Energy and AI 23 (2026) 100667

Contents lists available at ScienceDirect

Energy and AI

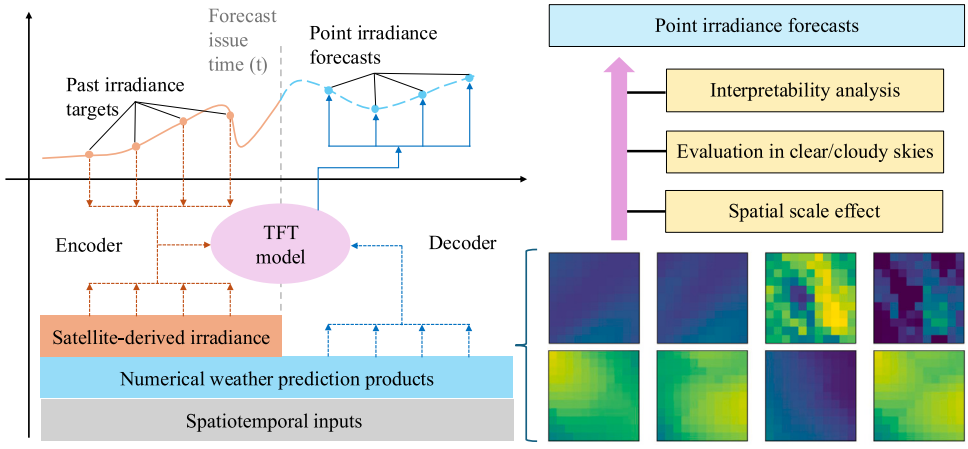
journal homepage: www.elsevier.com/locate/egyai

Interpretable transformer based intra-day solar forecasting with spatiotemporal satellite and numerical weather prediction inputs

Shanlin Chen [a,b,*], Tao Jing [c], Mengying Li [c], Hiu Hung Lee [b], Ming Chun Lam [d], Siqi Bu [b,e]

[a] Wind Engineering and Renewable Energy Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland
[b] Centre for Advances in Reliability and Safety, Hong Kong Special Administrative Region
[c] Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[d] Radiation Monitoring and Assessment Branch, Hong Kong Observatory, Hong Kong Special Administrative Region
[e] Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

GRAPHICAL ABSTRACT



HIGHLIGHTS

- Both satellite- and NWP-based data are used for solar forecasting with TFT.
- A simple yet effective method is applied to integrate spatiotemporal data into TFT.
- Spatiotemporal inputs tend to yield better solar forecasts than single-pixel data.
- Interpretability analysis is performed to help with the understanding of models.

ARTICLE INFO

Keywords:
Solar forecasting
Spatiotemporal information
Numerical weather prediction
Satellite data

ABSTRACT

With the increasing capacity addition of solar energy systems, solar forecasting is vital and cost-effective to mitigate solar variability and to support their operation. The temporal fusion transformer (TFT) has shown great potential in both solar irradiance and power output forecasting using multiple one-dimensional time series data. Since spatiotemporal information is more beneficial for solar forecasting, this work applies a simple yet effective way to incorporate two-dimensional spatiotemporal satellite- and numerical weather prediction (NWP)-based inputs with TFT for more skillful irradiance forecasts. Results show that spatiotemporal inputs

* Corresponding author.
E-mail address: shanlin.chen@epfl.ch (S. Chen).

https://doi.org/10.1016/j.egyai.2025.100667
Received 27 August 2025; Received in revised form 13 November 2025; Accepted 12 December 2025
Available online 15 December 2025

Interpretability analysis

with simple spatial averaging can generally lead to better irradiance forecasts with 4-h ahead skill scores up to 12.24%, compared to the use of single-location data. The benefit of using spatiotemporal information is more pronounced for forecasts under cloudy conditions, whereas it might result in some misrepresentations when the sky is clear or less cloudy. NWP data can generally be used to improve the intra-day solar forecasting performance with TFT, and the interpretability analysis shows that NWP irradiance products have a larger impact (up to 22.07%) on the overall results. Although NWP products are beneficial for intra-day solar forecasting when integrated with satellite-based data, their influences under different sky conditions and forecast horizons might be different. A proper analysis of these impacts should be performed and interpreted in practical applications for the reliability of energy systems. This work on improved irradiance forecasts with TFT and interpretability analysis is crucial for the operation of solar energy systems.

