



# Transformer approach to nowcasting solar energy using geostationary satellite data

Ruohan Li<sup>a</sup>, Dongdong Wang<sup>a,b,\*</sup>, Zhihao Wang<sup>a</sup>, Shunlin Liang<sup>c</sup>, Zhanqing Li<sup>d</sup>, Yiqun Xie<sup>a</sup>, Jienna He<sup>a</sup>

<sup>a</sup> Department of Geographical Sciences, University of Maryland, College Park, MD 20742, United States

<sup>b</sup> Institute of Remote Sensing and GIS, School of Earth and Space Sciences, Peking University, Beijing 100871, China

<sup>c</sup> Department of Geography, University of Hong Kong, Hong Kong, 999077, China

<sup>d</sup> Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, United States

## HIGHLIGHTS

- Transformer-based deep learning approach (SolarFormer) is proposed for near-real-time solar radiation nowcasting.
- SolarFormer uses geostationary satellite observations as input without need of ground data.
- SolarFormer outperforms other machine learning models in hourly predictions.
- SolarFormer offers extended lead-time forecasting potential due to its efficiency.

## ARTICLE INFO

### Keywords:

Solar energy forecasting  
Photovoltaic technology  
Deep learning  
Transformer  
Near real-time nowcasting  
Geostationary satellite

## ABSTRACT

Unpredicted spatial and temporal variability of global horizontal irradiance (GHI) reaching the photovoltaic panels presents a challenge for integrating solar power into the grid stably and cost-effectively at a regional scale. Therefore, there is a recognized demand for large-scale GHI nowcasting that is both timely and accurate, an area where most existing studies fall short. This study introduces the SolarFormer model, which utilizes satellite data and incorporates a gated recurrent unit for near real-time GHI estimation. It also includes a space-time transformer to provide forecasts with a 3-h lead time at 15-min intervals, maintaining accuracy without significant degradation over extended lead times. SolarFormer requires only the selected satellite band information shared by GOES-16 and Himawari-8 as the dynamic input, enabling near-real-time application across all areas covered by these satellites. This feature makes it accessible and efficient for large-scale energy planning. We validate the forecasting result with the ground-measured GHI over seven SURFRAD stations in 2018. The model achieves an hourly prediction root-mean-square error (relative root-mean-square error) of 93.8 W/m<sup>2</sup> (15.0 %), 118.9 W/m<sup>2</sup> (19.8 %), and 129.1 W/m<sup>2</sup> (24.2 %) with 1–3 h lead time respectively. These results demonstrate lower root-mean-square error compared to existing hourly updated numerical weather prediction modeling, such as High-Resolution Rapid Refresh, and deep learning models, such as ConvLSTM. Moreover, the study highlights the potential of SolarFormer for extended lead-time forecasting due to its high computation and memory efficiency compared with the above-mentioned models, potentially benefiting long-term energy planning and power market bidding and clearing. However, SolarFormer exhibits accumulated bias as the predicted lead time increases and faces challenges in predicting GHI in the early morning due to the invalid visible satellite bands during the night, suggesting areas for improvement in future studies.

## 1. Introduction

Solar energy has risen to prominence, comprising 27 % of renewable

energy generation in the past decade [5], due to advancements in photovoltaic (PV) technologies and the increasing demand for renewable energy sources. Global Horizontal Irradiance (GHI) which directly

\* Corresponding author at: Department of Geographical Sciences, University of Maryland, 2181 LeFrak Hall, College Park, MD 20742, USA.

E-mail address: [ddwang@umd.edu](mailto:ddwang@umd.edu) (D. Wang).

<https://doi.org/10.1016/j.apenergy.2024.124387>

Received 29 March 2024; Received in revised form 12 July 2024; Accepted 2 September 2024

Available online 10 September 2024

0306-2619/© 2024 Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.

influences the amount of solar radiation received by PV panels is one of the most important factors that affect the PV output. However, unlike more consistent energy sources such as hydropower or nuclear energy, GHI experiences significant variations. These variations stem from predictable factors such as solar zenith angles (SZA), which induce seasonal and diurnal changes, as well as dynamic factors with significant fluctuations caused by cloud cover and atmospheric conditions. For example, a disruption in dense and fragmented cloud cover can lead to a sudden surge, of approximately  $700 \text{ W/m}^2$  within a 30-min timeframe [35].

Solar PV devices respond rapidly to changes in GHI, and the swift fluctuations in PV output pose challenges for utility operators in balancing supply and demand in real time. Moreover, unforeseen short-term variations can lead to increased financial costs. For instance, operators may need to invest in larger battery storage systems [43] to store excess energy during periods of high generation and discharge it during dips in production. Additionally, they may incur higher operational expenses for peaking power plants [59] that are used to meet sudden spikes in demand or compensate for fluctuations in renewable energy output. Therefore, accurate and timely short-term nowcasting of GHI is necessary to maintain grid stability and cost-effective utilization of solar energy [14,35].

Moreover, with the increasing penetration of both macro and micro-scale PV, the need for GHI forecasting is no longer spatially insulated at specific solar sites. Nowcasting over broad geographic areas will support regional energy planning by optimizing energy deployment strategies [53]. Operators can strategically place renewable energy infrastructure to maximize efficiency and minimize costs [1]. Moreover, large-scale and timely forecasting will further refine electricity allocation strategies, mitigate risks associated with weather-related variability [18], and prevent emergency power supply issues [33,39,48].

Site-based forecasting has been dominant for short-term predictions (0–6 h) in existing studies. Surface sky camera networks are the primary data sources for ultra-short-term solar energy forecasts within 1 h [30]. Broader short-term forecasts incorporate historical site-specific solar radiation records and weather information. Forecasting methodologies include numerical weather prediction (NWP) [9], statistical approaches [20,57], machine-learning techniques [17,46,63], and deep learning strategies with data from neighboring sites [13,19,25,45,51,60]. However, due to their limited information on surrounding atmospheric conditions, site-based forecasts degrade in environments with significant variability or for predictions with longer lead times [46].

Satellite platforms provide insights into cloud patterns that surpass the capabilities of local, surface-based networks, making them particularly suitable for intraday GHI predictions [35]. Recognizing this, recent research has explored integrating satellite data with site-based observations to enhance short-term forecasting. For example, Choi et al. [8] applied a convolutional neural network (CNN) model to NOAA's GOES8 infrared channels for site-specific GHI predictions. Gallo et al. [16] employed 3D CNN and convolutional long short-term memory (ConvLSTM) with the spinning enhanced visible infra-red imager (SEVIRI) instrument on the Meteosat Second Generation satellite, combining it with ground measurements under clear conditions. Qin et al. [44] integrated contextual LSTM methodologies with satellite and ground data. Paletta et al. [40] achieved high accuracy for ultra-short prediction by combining satellite observations, sky images, as well as ground measurements using an encoder-decode structure. However, these methods still heavily rely on ground measurements, which limits their applicability in regional energy planning.

Existing regional-scale forecasting commonly employs NWP models, which utilize geographical and meteorological data with notable accuracy. However, despite recent advances in hourly-updated NWPs, their spatial prediction resolution remains less refined than ideal for photovoltaic grid operators [14]. Cloud motion vector (CMV) has been developed for satellite-image-based forecasting [32]. The State University of New York at Albany (SUNY Albany) solar forecasting model,

integrated into the SolarAnywhere software, refines CMV using NWP models and boasts a 25.5 % accuracy for 3-h forecasts [41]. However, this product is not freely accessible. Kosmopoulos et al. [24] and Carpentieri et al., [6] further refined the CMV model, achieving high accuracy across Europe and North Africa. These methods depend on extra satellite products for calculating cloud indexes based on radiative transfer models (RTM) which constrains its capabilities for near-real-time operations and its applicability over broader regions. Data-driven machine-learning methods, with cloud movement calculations at their core, offer alternative solutions. Pérez et al. [43] applied a CNN model to solar energy datasets from Meteosat's SEVIRI, although its applicability remains limited to a single test site. Nielsen, Iosifidis, & Karstoft [38] employed a ConvLSTM model on effective cloud albedo data from SEVIRI over Europe. Xia et al. [56] applied the predictive recurrent neural network (PredRNN++) to Advanced Himawari Imager (AHI) level-1B radiance data from the Himawari-8/9 satellite over China, but the study was validated over cloud cover fraction but not exact GHI values and the model degrades significantly after 2 h lead time. Nonetheless, both ConvLSTM and PredRNN encounter scalability issues, particularly with larger images and higher temporal resolutions that strain computational and memory capacities [15].

Though various methods and datasets have been utilized for intraday nowcasting of solar radiation, achieving near-real-time, accurate, regional-scale forecasting of GHI with an efficient model structure remains a challenge. Here, we propose a novel approach, SolarFormer, that integrates Bi-GRU and the Space-Time Transformer using GOES-16 data to address these challenges. The high spatial-temporal resolution of geostationary satellite GOES-16 provides extensive cloud movement information at large scales. SolarFormer utilized the capabilities of the existing Bi-GRU model, enabling near-real-time estimation of GHI images using satellite bands as the only dynamic inputs [26]. The Space-Time Transformer part enhances forecasting accuracy by incorporating a parallel cuboid attention mechanism, which selectively focuses on relevant spatial and temporal features, effectively capturing long-range dependencies in a computationally efficient manner [50].

This model is designed for direct application in real-time solar nowcasting. By leveraging satellite observations as the only dynamic inputs, SolarFormer achieves near-real-time GHI nowcasting within the satellite's typical update time, usually within 30 min. Moreover, its reliance solely on satellite data addresses data scarcity issues that often hinder refinements in many machine learning models. Additionally, SolarFormer's versatility allows it to utilize both GOES-16 and Himawari-8 data inputs, making it applicable across multiple continents. The novelty and key contributions of our paper can be summarized as follows:

- We proposed the SolarFormer, a novel integration of a space-time transformer with a gated recurrent unit (GRU) model, to forecast GHI up to 3 h ahead in 15-min intervals at a resolution of 5 km.
- The model utilizes geostationary satellite observations as dynamic inputs only, enabling near-real-time forecasting. and facilitating its use in large-scale PV management.
- The model demonstrates greater accuracy than baseline models, achieving stable forecasting skill of around 25 %, and exhibits improved time and memory efficiency over ConvLSTM.
- We conducted detailed investigations into the optimal spatiotemporal resolutions and attention mechanisms for transformers in GHI forecasting. The results show that divided space-time transformer presents the higher accuracy, thus adopted as the final model.

The structure of the paper is as follows: Section 2 examines *in situ* and satellite data. Section 3 details the methodology. Section 4 presents the validation results and discusses the limitations. Section 5 presents a conclusion.

## 2. Data

The input satellite data cover seven surface radiation budget observing network (SURFRAD) stations with  $600 \text{ km} \times 600 \text{ km}$  spatial scale each for the years 2018, 2019, and 2021 as shown in Fig. 1. Samples from 2019 and 2021 were designated as training data, whereas the 2018 samples were used for testing purposes. Additionally, we collected the ground measured GHI from the seven SURFRAD stations in 2018 for validation.

