K-means and grey relational analysis in [16]. Although SVR and K-means have the ability to process historical weather data to predict PV power, i.e. multiple regression, it is insufficient in extracting the temporal relationship between historical data. Studying future trends in in-situ meteorological data, such as irradiance, can help analyze the range of power output. Considering temporal correlations, literature [17] predicted the day-ahead

PV power based on a deep learning model, namely long-short-term memory [18], [19] recurrent neural network (LSTM-RNN) network, and on-site measurements of meteorological variables, e.g., wind speed, relative humidity, temperature, and global horizontal radiation. Nevertheless, given the lack of meteorological indicator measurement sensors in DPV stations, such as irradiance meters, the meteorological information available in the surroundings needs to be incorporated for DPV power prediction. Therefore, most of the above models cannot fully meet the requirements of fusing multi-source data.

Combining surrounding meteorological information to predict power poses a challenge for spatial correlation modelling. Literature [20] proposes a geo-sensory multi-level attention network for short-term PV forecasting based on the numerical weather forecast (NWP) and clearness index. This work adopts grey relational analysis to select nearby power stations but uses a separate attention mechanism model. The temporal attention mechanism [21] perceives the temporal correlation of data within a single CPV station, and the spatial attention mechanism only pays attention to the power information between different CPV stations. Based on a graph neural network, combined with LSTM and encoder-decoder architecture respectively, literature [22] proposes two deep learning models and uses NWP data for CPV power prediction. Despite previous studies having achieved good performance in CPV power forecasting, lack of optimized attention mechanisms and different data sources both limit their application in very short-term DPV power forecasting.

On the other hand, the available data sources for PV power forecasting are broadly divided into the following three aspects: local measurements data, NWP, satellite cloud images and others. First, local measurement data is not available in distributed PV plants due to the lack of a weather monitoring system. Second, NWP data is only applicable in short-term PV power forecasts [23], i.e. forecast steps longer than 4 hours. Last but not least, coarse-grained historical meteorological information can be obtained using satellite cloud images covering the range of the power station. Meanwhile, fine-grained real-time weather information can be accessed through CPV stations with monitoring systems in the surrounding area. Combining the above coarse and fine-grained data sources to jointly predict very short-term DPV power is considered a feasible solution. Literature [24] and [25] used shortwave radiation (SWR) data from Himawari-8 to predict short-term and ultra-short-term CPV power, respectively. Notably, the former is a regional power forecast and the latter uses locally measured data and additional NWP data. Literature [26] applies binary genetic algorithm and SVR for the tedious feature selection and then predicts short-term DPV power. Literature [27] utilizes the genetic algorithm and bidirectional LSTM model to predict short-term DPV power based on surrounding power stations but lacks the assistance of satellite cloud images.

However, deep learning methods have been shown to be superior to traditional machine learning models in their ability to learn from massive amounts of data and do not require complex feature engineering. High efficiency and flexibility of the attention mechanism in modelling multiple depen-

dencies make it preferable for PV power forecasting [28], [29], but application of attention mechanisms in DPV power forecasting still needs further modification. Subsequently, the graph convolutional network (GCN) can effectively extract representation and spatial features in non-Euclidean graph structures, e.g., PV station site [30]–[33]. More novel deep learning models (e.g., GraphSage [34], GAT [35], GIN [36], and DeeperGCN [37]) also focus on graph embedding and representation, but their scalability on DPV power forecasting remains to be studied. Consequently, incorporating external data and novel spatial-temporal modelling methods for very short-term DPV power forecasting still need to be explored.

In conclusion, to forecast very short-term DPV power based on multi-source data, e.g., satellite remote sensing data, surrounding on-site meteorological data, three major challenges are encountered:

- *Lack of in-situ meteorological measurements.* In-situ measurements are generally considered more reliable than numerical weather prediction and satellite remote sensing data [38]–[40]. However, local meteorological measurements are unavailable for DPV stations due to construction costs. Therefore, the first challenge lies in integration of multi-source and valid data for DPV power forecasting.
- *Time misalignment of multi-source data.* Typically, there is a 30-min delay in obtaining satellite data caused by transmission and preprocessing in real-world applications [41]. Conversely, on-site measured data of CPV stations are obtained in real-time. The second challenge is resolving the time lag between satellite and on-site measured data from CPV stations.
- *Spatial-temporal correlation modelling.* Meteorological measurements may have inherently complex time correlations. Additionally, the spatial correlation between regional satellite cloud images and surrounding power plants can affect the weather conditions of DPV stations [42]. Hence, modelling the dynamic spatial-temporal correlation of each station represents an additional challenge.

To address the aforementioned issues, we propose a Graph Spatial- Temporal Attention Neural Network (GSTANN) and employ satellite meteorological data and surrounding measurements of CPV stations to predict DPV power. Our contributions can be summarized as follows:

- We propose a novel and flexible network, GSTANN, for very short-term DPV power forecasting. This network applies an encoder-decoder architecture, which can leverage satellite remote sensing data and real-time solar irradiance measured by CPV stations for DPV power forecasting.
- We design data fusion and power approximation modules in this network. A data fusion block is designed to solve the time misalignment issue between the satellite-extracted SWR and measured solar irradiance from CPV stations. This block considerably improves the accuracy of DPV power forecasting. PV power approximation is designed to learn the conversion process from solar irradiance to PV power.

- We propose a spatial-temporal modelling approach for DPV power forecasting. Specifically, a GCN is adopted for extracting spatial features by modelling the SWR and various stations into an overall graph. Next, we develop a combined temporal and spatial attention (CTSA) to simultaneously perceive the connection between the spatial and temporal domains throughout the entire architecture.
- We evaluate the GSTANN on a real-world dataset, namely photovoltaic output dataset (PVOD), and compare a variety of classic models to verify the advantages of GSTANN in very short-term DPV power prediction. Further, we conduct extensive ablation studies.

The rest of this paper is summarized as follows: Section II presents notation for timelines, time series, spatial graphs, and problem definitions. Section III introduces details of each module in the GSTANN. Section IV discusses details of the dataset and results of the experiments. Finally, Section V presents the conclusions.

II. PRELIMINARIES AND PROBLEM DEFINITION

In this section, we define some notations and formally define our problem.

Definition 1 (Timeline). As illustrated in Fig. 1, a timeline is introduced. The time step $J+L$ stands for the current moment and $J+L+1$ to $J+K$ ($K \geq L+1$) corresponds to future time steps. Given there is a delay of L time step for the SWR points, the time lag is defined as T_l . Similarly, the historical time steps are defined as T_h , and the future time steps are defined as T_f . Thus, for the sake of simplicity, $T_{\text{seq}} = \{T_h, T_l, T_f\} = \{t_1, \dots, t_J, t_{J+1}, \dots, t_{J+L}, t_{J+L+1}, \dots, t_{J+K}\}$.

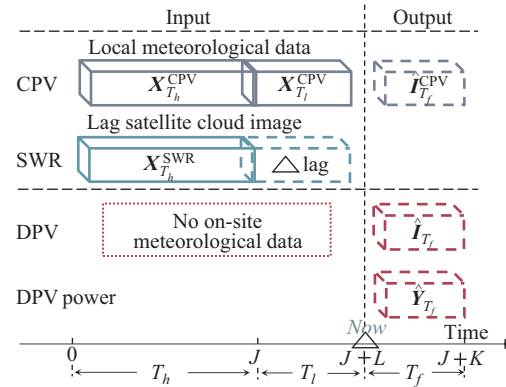


Fig. 1. Schematic of the definition and timeline.