1. Introduction

Renewable energy sources are on track to break new records in global power generation over the predicted period of 2025–2027 [1], where solar photovoltaic (PV) is set to be the second highest worldwide clean energy source (after hydropower) of electricity production by 2027 [1]. However, solar power generation exhibits a high variability due to its intermittent nature [2,3], which greatly hinders the operation and reliability of solar energy systems [4]. As the penetration of solar energy in the electricity generation mix continues to increase, driven by the supportive policies and lowered costs [5], it is paramount to understand periods with reduced solar power generation because of weather conditions [1,6]. Ground-level solar radiation, namely, direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), and global horizontal irradiance (GHI), is the "fuel" for all solar energy technologies [7]. Therefore, it is necessary to know the availability of future solar radiation to support the operation of solar energy systems [5]. Solar forecasting herein is to provide predictions up to several days ahead to facilitate the scheduling of various power production sources and to alleviate the solar variability [5,8].

Solar forecasting is generally known as both solar irradiance forecasting and solar power output forecasting [9,10]. The former focuses on irradiance quantities, while the latter deals with the power generation of solar energy systems. Solar irradiance and power output are intrinsically linked, since irradiance is the key driver for power generation of any solar energy systems [7]. Although there are a variety of studies on direct solar power forecasting [11,12], it is suggested to use high-accuracy irradiance forecasts to produce good solar power generation predictions via the solar power curve [7,10], which is known as the two-step framework [13]. This is because irradiance forecasts should consider exploiting physical laws governing the available amount of radiation reaching the ground through atmospheric processes [13]. Although direct data-driven power predictions can also incorporate spatiotemporal inputs to capture cloud motions, the performance of statistical and machine learning methods is highly dependent on the amount and quality of the training data [14], and such power dataset is often not practically available because of proprietary and privacy reasons [15].

Given the spatiotemporal nature of surface solar irradiance, forecasting with locally sensed data is incapable of producing accurate results for longer horizons [16]. Moreover, these single-location predictions that cannot capture the spatiotemporal cloud dynamics may have a negative impact on the solar power output modeling [5]. Spatiotemporal information should therefore be integrated in solar forecasting models for more skillful forecasts [5,16]. There are three basic methods to provide spatiotemporal data for solar forecasting in different forecast horizons: (1) locally sensed sky images are generally suitable for intra-hour forecasting [17]; (2) satellite-based data are particularly informative for intra-day forecasting [3,6]; and (3) numerical weather prediction (NWP) products are traditionally used for day-ahead forecasting [5]. Sky images are generally used to extract fine-grained cloud motions for the near future with various image processing techniques [18]; satellite imagery and satellite-derived products can provide atmospheric information (i.e., cloud and aerosol) for a larger scale due to their broad spatial coverage [19]; while NWP models simulate atmospheric dynamics with a set of physical equations and solar irradiance forecasts can be a few days ahead [20]. In various solar forecasting scenarios, machine learning based data-driven methods have been extensively used to extract features from different types of inputs [21]. Nevertheless, since the spatiotemporal resolution of satellite- and NWP-based products becomes increasingly finer, it ought to be beneficial to collectively consider different types of information that can infer the atmospheric process in solar forecasting [4,5]. For instance, sky images and 5-min satellite measurements can be incorporated for intra-hour forecasting [3,22]. Similarly, satellite-based data and hourly updated NWP products can be integrated for intra-day forecasting [5].

Numerous machine learning methods have been proposed for solar forecasting with various types of inputs [10,22], for instance, vision transformers with sky images [23], the deep learning model chain with satellite images [6], and post-processing NWP irradiance forecasts with deep learning [24]. However, limited studies explore the combination of satellite- and NWP-based data for intra-day solar forecasting, given their much increased spatial and temporal resolutions. In the integration of historical observations and known-future information for forecasting applications, temporal fusion transformer (TFT) has been widely used with accurate and interpretable results [25]. In TFT-based solar forecasting, observed inputs generally include on-site measurements of solar irradiance, solar power output, and some meteorological data (e.g., temperature and wind speed), while known-future inputs typically consist of clear-sky solar irradiance, calendar information, and weather forecasts (e.g., temperature and relative humidity) [8,26]. However, this actually does not leverage full advantages of operational NWP models since various NWP products can provide valuable information regarding the future atmospheric status [20]. Among the numerous NWP variables, the ones that are most related with solar radiation include forecasts of irradiance itself and cloud amount. While forecasts of temperature and wind speed can affect PV power outputs, their impacts on shortwave solar irradiance are limited [27]. Meanwhile, NWP models already take wind speed and direction into account when modeling cloud dynamics [20]. On this account, it could be beneficial to incorporate both past satellite observations and NWP forecasts of irradiance and cloud amount in TFT for improved intra-day solar forecasting.