### 2.1. In situ validation data

In situ data from seven SURFRAD stations within the continental United States from 2018 were used to validate the SolarFormer model (Fig. 1). SURFRAD captures GHI measurements at 1-min intervals. These data were subjected to a quality assessment following the processes proposed by Li et al. [28]. Data with a SZA exceeding 85 degrees or with zero or negative values were excluded. To reduce potential errors in point-scale solar radiation measurement, data were smoothed over 15-min windows, centered on the satellite's pass time, as recommended by Huang et al. [23]. During this smoothing process, missing values were ignored unless all data within the 15-min window were missing; in such cases, the specific sample was excluded from validation. Hourly measurements were directly averaged within the corresponding hour windows.

### 2.2. Input data

The input data for the SolarFormer are summarized in Table 1. The primary inputs include Geostationary-NASA Earth Exchange (GeoNEX) Level 1G (L1G) top-of-atmosphere (TOA) reflectance data and associated geometry from the Advanced Baseline Imager (ABI) on GOES-16. The GeoNEX L1G TOA data are obtainable from NASA's Advanced Supercomputing (NAS) Division GeoNEX platform, as detailed by Wang et al. [52]. These data provide full-disk scans every 10 or 15 min, varying by sensor and year. Fifteen bands, ranging from  $0.47 \mu\text{m}$  to  $13.3 \mu\text{m}$ , were selected. Following Li et al. [27], we included all shared bands

from ABI and AHI, which are 15 bands in total. The geometry angles include SZA, viewing zenith angle (VZA), and relative azimuth angle (RAA). Static elevation information was sourced from GTOPO30.

### 2.3. Ancillary data

The NOAA GOES-R Series Level 2 clear sky mask was used to analyze the performance of the SolarFormer under various cloud-changing conditions in Section 4.2. Full-disk versions of the products were generated with the same temporal resolution as those of ABI observations and were retrieved at a horizontal spatial resolution of 2 km. The product was matched with the ground measurements after conversion to the geographic coordinate system.

## 3. Methods

### 3.1. Data preprocessing

Before being fed into the BiGRU part of the SolarFormer, the 19 variables listed in Table 1 are aligned according to the GeoNEX tiling convention, which uses a  $600 \times 600 \text{ km}$  grid at a 1-km resolution. The missing value will be filled with  $-1$  before the normalization. The BiGRU will be applied on each grid. The output of BiGRU only have one channel for all time steps in one day representing the GHI values. The output was reshaped to  $600 \times 600$  images, subsequently aggregated to a 5-km resolution and divided into four patches, with each patch covering an area of  $300 \times 300 \text{ km}$  and containing  $60 \times 60$  pixels. With the whole year of observations shaped as  $35,040 \times 60 \times 60 \times 1$  (where 35,040 is derived from a 15-min interval over a year, calculated as 4 intervals per hour \* 24 h per day \* 365 days per year), we subsample each by using a window time series size of 24 and an interval of 8. Any subsamples with invalid input (SZA greater than 85) are excluded from use.

The input sample for the space-time transformer model consists of 8 time steps (representing 2 h of input data), and the output sample consists of the subsequent 12 time steps (representing 3 h of lead time for forecasting). Therefore, each input sample has a size of  $8 \times 60 \times 60 \times 1$ , and each output sample has a size of  $12 \times 60 \times 60 \times 1$ . To mitigate

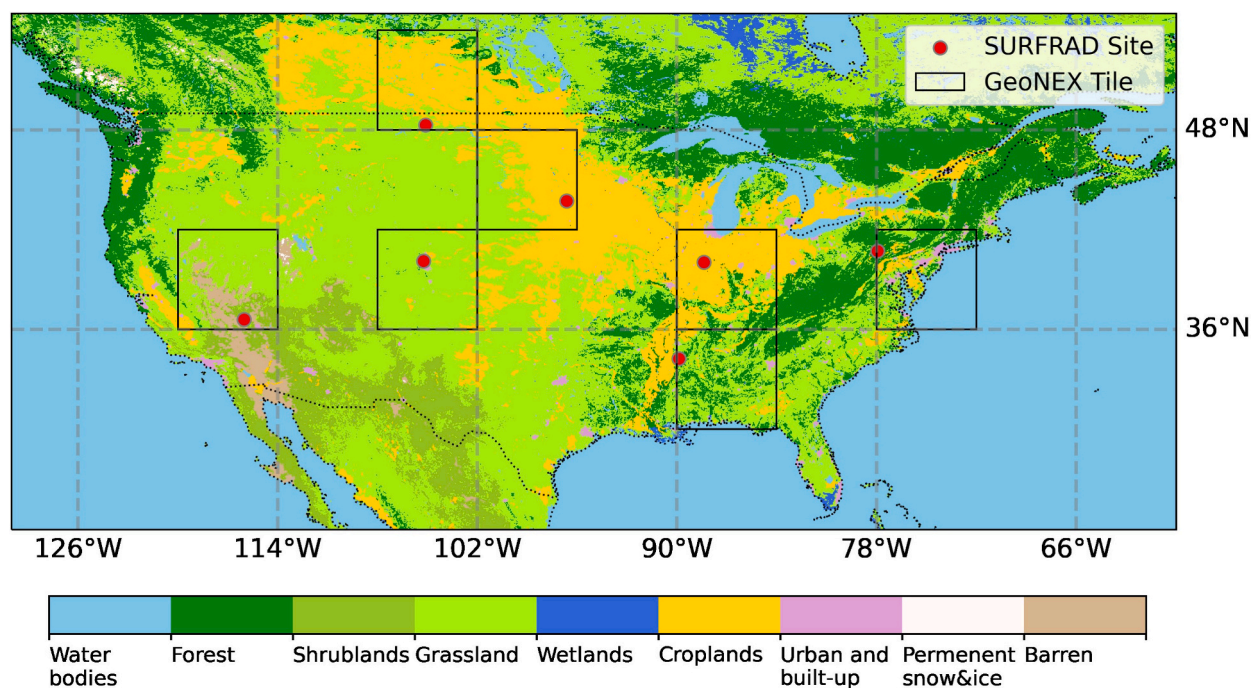


Fig. 1. The distribution of seven surface radiation budget observing network (SURFRAD) stations. The black rectangles outline the Geostationary-NASA Earth Exchange (GeoNEX) tile coverage areas.

**Table 1**  
Summary of input variables and corresponding datasets.

Variables	Dataset	Dimension	Spatial resolution	Temporal resolution	References
TOA reflectance	GOES16 L1G	15 (bands)	1 km	10 min or 15 min	[52]
Geometry angles	GOES16 L1G	3 (angles)	1 km	10 min or 15 min	[52]
Elevation	GTOPO30	1	30 arc sec	Static	[11]

the influence of SZA ( $\theta$ ) on diurnal GHI fluctuations, all GHI values retrieved from the BiGRUE part,  $R_i$ , are initially normalized by the theoretically clearest condition,  $E_{max,i}$ , at time  $i$ . The normalized tensor is represented as the clearness index or clear sky ratio (CSR),  $csr_i$ , calculated as follows:

$$E_{max,i} = E^* \cos(\theta_i) \quad (1)$$

$$csr_i = \frac{R_i}{E_{max,i}} \quad (2)$$

$E$  represents the solar constant, which is  $1360 \text{ W/m}^2$ . The closer the CSR value is to 0, the denser the cloud is.

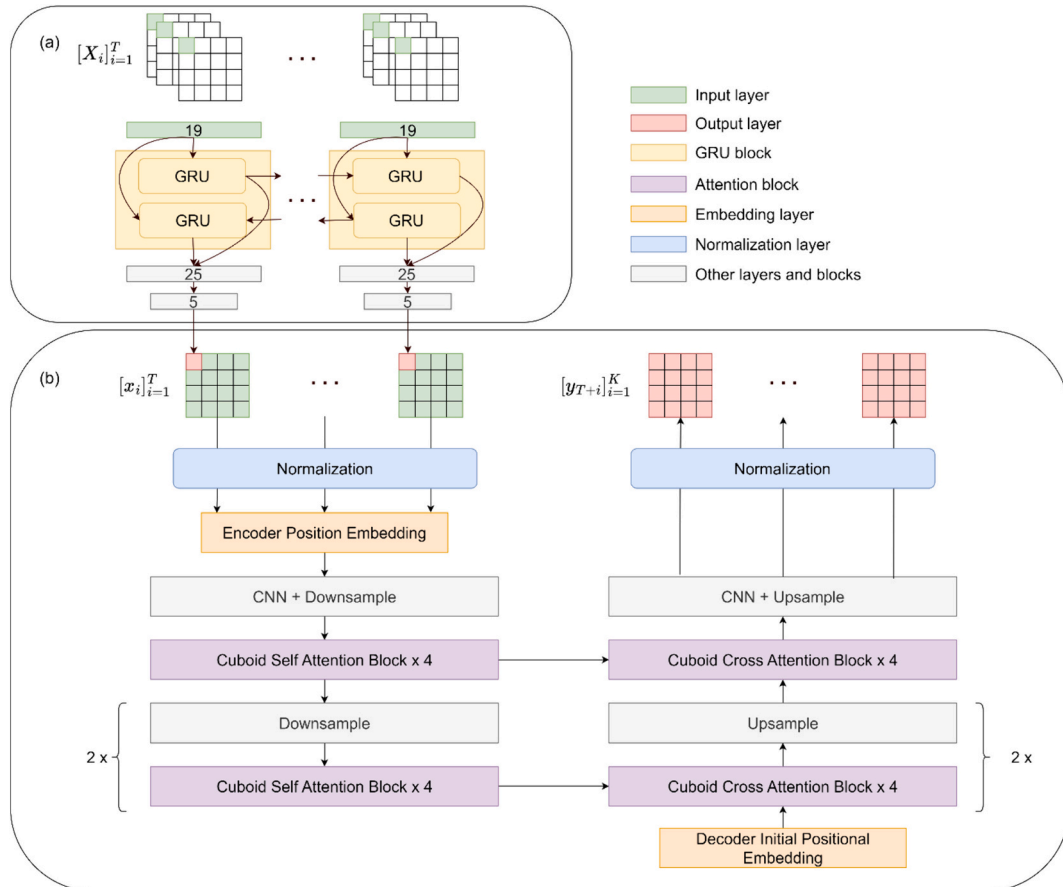
### 3.2. Models

The overall procedure of building SolarFormer is to first estimate past GHI from the satellite observations and then predict the future GHI based on the retrieved GHI images (Fig. 2). The estimation part builds the relationship between the TOA reflectance and surface radiation while the prediction part learns the spatial and temporal correlation of the atmospheric conditions from a time series of retrieved GHI images.

The near real-time GHI estimation is achieved by employing

developed and publicly accessible bi-directional gated recurrent unit (BiGRU) models (Fig. 2a) [26]. The model leverages temporal information within a single day to eliminate the need for additional products, such as surface albedo, thus enabling timely estimation of solar radiation [27]. It has been proved as one of the most accurate models to estimate GHI [26]. The estimation component includes a BiGRU layer with 32 units, two dense layers with 25 and 5 nodes, respectively, and an output layer. Li et al. [27] provided a comprehensive description and configuration of the BiGRU model.