Definition 2 (Time Series). Given the values of historical T_h time steps of the theoretical inferred irradiance from SWR points $X_{T_h}^{\text{SWR}} \in \mathbf{R}^{N^D \times M \times J}$ in addition to the measurements from CPV stations $X_{T_h, T_l}^{\text{CPV}} \in \mathbf{R}^{N^C \times M \times (J+L)}$, we aim to predict the DPV power of the next T_f time steps, denoted by predicted power $\hat{Y}_{T_f} \in \mathbf{R}^{N^D \times (K-L)}$, where N^S , N^D , and N^C represent the number of SWR points, DPV stations, and CPV stations, respectively, and M is the dimension of features such as solar irradiance and temperature. Specifically, we use $Y_{T_f} \in \mathbf{R}^{N^D \times (K-L)}$ to denote the real DPV power and use

$\mathbf{I}_{T_f}^{\text{CPV}} \in \mathbf{R}^{N^C \times (K-L)}$ to represent the label of CPV solar irradiance.

Definition 3 (Spatial Graph). Combined the N^S SWR points with N^D DPV stations and N^C CPV stations, the spatial graph is denoted by an unweighted graph $\mathbf{G} = (\mathbf{V}, \mathbf{A})$, where $\mathbf{V}$ is a set of $N = |\mathbf{V}| = N^S + N^D + N^C$ nodes. The adjacency matrix $\mathbf{A} \in \mathbf{R}^{N \times N}$ contains the connected state of all nodes.

Problem Definition: In summary, the goal is to predict DPV power $\hat{\mathbf{Y}}_{T_f}$ based on the delayed SWR points $\mathbf{X}_{T_h}^{\text{SWR}}$ and real-time surrounding CPV measurements $\mathbf{X}_{T_h, T_l}^{\text{CPV}}$. We formulate this target as the following:

$$f(\mathbf{X}_{T_h}^{\text{SWR}}; \mathbf{X}_{T_h, T_l}^{\text{CPV}}) \rightarrow \hat{\mathbf{Y}}_{T_f} \quad (1)$$

where $f(\cdot)$ is a mapping function we aim to train and learn.

III. GRAPH SPATIAL-TEMPORAL ATTENTION NEURAL NETWORK

In this section, we first introduce the architecture of the proposed GSTANN and then elaborate on each component from top to bottom.

A. Network Overview

On the top left of Fig. 2 lies the input data block, including the integrated spatial graph $\mathbf{G}$, theoretical inferred irradiance $\mathbf{X}_{T_h}^{\text{SWR}}$, and measured solar irradiance $\mathbf{X}_{T_h, T_l}^{\text{CPV}}$. On the bottom left is the output data and label, including DPV power $\hat{\mathbf{Y}}_{T_f}$ and its label $\mathbf{Y}_{T_f}$, CPV measured solar irradiance $\hat{\mathbf{I}}_{T_f}^{\text{CPV}}$

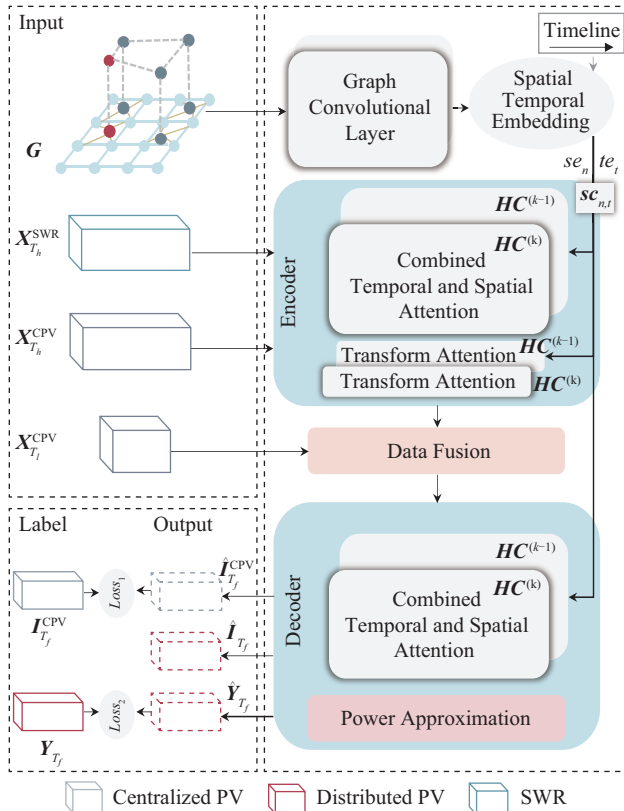


Fig. 2. GSTANN: Graph spatial-temporal attention neural network.

and its label $\mathbf{I}_{T_f}^{\text{CPV}}$, and predicted DPV theoretical inferred irradiance $\hat{\mathbf{I}}_{T_f}$.

On the right, we show the spatial-temporal dependency modelling with GCN and spatial-temporal embedding, an encoder-decoder architecture with data fusion, attention mechanism, and power approximation. Here, GCN extracts the spatial features from the spatial graph, and its output is embedded with the global timeline. The encoder is responsible for the acquisition of the spatial-temporal correlations in historical time T_h . Transform attention encodes the relationship between historical $\{T_h, T_l\}$ and future T_f time steps. The data fusion block integrates the $\mathbf{X}_{T_l}^{\text{CPV}}$ while considering the delayed $\mathbf{X}_{T_h}^{\text{SWR}}$. The decoder block models the global time T_{seq} and spatial features to predict solar irradiance, and the power approximation block completes the mapping of modelling from solar irradiance to DPV power in a future time T_f .

B. Spatial Dependency Modelling for Spatial Graph and Temporal Embedding

PV power output is influenced by local meteorological conditions and correlated to other stations in the power grid. Therefore, we employ GCN and spatial-temporal embedding to model spatial-temporal dependency.

Graph Convolutional Layer: Considering the effects of geographical factors on the graph and the suitability of GCN for modelling non-Euclidean sub-region correlations, we apply GCN to extract the geometrical and spatial features. We take $\mathbf{X}_g$ as the GCN input, and then the two-layer GCN can be expressed as:

$$g(\mathbf{X}_g, \mathbf{A}) = \sigma(\hat{\mathbf{A}}(\text{ReLU}(\hat{\mathbf{A}}\mathbf{X}_g\mathbf{W}_0))\mathbf{W}_1) \quad (2)$$

where $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$ denotes the pre-processing step, the degree matrix $\tilde{\mathbf{D}} = \sum_j \tilde{\mathbf{A}}_{ij}$, $\mathbf{W}_0$ and $\mathbf{W}_1$ represent the weight matrices in the first and second layers, respectively, and $\sigma(\cdot)$ and $\text{ReLU}(\cdot)$ represent the *sigmoid* and *rectified linear unit* activation functions, respectively. We also add the identity matrix $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ to generate a matrix with a self-connection structure, where $\mathbf{I}_N$ is an N -dimensional identity matrix. $\mathbf{A}$ is calculated based on the integrated power grid $\mathbf{G}$, and $a_{ij} \in \mathbf{A}$ is formulated as:

$$a_{ij} = \begin{cases} 1, & \text{distance}_{ij} < \kappa; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

where distance_{ij} is the Euclidean distance coefficient between graph nodes i and j , and κ is the distance threshold.

To cover the spatial features of the entire graph, we elaborate on input $\mathbf{X}_g$ with the great-circle distance. Given the haversine formula [43] determines the great-circle distance between two points on a sphere for a given set of longitudes and latitudes, feature matrix $\mathbf{X}_g \in \mathbf{R}^{N \times N}$ is defined to represent the great-circle distances of all the nodes to supplement the macroscale spatial feature of $\mathbf{A}$. (4) explicitly solves for the distance $x_{ij}^g \in \mathbf{X}_g$.

$$x_{ij}^g = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\varphi}{2} \right) + \cos(\varphi_i) \cos(\varphi_j) \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right) \quad (4)$$

where φ_i and φ_j are the latitudes of point i and point j (in radians), respectively, λ_i and λ_j are the longitudes of point i and point j (in radians), respectively, $\Delta\lambda = \lambda_j - \lambda_i$, $\Delta\varphi = \varphi_j - \varphi_i$, and r is the radius of the Earth in km.