As mentioned, spatiotemporal inputs are more advantageous than the single-location information for solar forecasting due to the spatiotemporal characteristics of solar irradiance [5,16]. However, TFT-based solar forecasting models usually use multiple one-dimensional time series data [25,28], while the two-dimensional spatiotemporal information cannot be directly incorporated. To address this challenge, a simple yet effective method is adopted to integrate the two-dimensional spatiotemporal information into TFT-based solar forecasting. Since on-site irradiance observations are typically one-dimensional data, regional satellite-derived irradiance and NWP forecasts of irradiance and cloud amount are used as the spatiotemporal inputs [6]. Moreover, advanced machine learning models may have a high-performance in solar forecasting, while the interpretability of such models are generally compromised [29]. Another advantage of TFT-based solar forecasting

Nomenclature

BON	Bondville
CSI	Clear-sky index
DHI	Diffuse horizontal irradiance
DNI	Direct normal irradiance
DRA	Desert Rock
ELU	Exponential linear unit
FPK	Fort Peck
GHI	Global horizontal irradiance
GLU	Gated linear unit
GRN	Gated residual network
GRU	Gated recurrent unit
GWN	Goodwin Creek
HRRR	High-Resolution Rapid Refresh
LSTM	Long-short-term memory
MAE	Mean absolute error
MLP	Multi-layer perceptron
nMAE	Normalized mean absolute error
nRMSE	Normalized root mean squared error
NSRDB	National Solar Radiation Database
NWP	Numerical weather prediction
PSU	Pennsylvania State University
PV	Photovoltaic
QC	Quality control
RMSE	Root mean squared error
SURFRAD	Surface Radiation Budget Network
SXF	Sioux Falls
SZA	Solar zenith angle
TBL	Table Mountain
TCN	Temporal convolution network
TFT	Temporal fusion transformer
UTC	Coordinated Universal Time
XGBoost	Extreme gradient boosting

Notations

$\bar{o}$	Mean of observations
$\circ$	Degree
η	Intermediate layer
$\hat{I}$	Clear-sky index forecast
$\hat{X}$	Known future inputs
$\mathbb{F}$	Forecasting model
$\odot$	Element-wise Hadamard product
$\sigma(.)$	Sigmoid activation function
a	Input
b_i	Bias
d_{attn}	Dimensionality
f	Forecast
H	Attention head
h	Number of heads
K	Key
N	Number of data points
o	Observation
Q	Query
R^2	Coefficient of determination
t	Time
V	Value
W_i	Weights
X	Historical inputs
DNI_sat	Satellite-derived direct normal irradiance
DSWRF	Downward shortwave radiation flux
GHI_sat	Satellite-derived global horizontal irradiance
HCDC	High cloud cover
LCDC	Low cloud cover
MCDC	Middle cloud cover
TCDC	Total cloud cover
VBDSF	Visible beam downward solar flux

Subscripts

f	Forecast of interest
r	Reference forecast
t	Time

Superscripts

i	The ith index
j	The jth index
T	Transpose

is its interpretability on the attention mechanism and variable selections [8,25]. The interpretability is generally known as the contribution of each input feature of the model to forecasting results [30]. To facilitate the understanding and design of forecasting models in solar energy systems, it is necessary to address their interpretability [29]. Many techniques (model-specific and model-agnostic) have been used in explainable machine learning for global or instance interpretability analysis, such as Shapley additive explanations [31], permutation feature importance [32], and local interpretable model-agnostic explanations [33]. While these techniques can provide explainability for the behavior of machine learning forecasting models, their interpretations on the final forecasting results are still unclear [34]. Since solar irradiance exhibits huge variations due to the changing weather conditions, this study also aims to investigate the interpretability of TFT-based solar forecasting. This can better support the integration of solar energy into the power system, supporting grid regulation, load-following generation, and unit commitment with reduced costs [3,17], as shown in Fig. 1. The major contributions are summarized as follows:

- High-resolution (3-km–1-hour) NWP products are used as both past and future covariates in TFT-based solar forecasting, which can lead to improved intra-day solar irradiance forecasts up to 4-hour ahead when integrated with satellite-derived irradiance products.
- Since spatiotemporal inputs are more beneficial for solar forecasting, a simple yet effective method is applied to incorporate such two-dimensional spatiotemporal satellite- and NWP-based information into TFT-based intra-day solar forecasting. The use of spatiotemporal information is more likely to yield better solar irradiance forecasts than single-location data.
- Interpretability analysis is performed to improve the understanding of solar forecasting models, which can ensure the transparency in decision-making process of data-driven models. Based on the interpretability analysis results at several locations, it is suggested that solar forecasting models should be developed and interpreted considering the local weather characteristics.

The rest of this study is organized as follows: The used data of ground-level measurements, satellite-derived irradiance, and NWP products are described in Section 2. The TFT-based intra-day solar

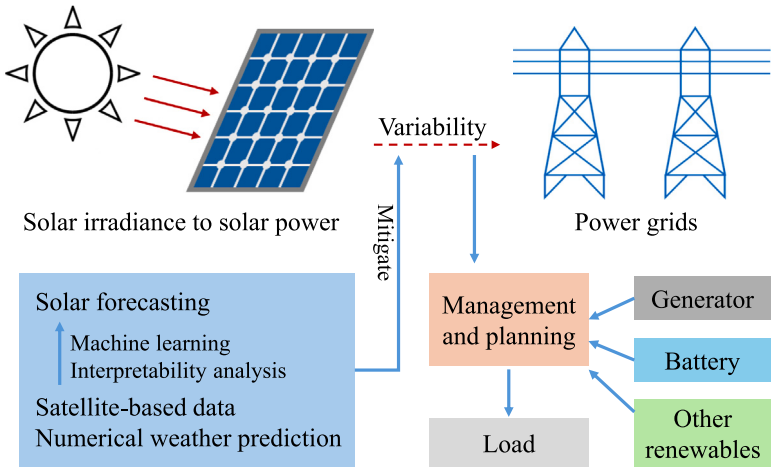


Fig. 1. The role of solar forecasting in supporting the integration of solar energy.

Table 1
The geography and climate information of the seven SURFRAD stations.

Abbr.	Station	Latitude (°)	Longitude (°)	Altitude (m)	Climate[a]	Time zone
BON	Bondville	40.05	−88.37	230	Dfa	UTC-6
DRA	Desert Rock	36.62	−116.02	1007	Bwk	UTC-8
FPK	Fort Peck	48.31	−105.10	634	Bsk	UTC-7
GWN	Goodwin Creek	34.25	−89.87	98	Cfa	UTC-6
PSU	Penn. State Uni.	40.72	−77.93	376	Dfb	UTC-5
SXF	Sioux Falls	43.73	−96.62	473	Dfa	UTC-6
TBL	Table Mountain	40.12	−105.24	1689	Bsk	UTC-7

[a] Köppen climate abbreviations: Bsk (arid, steppe, cold), Bwk (arid, desert, cold), Cfa (temperate, without dry season, hot summer), Dfa (continental, without dry season, hot summer), Dfb (continental, without dry season, warm summer).

forecasting model and the method to incorporate spatiotemporal inputs are detailed in Section 3. Section 4 presents verifications of the proposed method, related implications, and interpretability analysis. Finally, Section 5 concludes this study.

2. Data

2.1. Ground-level measurements

Ground-level data used as the target for training the forecasting model and the ground truth for evaluating model performance are obtained from the Surface Radiation Budget Network (SURFRAD) [35]. SURFRAD consists of seven stations located in diverse climate zones in the United States, which capture a wide range of surface and atmospheric conditions. Since its establishment in 1993, SURFRAD has provided high-quality meteorological measurements for various applications related with weather, climate, and energy. Table 1 details the geography and climate information of the seven SURFRAD stations.