The forecasting component utilizes the divided space-time transformer refined from EarthFormer to learn the cloud movements along both spatial and temporal dimensions (Fig. 2b). The core of the transformer is built on the attention mechanism, which allows the model to selectively focus on relevant features of the time series without requiring sequential processing of previous features. This enhances efficiency and overcomes the limitations of long-range dependencies inherent in traditional models like ConvLSTM [50]. An attention function computes a weighted sum of the values ( $V$ ), with weights determined by a compatibility function of the query ( $Q$ ) with the corresponding key ( $K$ ). For multi-head attention, projections  $w_Q$ ,  $w_K$ , and  $w_V$  are applied to  $Q$ ,  $K$ , and  $V$  for each head. Thus, multi-head attention can be summarized as follows:



**Fig. 2.** The structure of the SolarFormer: (a) bi-directional gated recurrent unit (BiGRU) estimation part and (b) space-time transformer forecasting part. CNN represents the convolutional neural network layer.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{w_Q Q w_K K}{\sqrt{C}}\right) w_V V \quad (3)$$

A detailed description of the attention can be found in the study conducted by Vaswani et al. [50].

The transformer is built through a hierarchical encoder–decoder attention structure (Fig. 2b). The encoder is responsible for processing the input sequence and creating a representation, while the decoder uses this representation to generate the output sequence. In the encoder, the input tensor first passes through CNN downsample blocks with a 2D convolutional layer (Conv  $3 \times 3$ ) featuring a  $3 \times 3$  kernel size. The downsample scale is set to 2. Subsequently, the resized tensors are encoded by position embedding before being fed into the cuboid attention block, which is the main mechanism of this space-time transformer. Each cuboid attention block contains several attention layers, the number of which depends on the chosen cuboid pattern for decomposing the input tensor. The cuboid decomposition is crucial to the transformer applied in earth science spatiotemporal studies, as it significantly reduces model complexity while still capturing various types of correlations [15]. In this study, we adopted the divided space–time pattern, which splits cuboids based on spatial and temporal axes separately, resulting in two attention layers per block. For each attention layer, the decomposed sequence is flattened, and multi-head self-attention is applied in parallel. The number of heads is set to 4. In the encoder, self-attention is applied where Q, K, and V are identical to the input tensor. The attention output maintains the same shape as each input tensor. Following each attention module, a feedforward network comprising two linear layers and a Gaussian error linear unit activation layer [21] is interspersed. The output from each cuboid attention block is combined with a residual connection. After four attention blocks, downscaling layers are introduced to reduce the tensor shape, thereby enhancing the encoder's efficiency. The PatchMerge function is utilized to achieve this downscaling [15].

The decoder mirrors the encoder process. It replaces self-attention with cross-attention, where Q is the input tensor, K and V are derived from the memory of the encoder's corresponding hierarchy. Upscaling layers are incorporated into the decoder, employing nearest neighbor interpolation. This sequence of attention blocks and scaling operations is repeated twice. The final layer's output is the CSR. We multiplied it by  $E_{max}$  to produce the ultimate GHI values for validation in Section 4.2. Detailed model configurations are presented in Table 2.

20 % of the training data is used as validation data. The mean square error of the predicted CSR for 12 time steps at each pixel of the validation data is calculated as the validation loss. The model uses the AdamW optimizer, which incorporates weight decay, to minimize the loss. Training stops when the validation loss does not decrease within 10 epochs. The model is trained with an initial learning rate of 0.001, using a cosine annealing schedule, gradient clipping at 1.0, and a warmup phase over the first 20 % of the 100 epochs.

### 3.3. Metrics

The models were evaluated using several metrics, including the coefficient of determination ( $R^2$ ), mean bias difference (Bias), relative bias (rBias), root-mean-square error (RMSE), relative root-mean-square error (rRMSE), and forecast skill (FS). rBias, rRMSE, and FS are relative results mainly used for comparison over different areas, seasons, and among different studies. RMSE and rRMSE served as the primary metrics for comparing the models. The formulas used to compute these evaluation metrics are as follows:

$$R^2 = \frac{\sigma_{R_p, R_g}}{\sigma_{R_p} \sigma_{R_g}} \quad (4)$$

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N (R_p^i - R_g^i) \quad (5)$$

**Table 2**

Detailed configuration of the space–time transformer part of the SolarFormer. Conv  $3 \times 3$  is the convolutional layer with a  $3 \times 3$  kernel. GroupNorm16 is the Group Normalization layer [55] with 16 groups. The LeakyReLU represents the Leaky Rectified Linear Unit activation function [34]. The LayerNorm represents a normalization layer. The FFN represents a feedforward network.

Encoder Block	Encoder Layer	Decoder Block	Decoder Layer
2D CNN + Downsembler	Conv $3 \times 3$ GroupNorm 16 LeakyReLU PatchMerge LayerNorm Linear		
Encoder Position Embedding	PosEmbed	Decoder Position Embedding	PosEmbed
Cuboid Attention Block x 4	LayerNorm Cuboid (8,1,1) FNN LayerNorm Cuboid (1,30,30) FNN	Cuboid Cross Attention Block x 4	LayerNorm Cuboid (8,1,1) FNN LayerNorm Cuboid (1,15,15) FNN LayerNorm CuboidCross (8,1,1) FNN LayerNorm
Downsembler	PatchMerge LayerNorm Linear	Upsampler	NearestNeighborInterp Conv $3 \times 3$
Cuboid Attention Block x 4	LayerNorm Cuboid (8,1,1) FNN LayerNorm Cuboid (1,15,15) FNN	Cuboid Cross Attention Block x 4	LayerNorm Cuboid (8,1,1) FNN LayerNorm Cuboid (1,30,30) FNN LayerNorm CuboidCross (8,1,1) FNN LayerNorm NearestNeighborInterp
		2D CNN + Upsampler	Conv $3 \times 3$ GroupNorm 16 LeakyReLU Linear

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (R_p^i - R_g^i)^2} \quad (6)$$

$$rBias = \text{Bias} / \overline{R_g} \quad (7)$$

$$rRMSE = RMSE / \overline{R_g} \quad (8)$$

where  $\sigma_{R_p}$  denotes the standard deviation of the predicted values for the test samples, and  $\sigma_{R_g}$  represents the standard deviation of the ground-truth measurements for the same samples. Furthermore,  $\sigma_{R_p, R_g}$  represents the covariance of the predicted values and ground-truth measurements for the testing samples. For a given observation,  $R_p^i$  and  $R_g^i$  denote the predicted value and the corresponding ground-truth measurement, respectively. The mean value of the ground-truth measurements for the test samples is denoted by  $\overline{R_g}$ , and N is the total number of testing samples.

Additionally, we constructed the smart persistence forecasting model (PFM) as the baseline model and calculated the forecast skill (FS) relative to this baseline. The PFM presupposes that the cloud conditions at the last input time,  $x_T$ , remain static throughout the forecast lead time, and that the predicted GHI ( $R^{PFM}$ ) is influenced solely by  $E_{max}$ . Yang [58] used a combination of PFM and climatology observations to determine

the upper boundary of RMSE for GHI forecasting. Following this study, given the limited impact of climatology data within a 3-h lead time, PFM serves as the primary reference to determine the acceptable values for RMSE. Also, given its widespread use as a baseline in other existing studies, we use PFM as the baseline for this study and in calculating the FS [38,43,44]. The FS represents the predictability of the proposed model in comparison to the PFM. A positive FS indicates that the SolarFormer outperforms the baseline model, and *vice versa*.

$$R_{T+i}^{PFM} = X_T * E_{max,T+i} \quad (9)$$

$$FS = 1 - \frac{RMSE}{RMSE_{PFM}} \quad (10)$$

### 3.4. Scalability

In this study, we train and test the results over seven study areas as a demonstration. In practical use, SolarFormer can be applied to all areas covered by GOES-16 and Himawari-8, which encompass several continents including North America, South America, Asia, and Oceania. The only dynamic input for SolarFormer is the 15 bands shared by both GOES-16 and Himawari-8. For Himawari-8, interpolating the 10-min interval satellite observations to 15-min intervals is the only additional procedure required, with the rest of the data preprocessing methods being the same as those introduced in this study.

Moreover, since both the input and output are satellite data, the model is not limited by the data scarcity issues that many deep learning models face, even when applied on a large scale. Using just two years of data for training has already shown good results without overfitting. Practically, for each area, more than four years of data can be used for training, and as more data becomes available, the dataset size continues to increase. Hence, this model can be easily scaled to larger areas.

## 4. Results and discussion

In this section, we conducted a comprehensive validation and analysis of the proposed model. We first validate the CRS at each pixel of the image against the values generated by Bi-GRU in section 4.1. By assuming Bi-GRU retrieved CRS as the ground truth, we can assess the forecasting capability of the transformer component. Higher FS and  $R^2$  is present than existing studies [56,63] and deep learning models used for image-scale forecasting, including widely used ConvLSTM and other transformer architectures [15,22,31,38]. To reveal its real-world accuracy and be fairly compared with the existing studies, we further

validate the forecasted GHI with the ground measurement at seven SURFRAD stations in section 4.2. The retrieved RMSE or rRMSE is lower than reported by the NWP models over the same sites but in different years [62], and comparable or lower than existing studies using site-based inputs but over different areas [8,43,44]. Additionally, we analyze the computational and memory efficiency of SolarFormer and ConvLSTM in section 4.3. The interference time of both models over CONUS is significantly faster than High-Resolution Rapid Refresh (HRRR) model. Among them, SolarFormer presents lower FLOPs than ConvLSTM. As the forecasting lead time increases, the superiority of SolarFormer in terms of FLOPs and parameter size becomes more prominent. The limitations and future directions of this study are also discussed. This section aims to provide a valid reference for real-world applications. The model's ability to operate fast, over large spatial areas, and with high accuracy makes it promising to combine with energy-dispatching models for future energy distribution and site selection decisions.

### 4.1. Image scale validation

Fig. 3 shows a comparison between the SolarFormer's predictions and the input images for the GWN site on January 13, 2018. The input timeframe spans from 9:00 AM to 10:45 AM, while the output timeframe extends from 11:00 AM to 1:45 PM. Overall, the SolarFormer's predicted images display a solar radiation pattern that is similar to the target over the 3-h forecast, capturing the evolution of solar radiation due to cloud dispersion. For example, the disappearance of low GHI on the right side of the image is evident. However, some detailed patterns are inaccurately predicted, particularly for longer forecast durations; the longer the forecast lead time, the more indistinct the images become.