Spatial Temporal Embedding: To transmit the global spatial-temporal representation to the following modules and improve the representation ability of GSTANN [44], we apply spatial-temporal embedding [45] to integrate these global spatial-temporal correlations, including spatial embedding of graph nodes and temporal embedding of time-of-day. Spatial embedding involves storage of graph structure by embedding node $n \in \mathbf{V}$ and transformation of spatial embedding to the D dimension vector, which is represented as $se_n \in \mathbf{R}^D$. Meanwhile, temporal embedding involves encoding each time step and mapping it to $\mathbf{R}^D$, where temporal embedding of a time step $t \in \mathbf{T}_{\text{seq}}$ is represented as $te_t \in \mathbf{R}^D$. Then, this skip-connected embedding of node n at time step t is represented by $sc_{n,t} = (se_n + te_t) \in \mathbf{R}^{N \times D \times (J+K)}$.

C. Spatial-temporal Dependency Modelling for Spatial Graph and Measurements

Combined Temporal and Spatial Attention: PV power of a station is typically correlative with several nearby nodes and closely related to time of day. For example, toward evening, power capacity of PV stations in the area gradually decreases. In other words, historical PV power and data of the surrounding stations or SWR points will affect current power of a DPV station. Moreover, this effect varies gradually over time. To model these spatial-temporal properties, both the encoder and decoder blocks are stacked by multiple CTSA to synchronously obtain the temporal and spatial correlation. However, the difference from the decoder is the encoder block combines the $\mathbf{T}_h$ data of SWR and CPV stations.

We assume the input of the k -th CTSA layer is $\mathbf{HC}^{(k-1)}$, and the hidden state result of point $n \in \mathbf{V}$ at time step $t \in \mathbf{T}_{\text{seq}}$ is represented by $hc_{n,t}^{(k-1)}$ as shown in Fig. 3. Output of the first layer is $\mathbf{HC}^{(k)}$, and the hidden state output of node n in layer l at time step t is given by:

$$hc_{n,t}^{(k)} = \sum_{t' \leq t} \sum_{n' \in \mathbf{V}} \alpha_{(n,t),(n',t')} \cdot hc_{n',t'}^{(k-1)} \quad (5)$$

$$\text{s.t.} \quad \sum_{t' \leq t} \sum_{n' \in \mathbf{V}} \alpha_{(n,t),(n',t')} = 1$$

where $\alpha_{(n,t),(n',t')}$ is the attention score of node n at time step t to node $n' \in \mathbf{V}$ at time step $t' \in \mathbf{T}_{\text{seq}}$. The attention scores are calculated by normalizing $s_{(n,t),(n',t')}$ with *softmax*.

The hidden state is concatenated to calculate the correlation between (n,t) and (n',t') . We formulate correlation $s_{(n,t),(n',t')}$ as:

$$s_{(n,t),(n',t')} = \frac{\langle hc_{n,t}^{(k-1)} \| sc_{n,t}, hc_{n',t'}^{(k-1)} \| sc_{n',t'} \rangle}{\sqrt{2D}} \quad (6)$$

$$\alpha_{(n,t),(n',t')} = \frac{\exp(s_{(n,t),(n',t')})}{\sum_{t^* \leq t} \sum_{n^* \in \mathbf{V}} \exp(s_{(n,t),(n^*,t^*)})} \quad (7)$$

where $\|$ represents the concatenation operation, $\langle \cdot, \cdot \rangle$ represents the inner product, and $2D$ represents the dimension of $hc_{n,t}^{(k-1)} \| sc_{n,t}$.

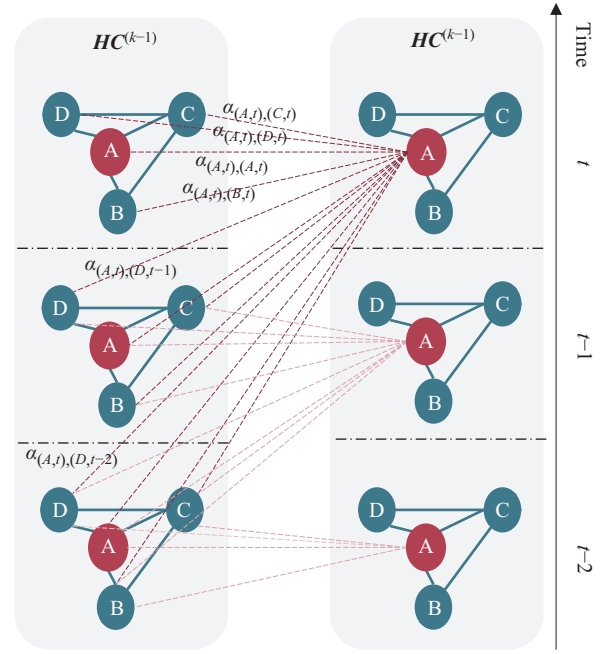


Fig. 3. Schematic of combined temporal and spatial attention.

Substituting the attention score α obtained by (7) into (5), we obtain the output of the k -th layer $hc_{n,t}^{(k)}$. Then, all $hc_{n,t}^{(k)}$ are concatenated to obtain $\mathbf{HC}^k$. When $k = 0$, $\mathbf{HC}^0$ is mapped to the result of $\mathbf{R}^{N \times D \times J}$ through two fully connected layers.

Transform Attention: To model the relationship between $\mathbf{T}_l$ and $\mathbf{T}_f$, transform attention is added before data fusion. For SWR data before the decoder, the future time of SWR is $\{\mathbf{T}_l, \mathbf{T}_f\}$ and the historical time of SWR is $\mathbf{T}_h$. Similar to CTSA, the correlation is calculated with the inner product of the skip-connected embedding $sc_{n,t_{lf}}$ and sc_{n,t_h} ($t_{lf} \in \{\mathbf{T}_l, \mathbf{T}_f\}$, $t_h \in \mathbf{T}_h$). Then the attention score is the result of correlation normalized by *softmax*. We assume the output of the first layer is $\mathbf{H}_{\mathbf{T}_l, \mathbf{T}_f}^{(k)}$, and $\mathbf{H}_{\mathbf{T}_l, \mathbf{T}_f}^{(k)}$ is calculated by the sum of the corresponding time step hidden state results in the $(k-1)$ -th layer with attention scores.

D. Time Misalignment Modelling and Power Approximation

Data Fusion: The features of the nodes corresponding to $\mathbf{X}_{\mathbf{T}_h}^{\text{SWR}}$ and $\mathbf{X}_{\mathbf{T}_h}^{\text{CPV}}$ before time step J are available. To use data closer to the current moment $J + L$ as soon as possible and obtain a more accurate prediction result, the measured solar irradiance $\mathbf{X}_{\mathbf{T}_l}^{\text{CPV}}$ is added to the data fusion.