For solar irradiance, SURFRAD offers measurements of GHI, DNI, and DHI at a time resolution of 1 minute, and the solar zenith angle (SZA) is calculated based on the solar positioning algorithm [36]. All the 1-minute data of GHI, DNI, DHI, and SZA in year 2020 are downloaded, and quality control (QC) is performed to filter out data points that are unlikely to occur in nature. The widely used QC steps include extremely rare limit test and three-component closure test [37], which are considered essential steps when using SURFRAD data and other radiometric measurements. More technical details on the QC steps in pre-processing raw measurements of solar irradiance can be found in [37,38]. Due to the large airmass effect and low values of solar irradiance at high zenith angles [39], data points with a SZA of 85° or greater are removed. After QC is performed on the raw SURFRAD data, the 1-minute measurements are resampled to the 1-hour resolution and indexed in Coordinated Universal Time (UTC) to match the temporal resolution of the NWP products [40]. The "center" scheme is used for resampling, as suggested by Yang et al. [40]. For instance, the GHI indexed at 14:00 represents the aggregation of 1-minute SURFRAD GHI from 13:31 to 14:30. The 1-hour GHI time series are normalized by the hourly clear-sky GHI of REST2 [41], and then used as targets in the forecasting model.

2.2. Spatiotemporal data

The spatiotemporal data used in this work include satellite-based irradiance estimations and NWP products with a regional size of 11 × 11 grids, as shown in Table 2. Satellite-derived irradiance includes both GHI and DNI, which are obtained from spectral satellite images (detailed in Section 2.2.1). NWP products are from the operational model of High-Resolution Rapid Refresh (HRRR) [42], the used variables are related with surface solar radiation flux and cloud amount (see Section 2.2.2).

2.2.1. Satellite-derived irradiance

Satellite-derived irradiance products have been extensively used in solar forecasting with data-driven approaches, as either the sole or a part of the inputs [6]. Owing to the advancement in modern remote sensing technologies, the spatiotemporal resolution of satellite-derived irradiance has been improved significantly. For example, the national solar radiation database (NSRDB) [43], as the state-of-the-art satellite-to-irradiance data, offers more than 20 years of complete satellite-based irradiance estimations of GHI, DNI, and DHI for the United States in the spatiotemporal resolution of 4-km–30-minute. This has been improved to 2-km–5-minute since 2018 due to the increased resolution of geostationary operational and environmental satellites [44].

The validation of 5-minute NSRDB data against SURFRAD measurement shows comparatively higher discrepancies due to the refined temporal resolution [44]. Therefore, an improved satellite-to-irradiance method based on the combination of deep learning and remote sensing is developed for both GHI and DNI estimations [45]. The estimated

Table 2
A summary of spatiotemporal data used for intra-day solar irradiance forecasting.

Data	Long name	Type	Resolution	
			Temporal	Spatial
GHI_{sat} [W/m^2]	Satellite-derived GHI	Satellite-based	5-min	2-km
DNI_{sat} [W/m^2]	Satellite-derived DNI	Satellite-based	5-min	2-km
DSWRF [W/m^2]	Downward shortwave radiation flux	NWP-based	1-h	3-km
VBDSF [W/m^2]	Visible beam downward solar flux	NWP-based	1-h	3-km
HCDC [%]	High cloud cover	NWP-based	1-h	3-km
MCDC [%]	Middle cloud cover	NWP-based	1-h	3-km
LCDC [%]	Low cloud cover	NWP-based	1-h	3-km
TCDC [%]	Total cloud cover	NWP-based	1-h	3-km

GHI and DNI based on deep learning show relatively lower uncertainties compared with those of NSRDB [45]. In a subsequent study [6], the deep learning method for satellite-based irradiance estimation is expanded to generate spatial GHI estimates, which are then used as spatiotemporal inputs for better intra-day solar forecasts [6].

Following the same methodology, the spatiotemporal satellite-derived GHI and DNI are obtained with two steps [6]: (1) the deep learning satellite-to-irradiance model is first developed for the location (usually with adequate on-site measurements) centered in the region of interest; (2) the spatial irradiance estimations are then obtained by using the pre-trained deep learning model with shifted input windows in a pixel-wise way. For more details on the acquisition of spatial irradiance estimations, the reader is referred to [6]. The obtained spatial GHI and DNI estimates are in the form of 5-minute–2-km clear-sky index (CSI) values with the regional size of 11 × 11 pixels in the year 2020, which might have gaps due to the failure or technical issue in the communication with satellite sensors. Nevertheless, these missing points are filled with data from NSRDB (normalized by the clear-sky irradiance of REST2). These CSI values are resampled to the time resolution of 1-hour using the "center" scheme (as mentioned in Section 2.1) and then used as parts of inputs for intra-day solar forecasting. Some sample spatial satellite-derived GHI and DNI estimates are presented in Fig. 2. It should be noted that the deep learning models for satellite-based GHI and DNI estimations are trained separately, and the obtained CSI is not converted to irradiance value at this step.