We further assessed the results by validating the normalized radiation (CSR values) pixel by pixel within the images. Validation was conducted from a 15-min to a 3-h lead time. We also produced images using a baseline model and compared them in this study (Fig. 4). The SolarFormer outperformed the baseline model at all steps, particularly at longer lead times. The  $R^2$  of the SolarFormer ranged from 0.992 to 0.744 from the 15 min to 3 h. This statistic for the SolarFormer's CSR validation is significantly better than that predicted by PredRNN++ [56]. Notably, our results were validated against the BiGRU-generated results, which may diverge from the ground truth, and were validated at 15-min intervals, whose statistics could be better when aggregated to 1 h [28]. The RMSE (rRMSE) of the SolarFormer ranged from 0.02 to 0.11 (3.2 to 18.3 %), while for the baseline model it ranged from 0.03 to

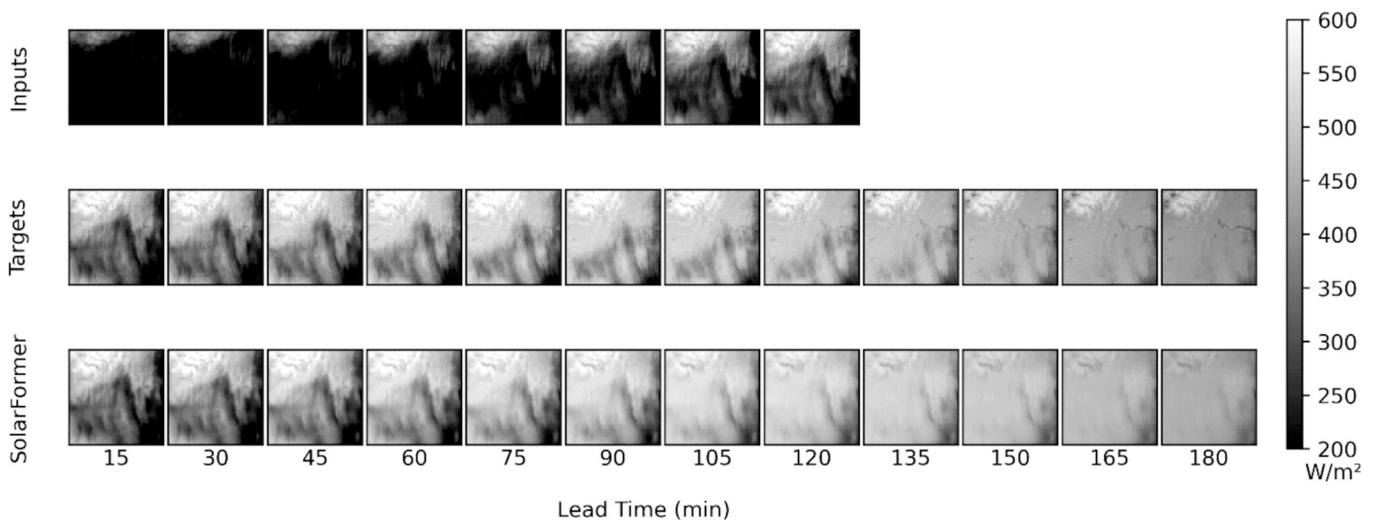


Fig. 3. One testing sample of the input, target, and SolarFormer predicted values over the GWN site on January 13, 2018. The inputs are BiGRU generated GHI images ranges from 9:00 AM to 10:45 AM and the outputs are BiGRU generated GHI images ranges from 11:00 AM to 1:45 PM.

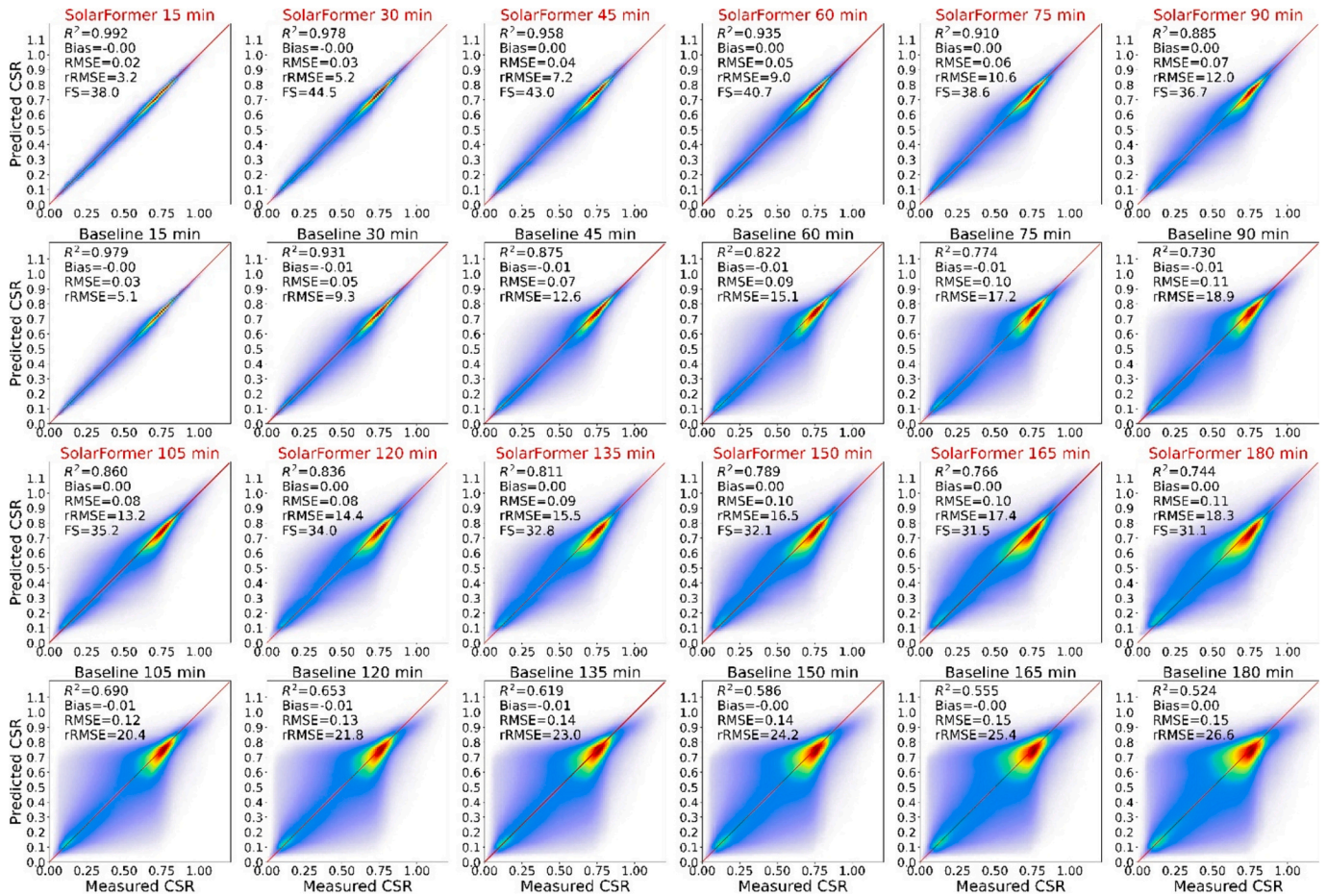


Fig. 4. The validation results of clear sky ratio (CSR) at each pixel within one image for both the SolarFormer and baseline models.  $R^2$ : coefficient of determination; Bias: mean bias difference RMSE: root-mean-square error; rRMSE: relative root-mean-square; FS: forecast skill.

0.15 (5.1 to 26.6 %). The RMSE (rRMSE) of both methods increased as the forecasting lead time increased, but the SolarFormer exhibited a significantly slower rate of increase. The SolarFormer maintained an rRMSE lower than 20 % when forecasting 3 h ahead, while the baseline model's rRMSE was 10 % higher. The FS of the SolarFormer ranged from 31 to 40 %, depending on the lead time. These results indicate the strong

predictability of the transformer and suggest that if we can achieve accurate real-time estimation of solar radiation, we can attain these rRMSE and FS levels in solar power predictions.

Fig. 5 compares the validation results at image scale with ConvLSTM [38] using the same input and training data. Both the transformer and ConvLSTM show lower rRMSE for all lead times. Among them, the

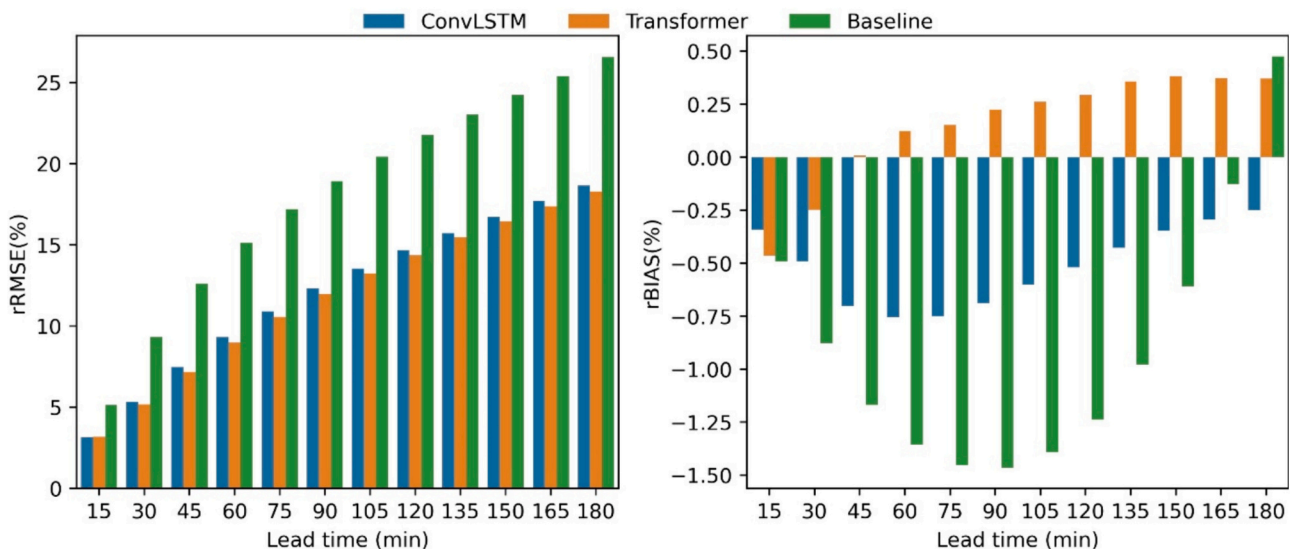


Fig. 5. The comparison of the validation results of clear sky ratio (CSR) at each pixel between Transformer, ConvLSTM, and Baseline models.

transformer presents lower rRMSE after the first 15 min. In terms of rBIAS, the transformer shows lower absolute values overall compared with the baseline and ConvLSTM. However, SolarFormer exhibits an accumulated bias as the forecasting lead time increases, whereas the other models exhibit the most bias in the middle of the forecast period and show some relief as the lead time extends further.

#### 4.2. In situ validation

The study also validated the SolarFormer's predictions with ground measurements from seven SURFRAD sites to assess the model's overall efficacy and compared it with the baseline model (Fig. 6). We first validated the BiGRU-retrieved GHI estimation at the last input time-frame, shown as 0 min. The estimated RMSE (rRMSE) is 92.1 W/m<sup>2</sup> (15.9 %), which aligns with the accuracy reported by Li et al. [26], representing the systematic uncertainties introduced by the estimation methods. Unlike the image-scale validation, the baseline model driven with *in situ* measurements exhibits high accuracy in the first 30 min, surpassing the estimation accuracy within the initial 15 min. This can be attributed to the relatively stable atmospheric conditions in such a short

period combined with the spatiotemporal representativeness issues of point-based ground measurements compared to satellite pixels as well as the systematic uncertainties introduced by the estimation part. After 30 min, the SolarFormer demonstrates superior performance. The overall RMSE (rRMSE) of the SolarFormer for the 3 h prediction is within 150 W/m<sup>2</sup> (30 %), compared to the baseline model, which achieved 198 W/m<sup>2</sup> (38 %). Regarding FS, the most significant increase occurs within the first hour, after which it largely stabilizes at approximately 25 % regardless of lead time, further demonstrating the transformer's advantages in handling long-time series data. However, SolarFormer exhibits increasing bias as the forecasting lead time extends, explained by the inevitable error accumulation in long time-series projections inherent to data-driven methods.