As shown in Fig. 4, we combine $\mathbf{X}_{\mathbf{T}_l}^{\text{CPV}}$ with the output feature of transform attentions $\mathbf{H}_{\mathbf{T}_l}^{(k)}$ and $\mathbf{H}_{\mathbf{T}_f}^{(k)}$, which provides prior knowledge to the decoder block and plays an important role in enhancing the prediction performance. $\mathbf{X}_{\mathbf{T}_l}^{\text{CPV}}$ is mapped to $\mathbf{R}^{N \times D \times L}$ through two fully connected networks $f_{c_0}(\cdot)$ and is assigned a weight $(1 - \gamma)$, while $\mathbf{H}_{\mathbf{T}_l}^{(k)}$ is assigned a weight of γ . The weighted sum result and $\mathbf{H}_{\mathbf{T}_f}^{(k)}$ are concatenated as the CTSA's input $\mathbf{HC}^{\text{in}}$ in the decoder block, as given in (8).

$$\mathbf{HC}^{\text{in}} = (\gamma \odot \mathbf{H}_{\mathbf{T}_l}^{(k)} + (1 - \gamma) \odot f_{c_0}(\mathbf{X}_{\mathbf{T}_l}^{\text{CPV}})) \| \mathbf{H}_{\mathbf{T}_f}^{(k)} \quad (8)$$

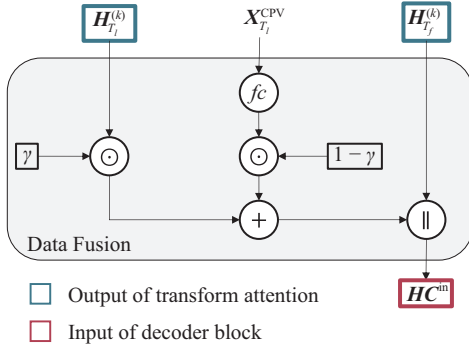


Fig. 4. Schematic of data fusion.

where $\odot$ denotes the element-wise product.

To perform data fusion with higher robustness, γ must be adjusted adaptively. Weight γ goes through a self-adaptation attention, and is subsequently normalized to $(0,1)$ by a σ function, as shown in (9).

$$\gamma = \sigma \left(H_{T_l}^{(k)} \mathbf{W}_{1\gamma} + f_{c1}(X_{T_l}^{CPV}) \mathbf{W}_{2\gamma} + \mathbf{b}_\gamma \right) \quad (9)$$

where $\mathbf{W}_{1\gamma}$, $\mathbf{W}_{2\gamma}$, and $\mathbf{b}_\gamma$ are the learnable parameters, trained by the optimization algorithms to realize self-adaptation. During the training process, adjustment of γ leads to a different preference between $H_{T_l}^{(k)}$ and $X_{T_l}^{CPV}$.

The decoder block re-encodes the features after fusing $X_{T_l}^{CPV}$ to predict solar irradiance of the PV stations. As in the case of the encoder block, this block is also stacked by multi-CTSA. Notably, following both the supervised learning and semi-supervised learning paradigm, all the decoder block outputs $\hat{\mathbf{Y}}_{T_f}$, $\hat{\mathbf{I}}_{T_f}$, and $\hat{\mathbf{I}}_{T_f}^{CPV}$, are used for input in power approximation.

Power Approximation: Given the inherent characteristics of PV power, power is directly proportional to global solar irradiance [46]. To accurately predict PV power, we propose a power approximation block that consists of a convolutional neural layer and two fully connected layers for learning a mapping approximation model from $\hat{\mathbf{I}}_{T_f}$ to $\hat{\mathbf{Y}}_{T_f}$.

$$\hat{\mathbf{Y}}_{T_f} = \sigma(f_{c2}(\mathbf{W}_p \hat{\mathbf{I}}_{T_f} + \mathbf{b}_p)) \quad (10)$$

where $\mathbf{W}_p$ and $\mathbf{b}_p$ are the learnable parameters.

In contrast to no-label $\hat{\mathbf{I}}_{T_f}$, both $\hat{\mathbf{Y}}_{T_f}$ and $\hat{\mathbf{I}}_{T_f}^{CPV}$ have labels. We design a multi-target loss function expressed by (11), (12), and (13).

$$\text{loss}_1 = \frac{1}{|\mathbf{T}_f|} \sum_{t \in \mathbf{T}_f} \left\| \mathbf{I}_t^{CPV} - \hat{\mathbf{I}}_t^{CPV} \right\|_2 \quad (11)$$

$$\text{loss}_2 = \frac{1}{|\mathbf{T}_f|} \sum_{t \in \mathbf{T}_f} \left\| \mathbf{Y}_t - \hat{\mathbf{Y}}_t \right\|_2 \quad (12)$$

Hyperparameters η_1 and η_2 are used to control the impact of CPV irradiance and DPV power on the loss function (13).

$$\text{loss} = \eta_1 \text{loss}_1 + \eta_2 \text{loss}_2 \quad (13)$$

IV. EXPERIMENTS

In this section, we first introduce datasets and then expound

on the data preprocessing. Further, we discuss the implementation and training model. Finally, we present the following experimental process and analyze the results.

A. Datasets

As shown in Fig. 5, the PVOD dataset integrates PV station records such as PV power, solar irradiance of the power grid (Hebei PV stations), and the preprocessed SWR from Himawari-8.

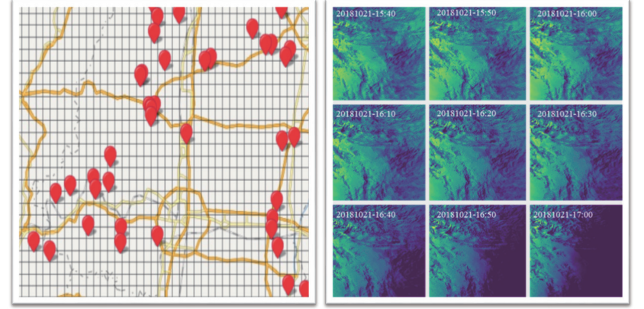


Fig. 5. Subset of PVOD (left) and satellite remote sensing data (right).

Hebei PV stations: We collect data from 6 PV stations at different latitudes and longitudes in Hebei, China. Each CPV station exhibits a 15-min interval in power and meteorological data measurements such as global irradiance, diffuse irradiance, temperature, pressure, wind direction, and wind speed, whereas only the value of power is provided for each DPV station. Both datasets range from July 1, 2018, to December 31, 2018, and available data of each day are acquired from 8:00 am to 5:00 pm (Coordinated Universal Time + 0800). We arrange the datasets in chronological order, and then data from each PV station for every 5 days were used for the test (20%); the rest were used for training (80%). This ensures the test set is evenly distributed and avoids training in summer and testing in winter. The first release of the measurements data can now be accessed through [47].

Shortwave Radiation: Himawari-8 is a geosynchronous satellite, launched by the Japan Meteorological Agency in 2015, which provides reports on weather forecasts, typhoons, and rainstorms around Japan, East Asia, and the Western Pacific. The detection range of Himawari-8 is 60° S–60° N latitude, 80° E–160° W longitude, and spatial resolution is 0.05°, which corresponds to approximately 5 km on the ground. All detected data are 2401 × 2401 matrixes. As one of the geophysical parameter data produced by the JAXA P-Tree system using the Himawari Standard Data, the SWR product is used in this work and is calculated based on the Himawari data [41], [48], [49]. The shape of the SWR product is the same as the shape of the detected matrices, corresponding to the theoretically inferred irradiance at each latitude and longitude position. There is a transport and preprocessing delay of 30 mins in obtaining the SWR data¹.

¹The SWR dataset is now publicly available <https://himawari8.nict.go.jp/> and <https://www.eorc.jaxa.jp/ptree/index.html> for nonprofit purposes.

B. Data Preprocessing

In this experiment, we choose a subset of the PVOD to validate GSTANN, which provides a 10×14 SWR grid. This subset of PVOD provides a dataset comprising 146 nodes, including 140 SWR points and 6 PV stations, whose range is 36.35°N – 36.8°N latitude and 113.5°W – 114.15°W longitude. We build the spatial graph G based on these 146 nodes. The SWR is a 10-min interval matrix, whereas each CPV station is a 15-min interval sequence. We extract a 10-min sequence of each point from the SWR and interpolate it into a 15-min sequence that aligns with the measured solar irradiance sequence of the CPV stations. As shown in Fig. 6, we calculated the Pearson correlation coefficient between all features and power output. Since the Pearson correlation coefficient between solar irradiance and power is the highest (0.98), we select global irradiance as the input of the encoder and data fusion.