2.2.2. HRRR forecasts

HRRR is an operational NWP model developed by National Oceanic and Atmospheric Administration in the United States [42], which started to provide high-resolution weather forecasts in 2014. The spatial resolution of HRRR is 3-km, and the forecasts are updated hourly. HRRR issues 0–18-hour-ahead forecasts in each forecasting cycle, where $t = 0$ indicates the reanalysis for initial conditions. Additionally, HRRR extends forecasts to 48-hour ahead in four runs at 00Z, 06Z, 12Z, and 18Z. In terms of intra-day solar forecasting, only the 1–16-hour-ahead forecasts released from 00Z and 12Z runs are used.

There are numerous weather variables in HRRR forecasts, of which surface radiation flux and cloud covers are most related with solar irradiance forecasting applications [27]. The variable corresponds to GHI in HRRR is named as "downward short-wave radiation flux" or DSWRF. DNI is known as "visible beam downward solar flux" or VBDSF. Both DSWRF and VBDSF are surface variables, while variables related with cloud amount are in different layers in the atmosphere. Cloud amount in the high, middle, and low cloud layers are named as "high cloud cover (HCDC), "middle cloud cover (MCDC)", and "low cloud cover (LCDC)", respectively. The total cloud amount in the entire atmosphere is represented by "total cloud cover" or TCDC. Table 2 presents the summary of the HRRR variables used in this work. To be compatible with the spatial size of satellite-derived irradiance data, all the mentioned NWP variables in year 2020 are downloaded for the 11 × 11 locations surrounding the target SURFRAD stations. Due to the technical issue and communication failure, there are some data gaps in HRRR forecasts, which are filled with linear interpolations. All HRRR forecasts are pre-processed before being fed into the forecasting model, specifically, irradiance forecasts are normalized using the hourly REST2 clear-sky irradiance and cloud amount values are converted from percentages into decimals. Some examples of the normalized regional radiation flux and cloud amount from HRRR forecasts are shown in Fig. 2.

3. Methods

3.1. Vanilla TFT model

TFT is an attention-based deep learning framework for multi-horizon time series forecasting, which has been extensively applied in solar irradiance and power forecasting [8,46]. TFT combines the advantages of feature selection, sequence-to-sequence models, and attention mechanism. This enables TFT to integrate diverse inputs, capture long-term dependencies, and model complex temporal relationships [25]. TFT is originally designed to build feature representation for different types of inputs (i.e., static, observed, and future known inputs). The major components of TFT include static covariate encoders, variable selection networks, gating mechanism, temporal processing, and prediction intervals [25]. Since only historical and future-known inputs are used for deterministic intra-day solar irradiance forecasting in this study, the TFT model is tailored accordingly. In specific, the static covariate encoders and prediction intervals are removed. Fig. 3 shows the detailed structure of TFT, and the main constitutes are briefly explained as follows:

- Gated residual network (GRN): gating mechanism is used to control the flow of information, which enables the model to focus on the most relevant features by applying non-linear processing. This ability on filtering information with "gates" can improve the model's performance on capturing complex patterns, processing long-term dependencies, and avoiding irrelevant inputs. The GRN, as shown in Fig. 3(b), is based on gating mechanism and residual connection,

$$\mathrm{GRN}(a) = \mathrm{LayerNorm}(a + \mathrm{GLU}(\eta_2)), \tag{1}$$

$$\eta_2 = W_2\eta_1 + b_2, \tag{2}$$

$$\eta_1 = \mathrm{ELU}(W_1 a + b_1), \tag{3}$$

$$\mathrm{GLU}(\eta_2) = \sigma(W_3\eta_2 + b_3) \odot (W_4\eta_2 + b_4), \tag{4}$$

  where LayerNorm is the standard layer normalization, a is the input, GLU is the gated linear unit, η_1 and η_2 are intermediate layers, W_i and b_i denote the weights and bias. ELU refers to the exponential linear unit activation function, $\sigma(.)$ represents the sigmoid activation function, and $\odot$ is the element-wise Hadamard product.
- Variable selection network: it is to select related input variables at each time step. Solar forecasting may involve indigenous and plenty of exogenous variables, while their impacts on the output are typically unknown. The use of variable selection is to provide information on what inputs are most relevant to the forecasting