We then assessed the results and depicted the rRMSE and FS for each SURFRAD station, along with their mean values, spanning time up to 3 h at 15-min intervals (Fig. 7). Among all the sites, DRA records the lowest rRMSE but also the lowest FS (Fig. 7). This is attributable to the DRA sites being located in Nevada's arid desert, where atmospheric conditions are comparatively stable and clear. Consequently, both the SolarFormer and the baseline models can achieve high prediction accuracy,

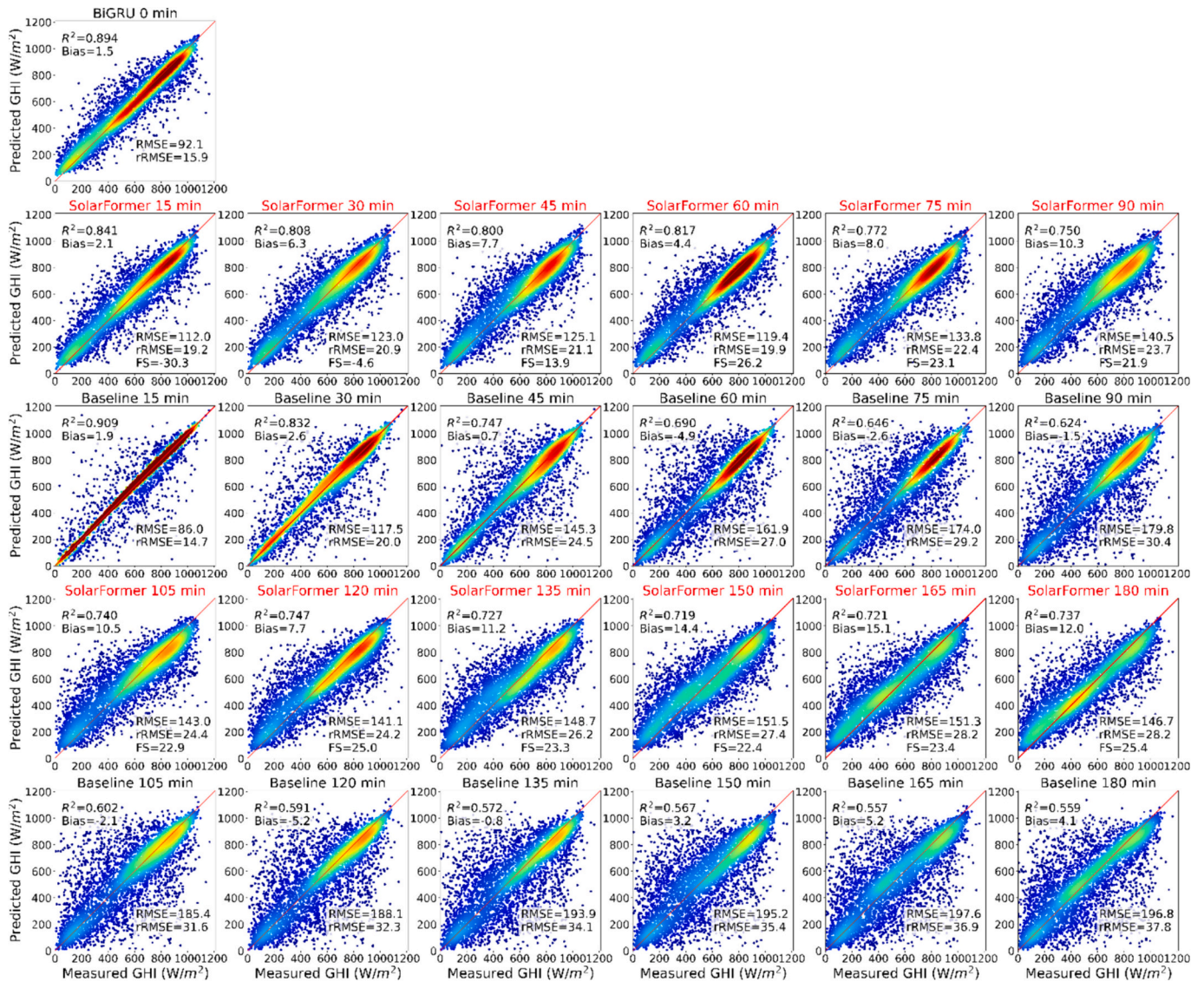


Fig. 6. The validation results of GHI over seven surface radiation budget observing network (SURFRAD) sites for both the SolarFormer and baseline models. 0 min represents the last input time frame, which we validate over bi-directional gated recurrent unit (BiGRU) generated GHI values. R<sup>2</sup>: coefficient of determination; Bias: mean bias difference RMSE: root-mean-square error; rRMSE: relative root-mean-square; FS: forecast skill.



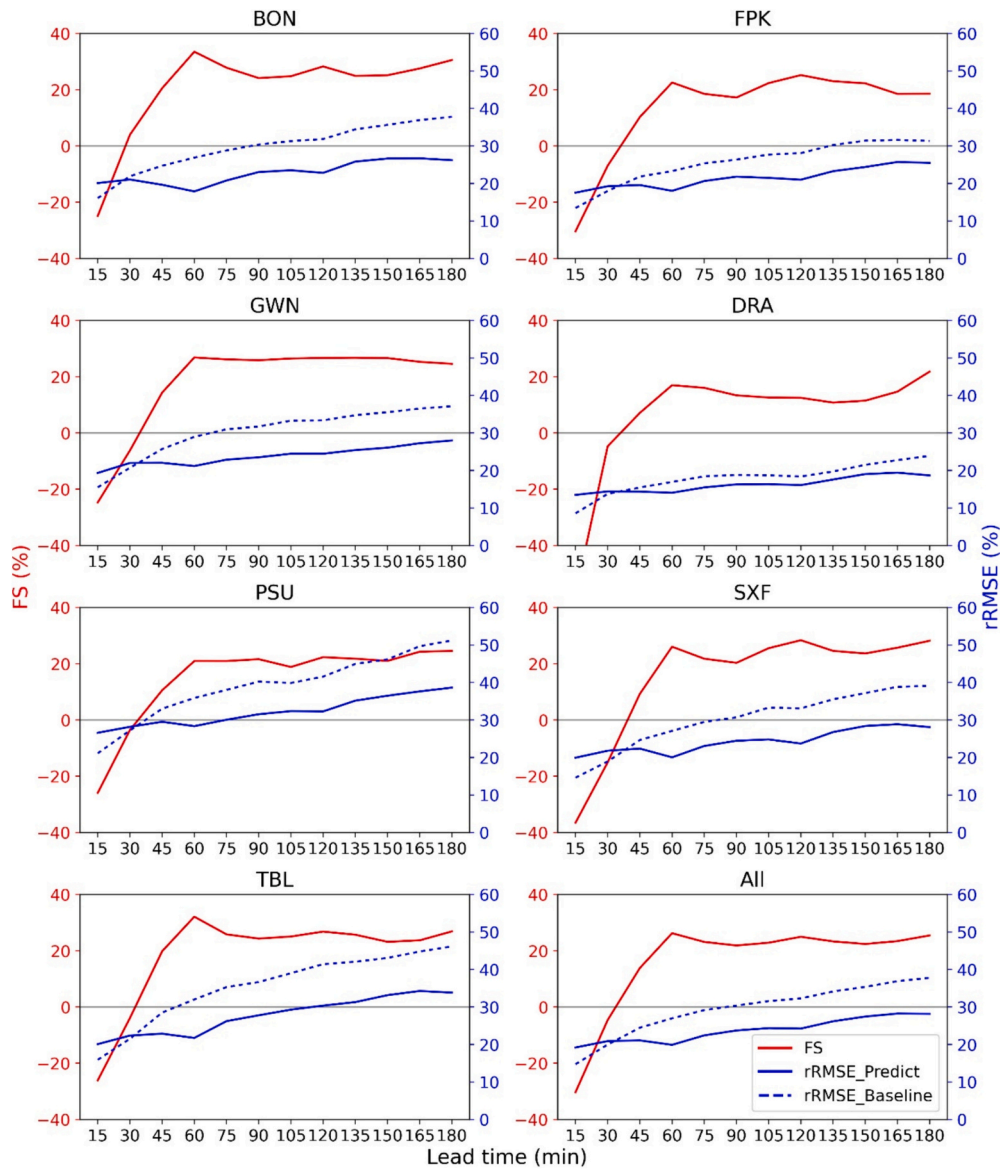


Fig. 7. The relative root-mean-square (rRMSE) and forecast skill (FS) values over seven surface radiation budget observing network (SURFRAD) stations and all combined data.

but the advantages of the Solarformer in dealing with dynamic cloud movement are also diminished in these regions. For the other sites, although the prediction rRMSE may vary slightly, the FS is similar, demonstrating the stability of the Solarformer across different climate types.

To facilitate comparison with existing studies, we aggregated the SolarFormer predictions to hourly granularity across all sites and compared them with the baseline model (Fig. 8). For the baseline model, the RMSE (rRMSE) values for the first, second, and third-hour forecasts are  $124.5 \text{ W/m}^2$  (19.9 %),  $159.1 \text{ W/m}^2$  (26.5 %), and  $170.5 \text{ W/m}^2$  (31.9 %), respectively. In contrast, the SolarFormer model exhibits improved RMSE (rRMSE) values of  $93.8 \text{ W/m}^2$  (15.0 %),  $118.9 \text{ W/m}^2$  (19.8 %), and  $129.1 \text{ W/m}^2$  (24.2 %) for the corresponding forecast durations. The FS for the SolarFormer remains consistent at approximately 25 % for all three forecast hours. The  $R^2$  values are 0.88, 0.8, and 0.78, respectively. Table 3 presents the corresponding metrics for each station. The RMSE for most sites, for both 1-h and 3-h forecasts, is within the upper and lower bounds defined by Yang [58], highlighting the effectiveness of the proposed models. Moreover, these hourly results exceeded the best outcomes previously reported in the state-of-art hourly updated NWP

modeling [62] and comparable with those for SolarAnywhere v2 in earlier research [41]. The  $R^2$  values are higher than those predicted by PredRNN++ over China at all lead times [56]. It is important to note that the results have not been compared with the identical datasets.

We conducted further analysis to understand whether the forecasting accuracy of the SolarFormer is influenced by its initial conditions, namely the estimation accuracy of the BiGRU at the last input time frame. We calculated the rRMSE across all lead times and the absolute error (AE) at the initial condition for each prediction, then aggregated the prediction rRMSE with the initial AE (Fig. 9). Our analysis revealed a significant correlation between the prediction rRMSE and the initial AE. A larger initial AE tends to result in a higher rRMSE and a greater standard deviation in the rRMSE of the SolarFormer predictions. Notably, the upper error bar of the prediction rRMSE corresponds to the highest value within each initial AE bin. This suggests that achieving superior accuracy at the initial condition can lead to more accurate forecasting outcomes.