C. Experimental Setting

1) Evaluation

We evaluate the performance of our model with respect to three widely used metrics, namely, the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The equations of metrics are expressed as:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t| \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2} \quad (15)$$

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (16)$$

2) Hyperparameters

For the 30-min delay corresponding to the SWR grid and the 15-min time interval of Hebei PV stations, we set time index $J = 4$ (1 h), $K = 4$ (1 h), time lag $L = 2$ (30 mins), $N = 146$, and $M = 1$. Then, we could predict the $K - L$ time steps (45 mins, 60 mins). In spatial dependency modelling, $\kappa = 0.05$ owing to the latitude and longitude resolution of SWR, $r = 6371$ km, and the embedding dimension is $D = 64$. We apply five CTSA layers in both the encoder and decoder blocks. We use the Adam optimizer, with a learning rate $\lambda = 0.0001$, to train 110 epochs, where $\eta_1 = 0.5$ and $\eta_2 = 0.5$ in the loss function.

3) Baseline Metrics

For comparison, we employ five classic baseline methods: (1) ARIMA: Auto-regressive integrated moving average is a statistic model for time series prediction task [13], (2) SVR: Support vector regression [50] uses specific training features to predict the target through regression analysis, (3) FC-LSTM: Sequence-to-sequence model with fully connected LSTM layers in both the encoder and decoder [51], (4) ST-GCN: Spatial-temporal graph convolutional network that uses graph convolutional layers for forecasting [31], and (5) GMAN:

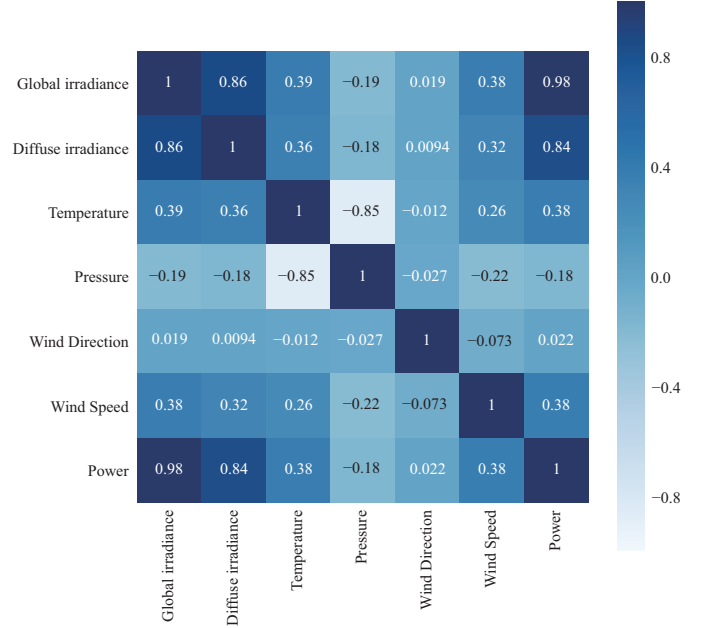


Fig. 6. Pearson correlation coefficient for all features.

Graph multi-attention network that leverages *node2vec* [52] to learn node representations [45].

Owing to the limitation of timeline misalignment, the above comparison experiments can only use T_h data but not T_l lag data. For fairness in considerations, five baselines and GSTANN-no-data-fusion (No-DF) are employed to predict PV power in T_f based on historical data in T_h . Additionally, we evaluate the performance of the original GSTANN.

D. Results and Discussion

Tables I and II present the overall results of and comparisons between the baselines on a subset of PVOD in terms of MAE, RMSE, and MAPE. No-DF outperforms five baselines in terms of MAE and RMSE, and performs comparably to MAPE. Notably, for the 45-min and 60-min power predictions, the average performance of GSTANN is significantly better than other methods. In particular, in short-term prediction, the 60-min RMSE of GSTANN still retains high accuracy compared with the 45-min RMSE. Moreover, we can observe those deep learning methods (e.g., FC-LSTM, ST-GCN, and GMAN, they pay attention to either spatial or temporal correlations) and perform better than traditional models (e.g., ARIMA, SVR).

In real-world applications, due to lack of local measurement devices for DPV, most existing methods (e.g., ARIMA, SVR, and FC-LSTM) are designed for CPV stations, then ST-GCN and GMAN are originally designed for traffic prediction problems, but encoder-decoder architecture has been well verified in time series forecasting. GSTANN focuses on DPV power forecasting exploiting neighboring CPV station nodes and SWR points, spatial or temporal features are both exhibited.

According to the weather record, we notice that December 8th, 13th, and 18th were almost sunny; first two were occasionally cloudy. As shown in Fig. 7, the GSTANN model can predict peaks and fluctuations pretty well on sunny days in both 45- and 60-min PV power prediction, whereas other

TABLE I
RESULTS OF 45-MIN PV POWER FORECASTING OF DIFFERENT STATIONS ON PVOD

Models	45-min (MAE (MW)/RMSE (MW)/MAPE (%))						Average
	Station-000	Station-001	Station-002	Station-003	Station-004	Station-005	
ARIMA	2.58/2.98/135	1.87/2.21/67	2.24/2.62/66	2.72/3.27/102	2.44/2.81/96	2.58/2.99/77	2.41/2.81/91
SVR	1.70/2.40/57	1.30/1.85/43	1.50/2.08/45	1.72/2.39/71	1.61/2.29/36	1.70/2.40/45	1.59/2.24/49
FC-LSTM	2.65/3.30/60	2.31/2.75/61	2.21/2.77/57	2.69/3.31/59	2.68/3.29/56	2.83/3.43/59	2.56/3.14/59
ST-GCN	1.40/1.95/35	1.05/1.42/37	1.18/1.49/25	1.98/2.54/49	1.47/1.91/108	1.60/2.28/33	1.45/1.93/48
GMAN	1.36/1.81/71	1.40/1.77/52	1.58/2.04/51	1.84/2.32/69	1.27/1.77/47	1.86/2.31/49	1.55/2.00/56
No-DF	1.40/1.91/39	1.62/2.16/46	1.48/2.08/37	1.84/2.39/53	1.36/1.93/34	1.84/2.38/34	1.59/2.14/40
GSTANN-S	1.33/1.80/46	1.27/1.64/40	1.40/1.88/38	1.95/2.42/51	1.19/1.69/34	1.76/2.21/36	1.48/1.94/41
GSTANN	1.31/1.81/45	1.06/1.45/29	1.03/1.49/28	1.28/1.77/67	1.25/1.73/30	1.26/1.85/29	1.20/1.68/38