The study further investigates the SolarFormer's performance under varying cloud conditions. We categorized the results based on the cloud mask's consistency, as depicted in Fig. 10. In cases where the cloud mask

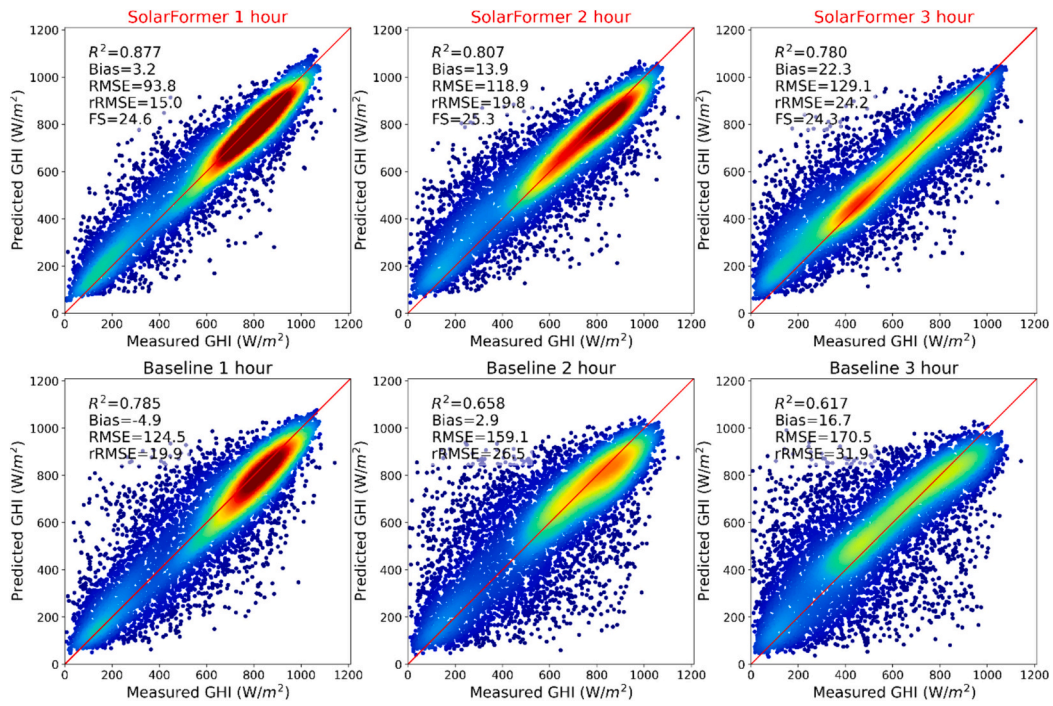


Fig. 8. The validation results on all ground truth for both the baseline and SolarFormer models at hourly granularity.  $R^2$ : coefficient of determination; Bias: mean bias difference RMSE: root-mean-square error; rRMSE: relative root-mean-square; FS: forecast skill.

Table 3  
Summary of the SolarFormer prediction results at hourly scale over seven SURFRAD sites.

	FS	1 h			2 h			3 h				
		FS	$R^2$	RMSE	rRMSE	FS	$R^2$	RMSE	rRMSE	FS	$R^2$	RMSE
	(%)			( $W/m^2$ )	(%)	(%)	( $W/m^2$ )	(%)	(%)	(%)	( $W/m^2$ )	(%)
ALL	24.63	0.88	93.82	14.99	25.26	0.81	118.87	19.80	24.27	0.78	129.10	24.18
BON	32.50	0.91	77.73	12.98	32.46	0.85	101.39	17.56	28.65	0.82	114.60	22.21
DRA	22.99	0.83	86.68	10.72	25.74	0.78	101.42	13.15	23.83	0.81	107.54	15.77
FPK	22.26	0.88	88.81	14.67	22.06	0.83	108.74	18.87	19.84	0.81	116.86	22.68
GWN	26.67	0.89	87.78	14.12	24.47	0.80	119.30	19.51	23.73	0.76	130.43	23.46
PSU	16.78	0.84	110.60	21.39	20.44	0.75	137.74	26.91	23.34	0.72	141.94	30.43
SXF	30.31	0.90	86.56	15.22	32.34	0.85	106.67	19.89	33.48	0.83	113.92	24.36
TBL	23.97	0.83	112.75	17.38	22.54	0.73	147.74	24.42	19.49	0.68	167.40	32.04

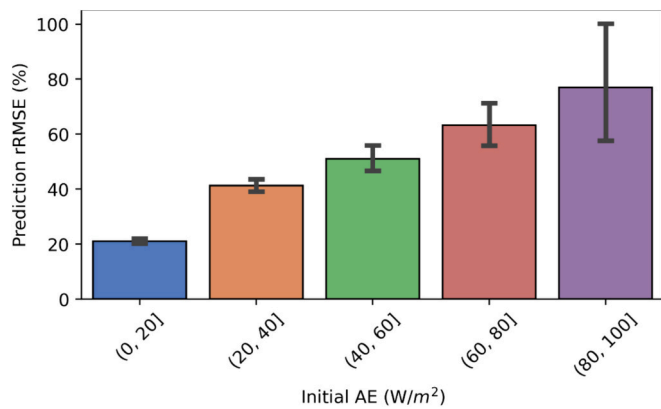


Fig. 9. The change of prediction relative root-mean-square (rRMSE, %) with different absolute error (AE) ranges at the initial condition. The error bar represents the standard deviation of the rRMSE.

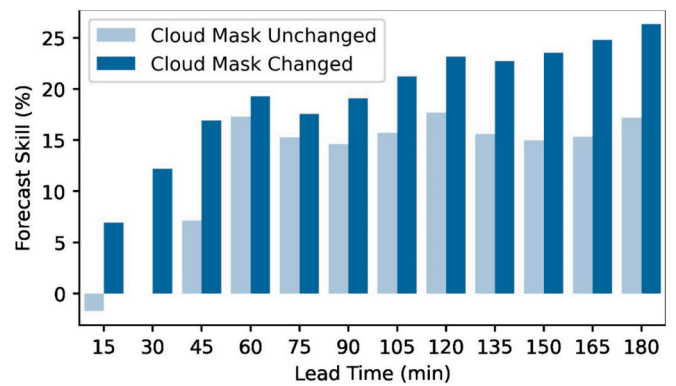


Fig. 10. The forecast skill (FS) of the SolarFormer at different lead time with cloud mask changed or unchanged.

was unchanged, the FS for the first 30 min was at or below 0. This suggests that the SolarFormer's performance does not surpass the baseline model within a 30-min forecast window under stable atmospheric conditions. Beyond the 1-h mark, the FS becomes positive and stabilizes

around 15 %. Conversely, when the cloud mask changed, the FS consistently exceeded 0 and improved with longer forecast lead times. This pattern emphasizes the SolarFormer's superior performance over the baseline model in dynamic cloud situations, especially as the forecast period lengthens. Additionally, the FS for changing cloud masks consistently surpassed that for stable masks at all intervals. These findings not only highlight the SolarFormer's effectiveness during significant atmospheric changes but also indicate its greater proficiency in managing variable conditions.

We also compare the forecasting rRMSE and FS of the SolarFormer in different months (Fig. 11). The SolarFormer's forecasting capabilities are robust throughout the year, with the rRMSE remaining consistent at approximately 25 %. The Forecast Skill (FS) exhibits significant variability across different months. It is observed to be higher during the winter months of November, December, and January, which can be attributed to the SolarFormer's enhanced ability to better handle complex cloud movements and varying irradiance compared with the PFM. Conversely, the FS is lower in February, April, May, and September, which could suggest that the SolarFormer's improvement is less pronounced during these months. The enhanced performance of the SolarFormer in winter could be beneficial for grid stability and energy planning in winter when the demand for heating increases.

#### 4.3. Analysis and evaluation of the SolarFormer components and forecasting potentials

##### 4.3.1. Ablation study of model structure

In our study, we tested various model structures to identify the optimal configuration for the space–time transformer when applied to solar forecasting. We assessed the model in terms of accuracy, measured by rRMSE, and model size, measured by the number of parameters (#Para) and floating point operations (FLOPs), across different cuboid attention patterns, global vectors, and downsample scales (Table 4). As delineated by Gao et al. [15], cuboid types determine the design of the elementary attention layer, which is instrumental in calculating data correlation for efficiently and effectively extracting space–time information for forecasting. Our analysis compared several pioneering space–time attention types: divided space–time attention (divided\_st), which separates attention along temporal and spatial dimensions [3]; axial attention, which segments along the tensor's temporal, height, and width dimensions [22]; video shifted window (video\_swin), which specifies a non-overlapping window size of  $2 \times 4 \times 4$ -sized patches as one attention unit [28]; and spatial local dilate attention (spatial\_lg), which segments tensors along temporal dimensions and  $4 \times 4$  patches along spatial dimensions with dilated aggregation approaches [15]. We also compared the effects of including or excluding the global vector, which links various cuboids and gathers global information [15]. In

**Table 4**

Summary of accuracy and complexity of different model structures. The chosen model configuration is underlined.

Model Structure		rRMSE (%)	FLOPs	#Para
Cuboid Attention Patterns	<u>divided_st</u>	12.7	22.95 G	4.81 M
	spatial_lg	13.5	30.15 G	6.59 M
	video_swin	13.4	23.14 G	4.81 M
	axial	13.4	29.55 G	6.59 M
Global Vector	<u>0</u>	12.7	22.95 G	4.81 M
	8	12.8	22.96 G	5.48 M
Downsample Scales	<u>2</u>	12.7	22.95 G	4.81 M
	3	13.0	13.70 G	4.87 M
	4	13.3	11.40 G	4.93 M

contrast to the Earthformer, which was originally designed for precipitation forecasting and is best paired with an axial cuboid type and eight global vectors, our results indicated that the divided\_st type without a global vector yielded the highest accuracy with the lowest model complexity for solar forecasting. Downsample scales represent preliminary steps that reduce the spatial resolution of the input. Our results show that a lower downsample scale leads to higher accuracy, but increasing the number of downsample layers tends to decrease accuracy while significantly reducing FLOPs. Future studies should thoughtfully consider these trade-offs when determining the choice of convolutional layers.

##### 4.3.2. Long-term forecasting potential of the SolarFormer

We here evaluate the SolarFormer's capability to forecast over extended periods and identifies factors influencing its accuracy. Fig. 12 depicts how varying combinations of input and output frames, along with spatial dimensions, affect performance. Longer output frames are correlated with decreased prediction accuracy but exhibit consistence accuracy across all forecasting lead times. The model appears to trade off short-term forecasting accuracy to maintain overall precision, particularly for longer lead times. Additionally, a larger spatial size can enhance accuracy for forecasts with longer lead times. However, for models forecasting only three hours ahead, a larger input spatial size has the opposite effect. Furthermore, longer input time frames lead to higher accuracy with smaller spatial sizes and at longer forecasting lead times. The results indicate that the SolarFormer consistently maintains its forecasting capability over longer durations without a significant accuracy decline. For short-term forecasts of up to 3 h, the configuration used in this study is optimal. For longer forecasting horizons, increasing both the input frames and spatial size can improve reliability.

The SolarFormer not only offers improved prediction accuracy but

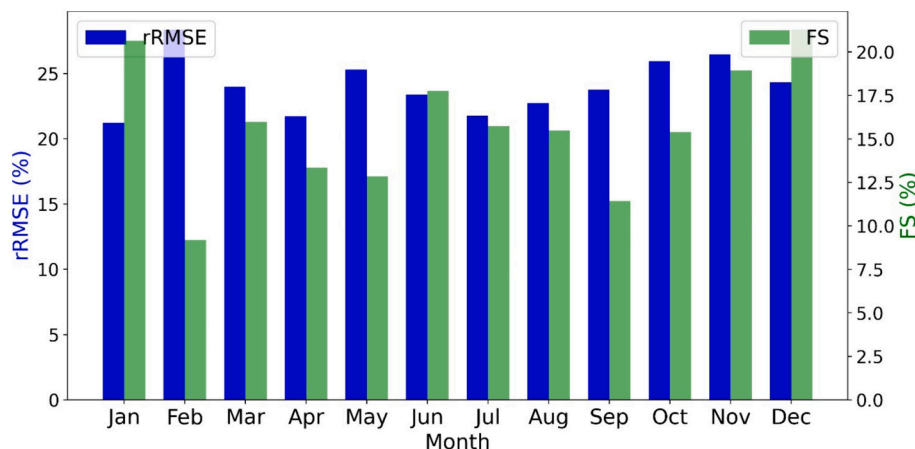


Fig. 11. The forecast skill (FS) and rRMSE of the SolarFormer in different months.

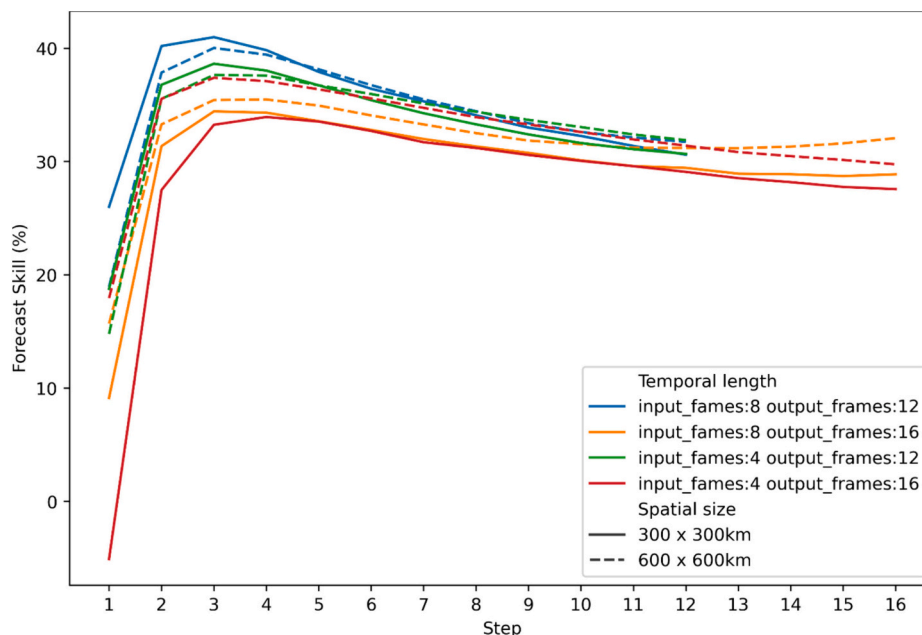


Fig. 12. Ablation study of the SolarFormer for input/output time frames and spatial sizes.

Table 5

Comparison of the accuracy and complexity of the SolarFormer and ConvLSTM for different input/output frames and spatial dimensions.

Input frames	Output frames	Spatial size	SolarFormer				ConvLSTM			
			FLOPs (G)	#Para (M)	rRMSE (%)	FS (%)	FLOPs (G)	#Para (M)	rRMSE (%)	FS (%)
8	12	300	23	4.8	12.7	34.2	106	2.6	13.0	32.8
8	16	300	28	4.8	15.6	30.1	131	2.6	15.4	31.0
4	12	300	19	4.8	12.9	34.2	90	2.6	13.3	32.4
4	16	300	24	4.8	15.1	31.8	116	2.6	15.4	30.6
8	12	600	114	6.6	12.7	34.2	1697	10.6	-	-
8	16	600	142	6.6	15.1	32.4	2101	10.6	-	-
4	12	600	96	6.6	13.0	33.9	1455	10.6	-	-
4	16	600	121	6.6	15.0	32.6	1860	10.6	-	-

also boasts reduced computational demands and memory consumption (Table 5). To demonstrate this, we compared the FLOPs and #Para across various input frames, output frames, and input image spatial dimensions for the SolarFormer and ConvLSTM [47]. With the spatial size set at 300 km, the SolarFormer consistently requires fewer FLOPs, although it has a larger parameter count. Despite the increased number of parameters, the SolarFormer attains a lower rRMSE and a higher FS in nearly all tested scenarios. When the input spatial dimension is expanded to 600 km, both the FLOPs and parameter count for the SolarFormer increase, yet they remain within the processing capabilities of a single 16GB NVIDIA V100 GPU. In contrast, the FLOPs and parameters for ConvLSTM increase exponentially, exceeding the memory capacity of a single GPU. As Fig. 11 emphasizes, future forecasting efforts, particularly for longer lead times, will require the expansion of both input timeframes and image dimensions. In this context, the SolarFormer stands out as a promising option, demonstrating considerable potential for long-term forecasting.

#### 4.4. Limitations and future directions

Though geostationary satellite data is a unique data source for solar forecasting due to its spatial information of clouds, some limitations arise from using solely satellite data. First, visible bands become invalid during nighttime, rendering SolarFormer unable to predict early morning GHI as the input data is unavailable. This also poses challenges

when extending the method to predict day-ahead GHI. Moreover, satellite observations suffer from pixel-point mismatch due to their coarse resolution, leading to unavoidable systematic uncertainties [23,29]. Lastly, the latency of nowcasting in real-world applications cannot be less than the release time of the satellite observations.

The proposed algorithm shows promise but could benefit from further enhancements. As shown in Fig. 9, the initial conditions, namely the estimation accuracy of the BiGRU at the last input time frame, will significantly influence the overall forecasting accuracy. If we can retrieve highly accurate GHI at image scale at the initial conditions, the forecasting part of the SolarFormer can achieve outstanding performance as shown in sections 4.1. This suggests that improving the near-real-time estimation part can be one way to enhance forecasting accuracies further.

Another notable issue is the accumulation of bias as the forecasting lead time extends compared with the baseline model shown in Fig. 5. The biased results will lead to biased decision in energy management. This error accumulation issue is common in time-series projections inherited in data-driven methods [7]. Future studies can adopt multi-step forecasting methods such as direct strategies and multiple-input multiple-output methods to mitigate these errors [49,53].

Furthermore, future models should enable probabilistic forecasting rather than determinate results to provide more practical meaningful results. This has been achieved by precipitation nowcasting studies [12] and a recent study in day-ahead solar radiation forecasting [4]. Adding

components like multi-quantiles in machine learning models can provide prediction intervals, that can assist more in decision making.

Additionally, as discussed in section 4.3.2, the SolarFormer model holds significant potential in long-term forecasting. However, several challenges remain. Firstly, regarding the input data, incorporating large spatial images and longer input time frames is necessary (Fig. 12). Satellite visible bands will be invalid during nighttime, so appropriately handling these invalid values and effectively using continuous bands like infrared bands for forecasting will be important. Furthermore, the increasing input data size will also increase the data storage burden. Moving the entire computation process to online platforms like Google Earth Engine would further enhance the scalability of this study.

## 5. Conclusions

Accurate, near-real-time nowcasting of GHI with high temporal updates at the regional scale is crucial for cost-effective and sustainable management of regional energy systems. However, existing studies either rely on ground-measured meteorological data, which are lack of spatial cloud information and limited to areas with ground measurements, or they rely on specific or additional inputs with computationally heavy models that cannot be easily scaled both spatially and temporally. In this study, we developed the SolarFormer model to address these gaps.

The model combines a BiGRU with a space-time transformer. It takes geostationary satellite observations from the past two hours as dynamic inputs only, enabling near-real-time forecasting up to 3 h at 15-min intervals. SolarFormer eliminates the requirement for ground measurement data as inputs and hence addresses data scarcity issues that hinder many machine-learning models in large-scale deployments. Consequently, the model can be applied directly in areas covered by both AHI and ABI. These features make the SolarFormer particularly advantageous for the dispatch modeling of both macro- and micro-energy systems. Due to its computational efficiency, the model can be operated on a single 16GB NVIDIA V100 GPU.

The SolarFormer model has been rigorously validated with satellite-derived CSR and ground-measured GHI with metrics including  $R^2$ , Bias (rBias), RMSE (rRMSE), and FS using PFM as the baseline model. At the image scale, the forecasted GHI is compared with BiGRU-generated CSR at each pixel for each timeframe at 15-min intervals. The rRMSE ranges from 3.2 % to 18.3 %, and the FS ranges from 31.1 to 44.5 (%). When validated against ground-measured GHI over seven SURFRAD stations at 15-min intervals, the rRMSE ranges from 19.2 % to 28.2 %. Hourly validation of RMSE (rRMSE) at these stations yields results of 93.8 W/m<sup>2</sup> (15.0 %), 118.9 W/m<sup>2</sup> (19.8 %), and 129.1 W/m<sup>2</sup> (24.2 %) for 1-h, 2-h, and 3-h lead times, respectively. Forecast skills were recorded at 24.6, 25.3, and 24.3 (%) for the corresponding intervals. The validation accuracy surpasses not only the smart persistent model, but also those reported from existing hourly updated NWP methods such as the National Oceanic and Atmospheric Administration's Rapid Refresh (RAP), HRRR, and post-processed HRRR. Moreover, the model shows lower rRMSE and relative rBias than ConvLSTM with the same input data.

Our analysis underscores the critical role of BiGRU's estimation accuracy in enhancing forecasting precision as shown in Fig. 9. Higher accuracy in initial conditions invariably leads to improved forecasting outcomes, a dimension that merits further investigation. Furthermore, the SolarFormer's performance is outstanding under highly variable cloudy conditions, especially when the lead time extends, with FS increasing from 7 % to 25 %, as shown in Fig. 10. Under stable atmospheric conditions beyond the initial 30 min, SolarFormer also presents high and stable FS ranging from 15 to 20 %. Moreover, SolarFormer achieves stable forecasting accuracy throughout the year and enhanced performance in winter compared with the baseline as shown in Fig. 11. This would be beneficial for grid stability and energy planning in winter when the demand for heating increases.

The study reveals the potential of SolarFormer to forecast over

extended periods. The findings indicate that the SolarFormer reliably sustains its forecasting accuracy over longer durations without significant deterioration (Fig. 12). For longer time series forecasting, incorporating broader spatial dimensions and longer input time frames can enhance predictive precision. Comparative analysis also demonstrates SolarFormer's superior performance relative to ConvLSTM regarding accuracy, computational requirements, and memory efficiency at different forecasting lead time (Table 5). The SolarFormer's efficient architecture enhances its prospects, positioning it as a viable option for longer forecasting horizons in the future, which will benefit applications in transmission and distribution scheduling, as well as power market bidding and clearing [2].

## CRedit authorship contribution statement

**Ruohan Li:** Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Dongdong Wang:** Writing – review & editing, Supervision, Conceptualization. **Zhihao Wang:** Writing – review & editing, Software. **Shunlin Liang:** Writing – review & editing. **Zhanqing Li:** Writing – review & editing. **Yiqun Xie:** Writing – review & editing. **Jiena He:** Writing – review & editing.

## Declaration of competing interest

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

## Data availability

Data will be made available on request.