TABLE II
RESULTS OF 60-MIN PV POWER FORECASTING OF DIFFERENT STATIONS ON PVOD

Models	60-min (MAE (MW)/RMSE (MW)/MAPE (%))						Average
	Station-000	Station-001	Station-002	Station-003	Station-004	Station-005	
ARIMA	3.00/3.43/270	2.19/2.56/116	2.54/2.93/83	3.13/3.70/138	2.95/3.32/164	3.08/3.52/247	2.81/3.24/170
SVR	1.85/2.64/82	1.46/2.07/57	1.58/2.21/51	1.89/2.69/79	1.78/2.52/43	1.9/2.68/86	1.74/2.47/66
FC-LSTM	2.57/3.20/52	2.21/2.65/53	2.15/2.70/45	2.61/3.22/48	2.62/3.22/51	2.73/3.31/51	2.48/3.05/50
ST-GCN	1.50/2.05/41	1.14/1.53/44	1.25/1.58/28	2.11/2.71/66	1.48/1.93/69	1.62/2.30/36	1.51/2.02/47
GMAN	1.50/1.99/125	1.64/2.04/84	1.69/2.20/61	1.93/2.44/80	1.43/1.93/72	2.00/2.47/125	1.70/2.18/91
No-DF	1.43/1.98/55	1.69/2.25/56	1.52/2.14/42	1.88/2.46/58	1.39/1.98/41	1.84/2.39/53	1.63/2.20/51
GSTANN-S	1.40/1.88/69	1.43/1.83/61	1.38/1.92/44	1.99/2.53/60	1.29/1.80/51	1.86/2.37/69	1.56/2.05/59
GSTANN	1.36/1.89/55	1.09/1.51/42	1.07/1.56/31	1.36/1.85/76	1.24/1.74/33	1.29/1.91/55	1.24/1.74/49

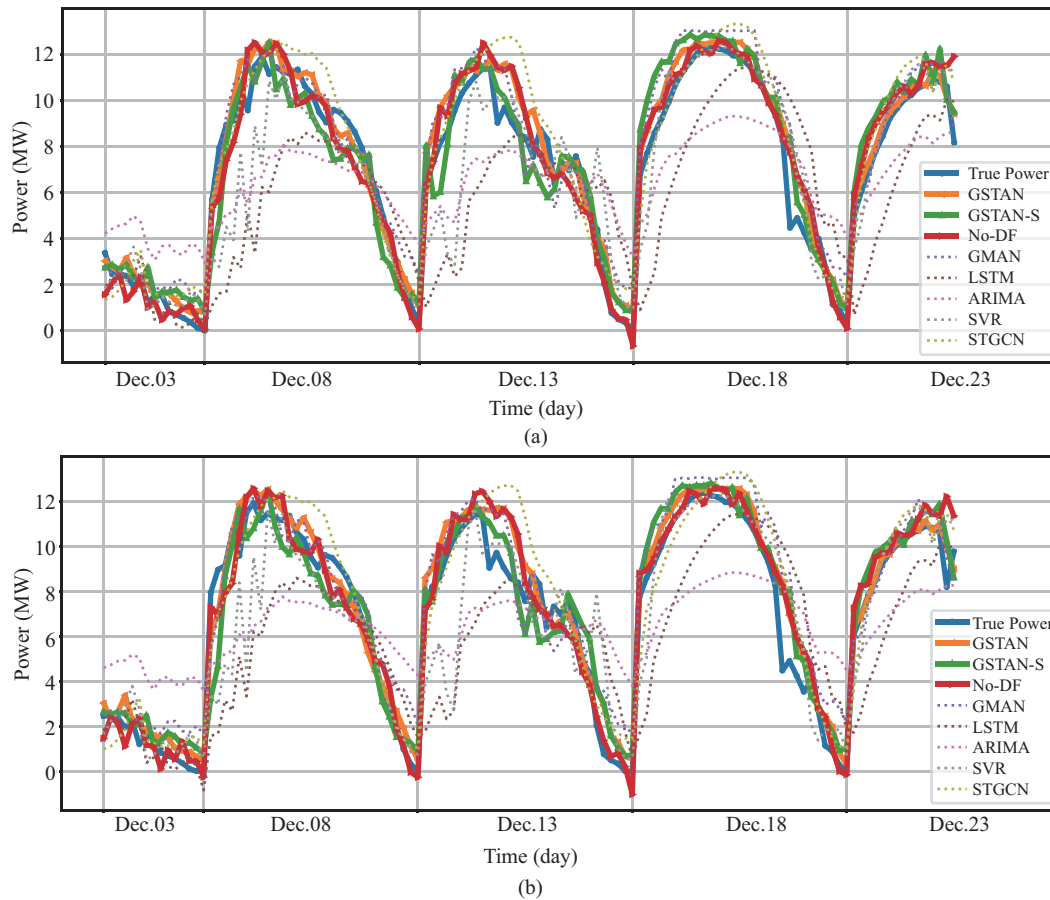


Fig. 7. Schematic of PV power prediction performance (Station 0). (a) Schematic of the 45-min prediction experiment. (b) Schematic of the 60-min prediction experiment.

models cannot, especially in the small-range fluctuations on December 8th. In addition, on December 13th and 18th, GMAN and ST-GCN incorrectly predicted the peak position of the PV power curve, and were unable to perceive the

fluctuation of power due to cloud flow. Experimental results also demonstrate the GSTANN model can timely sense rapid changes in power under these weather conditions, including increase and decrease of power, but other models (e.g., ST-

GCN, LSTM, SVR, ARIMA) have obvious perception delays. Although the 60-min prediction error of ST-GCN is 2 percentage points lower than of GSTANN in terms of MAPE, from Fig. 7(b), the ST-GCN prediction curve still has an offset. Considering the three evaluation metrics, GSTANN is superior to other models in overall prediction performance. In particular, the prediction time of all models in this paper, including GSTANN, is less than 1 millisecond, thus the model inference time is in line with the time requirement of PV power prediction.

E. Ablation Studies

To investigate the effect of each component on our model, we evaluate the model by removing data fusion, and CTSA, hereafter represented by No-DF, and GSTANN-S, respectively. Then we also develop four graph neural network (GNN) variants based on GSTANN to show the scalability of the proposed network.

In GSTANN-S, the CTSA mechanism is divided into two separate modules, i.e., temporal attention, and spatial attention. The former considers the historical state of the current node, and the latter only considers the states of the surrounding nodes at one moment. In four GNN variants experiments, module GCN is replaced by GraphSAGE, GATconvs, GINconvs, and DeeperGCN, respectively. Their inputs are simplified to $\mathbf{A}$ and the identity matrix $\mathbf{I}_N$, then we retrain every model to include GSTANN without feature matrix $\mathbf{X}_g$. In No-GCN, the

output features of GCN are all filled with zeros in the case of model integrity. Hidden channels of GCN, GraphSAGE, and DeeperGCN are all 146, and the hidden dimension of GIN adopts default setting 32. The first hidden channel and heads of GAT are set to 16 and 8, respectively. The output of all GNN variants is the same as GCN ($D = 64$).

Table I and II clearly indicate that when the data fusion module is removed, the evaluation error significantly increases both within 46 and 60 mins. This proves the proposed data fusion can effectively handle time misalignment between delayed SWR and measured solar irradiance of CPV stations to obtain more accurate power prediction. Compared with GSTANN-S, GSTANN with CTSA has shown advantages in 60-min time step prediction in most power stations.

The PV power prediction task is more concerned with some fluctuations. According to Fig. 8(a), (b), on one hand, it can be seen that GCN and GIN can effectively extract spatial embedding features and assist in learning peak power. On the other hand, without the GCN module, the model can hardly perceive peak power from the surrounding nodes. Lacking edges weights $\mathbf{X}_g$ in this ablation study, performance of some GNN variants (e.g., GraphSAGE, GATconvs, and DeeperGCN) does not perform as well as GCN and GIN. Considering lightweight deployment and real-time inference, some GNN variant models, which are too redundant or are not suitable for time series prediction, may not apply to actual DPV prediction application.

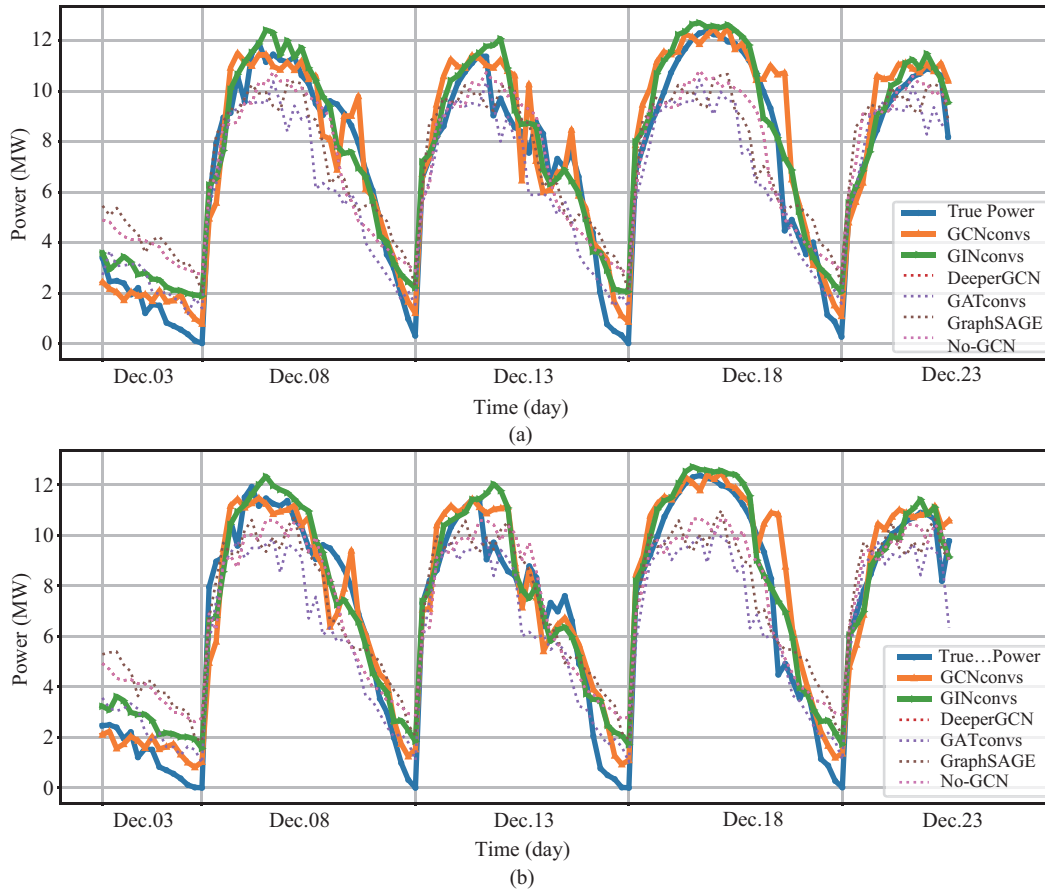


Fig. 8. Schematic of ablation study (Station 0). (a) Schematic of the 45-min GNN ablation study. (b) Schematic of the 60-min GNN ablation study.

V. CONCLUSION

In this study, we have proposed GSTANN to predict very short-term DPV power. First, we used satellite remote sensing data covering a specific geographical area, to enhance the meteorological features of all the power stations. Second, we incorporated a graph convolution block and modelled the non-Euclidean local and global spatial dependence of the SWR grid and power stations. Moreover, temporal and spatial correlations were simultaneously obtained by using the CTSA. Then, we proposed a data fusion to resolve the time lag between solar irradiance and satellite remote sensing data and applied a power approximation module to capture the mapping function of DPV power. Experiments on real-world case studies with multiple power stations demonstrated the prediction performance of GSTANN was significantly better than those of the five state-of-the-art baselines. Furthermore, this study is partially suitable for transfer to improve the calculation accuracy of satellite remote sensing data, based on the data fusion of PV power and ground observation.

Deep learning methods are indeed superior to traditional methods in DPV power forecasting, but they are also regarded as a black box forecasting method. One future work is to study the internal interpretability of the model, such as visualizing and explaining the significance of the weight of each module. Other work is to experiment with more power stations and explore more novel graph neural networks to improve accuracy of DPV power forecasting.

REFERENCES

- [1] Y.-C. Tsai, Y.-K. Chan, F.-K. Ko, and J.-T. Yang, "Integrated operation of renewable energy sources and water resources," *Energy Conversion and Management*, vol. 160, pp. 439–454, Mar. 2018.
- [2] C. Wan, J. Zhao, Y. H. Song, Z. Xu, J. Lin, and Z. C. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38–46, Dec. 2015.
- [3] B. Yang, T. Zhu, P. Cao, Z. Guo, C. Zeng, D. Li, Y. Chen, H. Ye, R. Shao, H. Shu, and T. Yu, "Classification and summarization of solar irradiance and power forecasting methods: A thorough review," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 3, pp. 978–995, 2023.
- [4] D. Z. Yang, W. X. Li, G. M. Yagli, and D. Srinivasan, "Operational solar forecasting for grid integration: Standards, challenges, and outlook," *Solar Energy*, vol. 224, pp. 930–937, Aug. 2021.
- [5] S. C. Wang, "Current status of PV in China and its future forecast," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 72–82, Mar. 2020.
- [6] Y. P. Qu, J. Xu, Y. Z. Sun, and D. Liu, "A temporal distributed hybrid deep learning model for day-ahead distributed PV power forecasting," *Applied Energy*, vol. 304, pp. 117704, Dec. 2021.
- [7] H. F. Li, H. Y. Lin, Q. K. Tan, P. Wu, C. J. Wang, G. De, and L. L. Huang, "Research on the policy route of China's distributed photovoltaic power generation," *Energy Reports*, vol. 6, pp. 254–263, Nov. 2020.
- [8] T. Zhong, Z. X. Zhang, M. Chen, K. Zhang, Z. X. Zhou, R. Zhu, Y. J. Wang, G. N. Lü, and J. Y. Yan, "A city-scale estimation of rooftop solar photovoltaic potential based on deep learning," *Applied Energy*, vol. 298, pp. 117132, Sep. 2021.
- [9] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, Oct. 2016.
- [10] R. R. Rao, M. Mani, and P. C. Ramamurthy, "An updated review on factors and their inter-linked influences on photovoltaic system performance," *Heliyon*, vol. 4, no. 9, pp. e00815, Sep. 2018.
- [11] A. Ahmed and M. Khalid, "A review on the selected applications of forecasting models in renewable power systems," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 9–21, Feb. 2019.
- [12] T. Ahmad, H. C. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," *Sustainable Cities and Society*, vol. 55, pp. 102052, Apr. 2020.
- [13] S. Makridakis and M. Hibon, "ARMA models and the Box–Jenkins methodology," *Journal of Forecasting*, vol. 16, no. 3, pp. 147–163, May 1997.
- [14] Y. T. Li, Y. Su, and L. J. Shu, "An ARMAX model for forecasting the power output of a grid connected photovoltaic system," *Renewable Energy*, vol. 66, pp. 78–89, Jun. 2014.
- [15] W. VanDeventer, E. Jamei, G. S. Thirunavukkarasu, M. Seyedmahmoudian, T. K. Soon, B. Horan, S. Mekhilef, and A. Stojcevski, "Short-term PV power forecasting using hybrid GASVM technique," *Renewable Energy*, vol. 140, pp. 367–379, Sep. 2019.
- [16] P. J. Lin, Z. N. Peng, Y. F. Lai, S. Y. Cheng, Z. C. Chen, and L. J. Wu, "Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets," *Energy Conversion and Management*, vol. 177, pp. 704–717, Dec. 2018.
- [17] F. Wang, Z. M. Xuan, Z. Zhen, K. P. Li, T. Q. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Conversion and Management*, vol. 212, pp. 112766, May 2020.
- [18] P. T. Li, K. L. Zhou, X. H. Lu, and S. L. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Applied Energy*, vol. 259, pp. 114216, Feb. 2020.
- [19] M. Yang, X. Shen, D. Huang, and X. Su, "Fluctuation classification and feature factor extraction to forecast very short-term photovoltaic output powers," *CSEE Journal of Power and Energy Systems*, doi: 10.17775/CSEEJPES.2022.03760.
- [20] M. Shi, K. Xu, J. Wang, R. Yin, and P. Zhang, "Short-term photovoltaic power forecast based on grey relational analysis and GeoMAN model," *Transactions of China Electrotechnical Society*, vol. 36, no. 11, pp. 2298–2305, Jun. 2021.
- [21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 6000–6010.
- [22] J. Simeunović, B. Schubnel, P. J. Alet, and R. E. Carrillo, "Spatio-temporal graph neural networks for multi-site PV power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1210–1220, Apr. 2022.
- [23] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, Nov. 2013.
- [24] Z. B. Liu, C. B. Xu, X. L. Wang, Q. Chen, X. Zuo, Y. Wu, and D. L. Xu, "A photovoltaic power prediction method based on remote sensing data," *Spacecraft Recovery & Remote Sensing*, vol. 42, no. 5, pp. 85–95, Oct. 2021.
- [25] T. C. Yao, J. Wang, H. Y. Wu, P. Zhang, S. G. Li, K. Xu, X. Y. Liu, and X. B. Chi, "Intra-hour photovoltaic generation forecasting based on multi-source data and deep learning methods," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 1, pp. 607–618, 2022.
- [26] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. J. Millar, "Adaptive predictor subset selection strategy for enhanced forecasting of distributed PV power generation," *IEEE Access*, vol. 7, pp. 90652–90665, 2019.
- [27] H. Zhen, D. X. Niu, K. K. Wang, Y. C. Shi, Z. S. Ji, and X. M. Xu, "Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information," *Energy*, vol. 231, pp. 120908, Sep. 2021.
- [28] H. X. Zhou, Y. J. Zhang, L. F. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, 2019.
- [29] D. Yu, W. Choi, M. Kim, and L. Liu, "Forecasting day-ahead hourly photovoltaic power generation using convolutional self-attention based long short-term memory," *Energies*, vol. 13, no. 15, pp. 4017, Aug. 2020.
- [30] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2017.
- [31] B. Yu, H. T. Yin, and Z. X. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," arXiv preprint arXiv:1709.04875, 2018.
- [32] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. H. Wang, and G. Y. Liu, "Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 571–583, Apr. 2020.
- [33] C. Dinesh, S. Makonin, and I. V. Bajić, "Residential power forecasting using load identification and graph spectral clustering," *IEEE Trans-*