## References

- [1] Al Garni HZ, Awasthi A. Solar PV power plants site selection: a review. *Advances in renewable energies and power technologies* 2018;57–75.
- [2] Behera MK, Majumder I, Nayak N. Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Engineering Science and Technology, an International Journal* 2018;21(3):428–38.
- [3] Bertasius G, Wang H, Torresani L. Is space-time attention all you need for video understanding? In *ICML 2021*;2(3):4.
- [4] Boussif O, Boukachab G, Assouline D, Massaroli S, Yuan T, Benabbou L, et al. Improving day-ahead solar irradiance time series forecasting by leveraging Spatio-temporal context. *Advances in Neural Information Processing Systems* 2024;36.
- [5] British Petroleum (BP). *Statistical Review of World Energy*. 2021. p. 2021.
- [6] Carpentieri A, Folini D, Nerini D, Pulkkinen S, Wild M, Meyer A. Intraday probabilistic forecasts of surface solar radiation with cloud scale-dependent autoregressive advection. *Appl Energy* 2023;351:121775. <https://doi.org/10.1016/j.apenergy.2023.121775>.
- [7] Cheng H, Tan PN, Gao J, Scripps J. Multistep-ahead time series prediction. In: *In Advances in Knowledge Discovery and Data Mining: 10th Pacific-Asia Conference, PAKDD 2006, Singapore, April 9–12, 2006. Proceedings* 10. Berlin Heidelberg: Springer; 2006. p. 765–74.
- [8] Choi M, Rachunok B, Nateghi R. Short-term solar irradiance forecasting using convolutional neural networks and cloud imagery. *Environ Res Lett* 2021;16(4):044045.
- [9] Chow CW, Urquhart B, Lave M, Dominguez A, Kleissl J, Shields J, et al. Intra-hour forecasting with a total sky imager at the UC san Diego solar energy testbed. *Solar Energy* 2011;85(11):2881–93.
- [11] Earth Resources Observation And Science (EROS) Center. Global 30 Arc-Second Elevation (GTOPO30), U.S. 465 Geological Survey. 2017. <https://doi.org/10.5066/F7DF6PQS>.
- [12] Espeholt L, Agrawal S, Sønderby C, Kumar M, Heek J, Bromberg C, et al. Deep learning for twelve hour precipitation forecasts. *Nat Commun* 2022;13(1):1–10.
- [13] Etxegarai G, López A, Aginako N, Rodríguez F. An analysis of different deep learning neural networks for intra-hour solar irradiation forecasting to compute solar photovoltaic generators' energy production. *Energy Sustain Dev* 2022;68: 1–17.
- [14] Fu CL, Cheng HY. Predicting solar irradiance with all-sky image features via regression. *Solar Energy* 2013;97:537–50.
- [15] Gao Z, Shi X, Wang H, Zhu Y, Wang YB, Li M, et al. Earthformer: exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems* 2022;35:25390–403.
- [16] Gallo R, Castangia M, Macii A, Macii E, Patti E, Aliberti A. Solar radiation forecasting with deep learning techniques integrating geostationary satellite images. *Eng Appl Artif Intel* 2022;116:105493.

- [17] Ghimire S, Deo RC, Downs NJ, Raj N. Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia. *J Clean Prod* 2019;216:288–310.
- [18] Germaat DE, de Boer HS, Daioglou V, Yalew SG, Müller C, van Vuuren DP. Climate change impacts on renewable energy supply. *Nature Climate Change* 2021;11(2):119–25.
- [19] Ghimire S, Deo RC, Raj N, Mi J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Appl Energy* 2019;253:113541.
- [20] Haider SA, Sajid M, Sajid H, Uddin E, Ayaz Y. Deep learning and statistical methods for short-and long-term solar irradiance forecasting for Islamabad. *Renew Energy* 2022;198:51–60.
- [21] Hendrycks D, Gimpel K. Gaussian error linear units (gelus). arXiv preprint 2016. arXiv:1606.08415.
- [22] Ho J, Kalchbrenner N, Weissenborn D, Salimans T. Axial attention in multidimensional transformers. arXiv preprint 2019. arXiv:1912.12180.
- [23] Huang G, Li Z, Li X, Liang S, Yang K, Wang D, et al. Estimating surface solar irradiance from satellites: past, present, and future perspectives. *Remote Sens Environ* 2019;233:111371.
- [24] Kosmopoulos P, Dhake H, Melita N, Tagarakis K, Georgakis A, Stefan A, et al. Multi-layer cloud motion vector forecasting for solar energy applications. *Appl Energy* 2024;353:122144. <https://doi.org/10.1016/j.apenergy.2023.122144>.
- [25] Kumari P, Toshniwal D. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Appl Energy* 2021;295:117061.
- [26] Li R, Wang D, & Liang S. (2023a). Comparison between deep learning architectures for the 1 km, 10/15-min estimation of downward shortwave radiation from AHI and ABI. *Remote Sens Environ*, 295, 113697. doi:10.1016/j.rse.2023.113697.
- [27] Li R, Wang D, Wang W, Nemani R. A GeoNEX-based high-spatiotemporal-resolution product of land surface downward shortwave radiation and photosynthetically active radiation. *Earth Syst Sci Data* 2023;15:1419–36. <https://doi.org/10.5194/essd-15-1419-2023>.
- [28] Li R, Wang D, Liang S. Comprehensive assessment of five global daily downward shortwave radiation satellite products. *Science of Remote Sensing* 2021;4:100028. <https://doi.org/10.1016/j.srs.2021.100028>.
- [29] Li Z, Whitlock CH, Charlock TP. Assessment of the global monthly mean surface insolation estimated from satellite measurements using global energy balance archive data. *J Climate* 1995;8(2):315–28.
- [30] Lin F, Zhang Y, Wang J. Recent advances in intra-hour solar forecasting: a review of ground-based sky image methods. *International Journal of Forecasting* 2023;39(1):244–65.
- [31] Liu Z, Ning J, Cao Y, Wei Y, Zhang Z, Lin S, et al. Video swin transformer. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2022:3202–11.
- [32] Lorenz E, Heinemann D, Wickramaratne H, Beyer HG, Bofinger S. Forecast of ensemble power production by grid-connected PV systems. In: Proc. 20th European PV conference, Milano, Italy; 2007.
- [33] Ma S, Xiang T, Hou K, Liu Z, Tang P, Qi N. Spatial–temporal optimal dispatch of mobile energy storage for emergency power supply. *Energy Rep* 2022;8:322–9.
- [34] Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In *proc icml* 2013, June;30(1):3.
- [35] Miller SD, Rogers MA, Haynes JM, Sengupta M, Heidinger AK. Short-term solar irradiance forecasting via satellite/model coupling. *Solar Energy* 2018;168:102–17.
- [36] Nielsen AH, Iosifidis A, Karstoft H. IrradianceNet: spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. *Solar Energy* 2021;228:659–69.
- [37] Nycander E, Morales-España G, Söder L. Power-based modelling of renewable variability in dispatch models with clustered time periods. *Renew Energy* 2022;186:944–56.
- [38] Paletta Q, Arbod G, Lasenby J. Omnivision forecasting: combining satellite and sky images for improved deterministic and probabilistic intra-hour solar energy predictions. *Appl Energy* 2023;336:120818. <https://doi.org/10.1016/j.apenergy.2023.120818>.
- [39] Perez R, Kivalov S, Schlemmer J, Hemker Jr K, Renné D, Hoff TE. Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy* 2010;84(12):2161–72.
- [40] Pérez E, Pérez J, Segarra-Tamarit J, Beltran H. A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. *Solar Energy* 2021;218:652–60.
- [41] Qin J, Jiang H, Lu N, Yao L, Zhou C. Enhancing solar PV output forecast by integrating ground and satellite observations with deep learning. *Renew Sustain Energy Rev* 2022;167:112680.
- [42] Reikard G. Predicting solar radiation at high resolutions: a comparison of time series forecasts. *Solar energy* 2009;83(3):342–9.
- [43] Sharma V, Yang D, Walsh W, Reindl T. Short term solar irradiance forecasting using a mixed wavelet neural network. *Renew Energy* 2016;90:481–92.
- [44] Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 2015;28.
- [45] Su Y, Chan LC, Shu L, Tsui KL. Real-time prediction models for output power and efficiency of grid-connected solar photovoltaic systems. *Appl Energy* 2012;93:319–26.
- [46] Taieb SB, Bontempi G, Atiya AF, Sorjamaa A. A review and comparison of strategies for multi-step ahead time series forecasting based on the NNS forecasting competition. *Expert systems with applications* 2012;39(8):7067–83.
- [47] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Advances in neural information processing systems* 2017;30.
- [48] Wang F, Lu X, Mei S, Su Y, Zhen Z, Zou Z, et al. A satellite image data based ultra-short-term solar PV power forecasting method considering cloud information from neighboring plant. *Energy* 2022;238:121946.
- [49] Wang W, Li S, Hashimoto H, Takenaka H, Higuchi A, Kalluri S, et al. An introduction to the geostationary-NASA earth exchange (GeoNEX) products: 1. Top-of-atmosphere reflectance and brightness temperature. *Remote Sens (Basel)* 2020;12(8):1267.
- [50] Wang Y, Wang R, Tanaka K, Ciaia P, Penuelas J, Balkanski Y, et al. Accelerating the energy transition towards photovoltaic and wind in China. *Nature* 2023;619(7971):761–7.
- [51] Wu Y, He K. Group normalization. In proceedings of the European conference on computer vision (ECCV) 2018:3–19.
- [52] Xia P, Zhang L, Min M, Li J, Wang Y, Yu Y, et al. Accurate nowcasting of cloud cover at solar photovoltaic plants using geostationary satellite images. *Nat Commun* 2024;15(1):510.
- [53] Yang D, Jirutitijaroen P, Walsh WM. Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy* 2012;86(12):3531–43.
- [54] Yang D. Correlogram, predictability error growth, and bounds of mean square error of solar irradiance forecasts. *Renew Sustain Energy Rev* 2022;167:112736.
- [55] Yin J, Molini A, Porporato A. Impacts of solar intermittency on future photovoltaic reliability. *Nat Commun* 2020;11(1):4781.
- [56] Ying C, Wang W, Yu J, Li Q, Yu D, Liu J. Deep learning for renewable energy forecasting: a taxonomy, and systematic literature review. *J Clean Prod* 2022;135414.
- [57] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. *Renew Sustain Energy Rev* 2022;154:111768.
- [58] Zhou Y, Liu Y, Wang D, Liu X, Wang Y. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. *Energy Convers Manage* 2021;235:113960.

### Further-reading

- [10] Dong Z, Yang D, Reindl T, Walsh WM. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy* 2013;55:1104–13.
- [36] Neuhoff K, Ritter N, Salah-Abou-El-Enien A, Vassilopoulos P. Intraday markets for power: Discretizing the continuous trading. 2016.
- [37] Newell R, Raimi D, Villanueva S, Prest B. Global energy outlook 2021: pathways from Paris. Resources for the Future 2021;8.
- [42] Perez R, Kankiewicz A, Schlemmer J, Hemker K, Kivalov S. A new operational solar resource forecast model service for PV fleet simulation. In: In 2014 IEEE 40th photovoltaic specialist conference (PVSC). IEEE; 2014. p. 0069–74.
- [54] Wang Z, Xie Y, Jia X, Ma L, Hurtt G. High-Fidelity Deep Approximation of Ecosystem Simulation over Long-Term at Large Scale. In Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems 2023:1–10.
- [61] Wu Yuxin, He Kaiming. Group normalization. In Proceedings of the European conference on computer vision (ECCV) 2018:3–19.