actions on Circuits and Systems II: Express Briefs, vol. 66, no. 11, pp. 1900–1904, Nov. 2019.

- [34] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 1025–1035.
- [35] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” arXiv preprint arXiv:1710.10903, 2018.
- [36] K. Xu, W. H. Hu, J. Leskovec, and S. Jegelka, “How powerful are graph neural networks?” arXiv preprint arXiv:1810.00826, 2019.
- [37] G. H. Li, C. X. Xiong, A. Thabet, and B. Ghanem, “DeeperGCN: All you need to train deeper GCNs,” arXiv preprint arXiv:2006.07739, 2020.
- [38] P. Mathiesen and J. Kleissl, “Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states,” *Solar Energy*, vol. 85, no. 5, pp. 967–977, May 2011.
- [39] K. Bakker, K. Whan, W. Knap, and M. Schmeits, “Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation,” *Solar Energy*, vol. 191, pp. 138–150, Oct. 2019.
- [40] Y. W. Fu, H. Chai, Z. Zhen, F. Wang, X. J. Xu, K. P. Li, M. Shafie-Khah, P. Dehghanian, and J. P. S. Catalão, “Sky image prediction model based on convolutional auto-encoder for minutely solar PV power forecasting,” *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 3272–3281, Jul./Aug. 2021.
- [41] R. Frouin and H. Murakami, “Estimating photosynthetically available radiation at the ocean surface from ADEOS-II global imager data,” *Journal of Oceanography*, vol. 63, no. 3, pp. 493–503, Jun. 2007.
- [42] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, and C. Cornaro, “Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data,” *Solar Energy*, vol. 158, pp. 1026–1038, Dec. 2017.
- [43] C. C. Robusto, “The cosine-haversine formula,” *The American Mathematical Monthly*, vol. 64, no. 1, pp. 38–40, Jan. 1957.
- [44] A. E. Orhan and X. Pitkow, “Skip connections eliminate singularities,” arXiv preprint arXiv:1701.09175, 2018.
- [45] C. P. Zheng, X. L. Fan, C. Wang, and J. Z. Qi, “GMAN: A graph multi-attention network for traffic prediction,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 1, pp. 1234–1241, Apr. 2020.
- [46] W. Zhou, H. X. Yang, and Z. H. Fang, “A novel model for photovoltaic array performance prediction,” *Applied Energy*, vol. 84, no. 12, pp. 1187–1198, Dec. 2007.
- [47] T. C. Yao, J. Wang, H. Y. Wu, P. Zhang, S. G. Li, Y. G. Wang, X. B. Chi, and M. Shi, “A photovoltaic power output dataset: Multi-source photovoltaic power output dataset with python toolkit,” *Solar Energy*, vol. 230, pp. 122–130, Dec. 2021.
- [48] D. Z. Yang and J. M. Bright, “Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years,” *Solar Energy*, vol. 210, pp. 3–19, Nov. 2020.
- [49] Y. C. Yu, J. C. Shi, T. X. Wang, H. S. Letu, P. F. Yuan, W. Zhou, and L. Hu, “Evaluation of the himawari-8 shortwave downward radiation (SWDR) product and its comparison with the CERES-SYN, MERRA-2, and ERA-interim datasets,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 2, pp. 519–532, Feb. 2019.
- [50] C.-H. Wu, J.-M. Ho, and D.-T. Lee, “Travel-time prediction with support vector regression,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276–281, Dec. 2004.
- [51] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014, pp. 3104–3112.
- [52] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 855–864.



Jue Wang received the Ph.D. degree in Computer Science from the University of Science and Technology of China, Beijing, China. He is currently a Research Fellow of Computer Network Information Center of the Chinese Academy of Sciences. His main research interests are artificial intelligence algorithms and high performance computation.



Yangang Wang received a Ph.D. degree in Computational Mathematics from the Jilin University, Jilin, China. He is currently a Research Fellow of Computer Network Information Center of the Chinese Academy of Sciences. His main research interests are artificial intelligence algorithms and high performance computation.



Pei Zhang received a Ph.D. degree at Imperial College London, London, United Kingdom. He is currently a professor at School of Electrical Engineering, Beijing Jiaotong University. His main research interest includes renewable energy forecasting and integration.



Haizhou Cao received a bachelor's degree in Computer Science and Technology from Zhengzhou University, Zhengzhou, China, in 2020. He is an M.S. student at the Computer Network Information Center of the Chinese Academy of Sciences. He is also an M.S. student at the University of Chinese Academy of Sciences. His research interests include time series forecasting, PV power forecasting, computer vision, and data mining.



Xuebin Chi received Ph.D. and M.S. degrees in Computational Mathematics from the Computing Center, Chinese Academy of Sciences, where he is currently a Research Fellow of the Computer Network Information Center of the Chinese Academy of Sciences. His research interests include high-performance computing, machine learning, and grid technology.



Tiechui Yao is a Ph.D. student at the Computer Network Information Center of the Chinese Academy of Sciences. He is also a Ph.D. student at the University of Chinese Academy of Sciences. His research interests include PV power forecasting, time series forecasting, high-performance computing, and artificial intelligence application.



Min Shi is currently a Senior Engineer with State Grid Hebei Electric Power Co., Ltd., Shijiazhuang, China. He graduated from the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include power system operation and control, renewable energy generation systems.