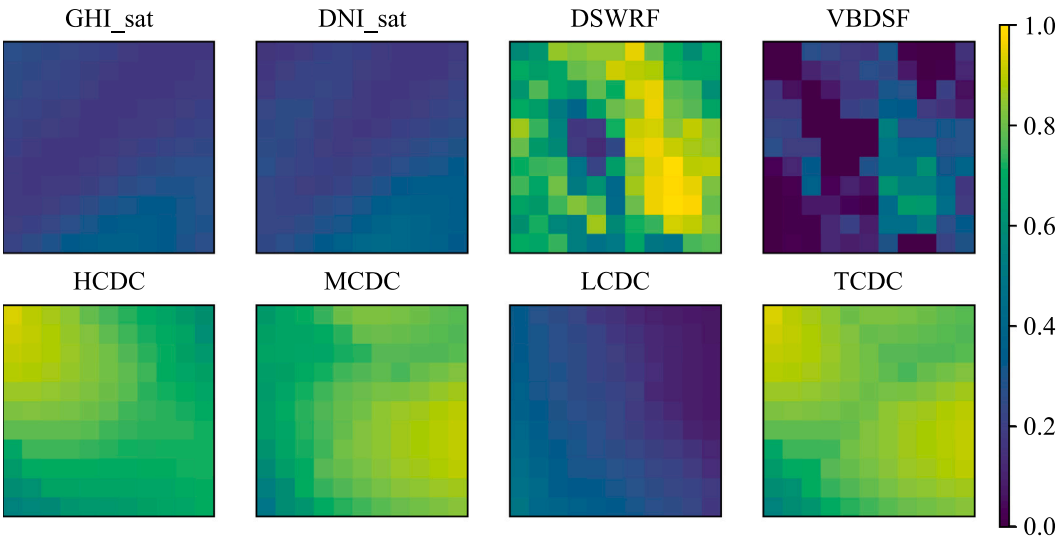


Fig. 2. Sample images of the normalized spatiotemporal data for Bondville (BON) station (40.05°, −88.37°) at 2020-04-23 18:00:00 (UTC). BON is at the center in the region of 11 × 11 grids. Note that satellite- and NWP-based data have different spatial resolutions but with the same 11 × 11 surrounding grids.

problem, and to remove noisy and redundant features. As shown in Fig. 3(c), each variable x_t^j of the transformed inputs at time t is fed through its own GRN to get $\tilde{x}_t^j$ (the processed feature vector of variable j):

$$\tilde{x}_t^j = \mathrm{GRN}_{x(j)}(x_t^j). \tag{5}$$

The flattened inputs $\left[(x_t^1)^T, \ldots, (x_t^m)^T\right]^T$, where m is number of inputs) at time t are then processed by a GRN and a Softmax layer to get the variable selection weights v_{x_t}, which are then combined with the processed features $\tilde{x}_t$ from Eq. (5):

$$\tilde{x}_t = \sum_{j=1}^{m} v_{x_t}^j \tilde{x}_t^j, \tag{6}$$

where j indicates the jth element of vectors v_t and $\tilde{x}_t$.

- Interpretable multi-head attention: it is a key component in TFT to capture long-term dependencies in different time steps and to provide explainability (e.g., importance of features and time steps). Generally, the attention mechanism includes values V, keys K, and queries Q, the V values are scaled based on K and Q, which is expressed as follows:

$$\mathrm{Attention}(Q, K, V) = \mathrm{Softmax}(Q \cdot K^T / \sqrt{d_{attn}}) \cdot V, \tag{7}$$

where d_{attn} is the dimensionality of Q and K, and Softmax is the function to get attention weights. Then multi-head attention is:

$$\mathrm{MultiHead}(Q, K, V) = [H_1, H_2, \ldots, H_h] \cdot W, \tag{8}$$

where $H_i = \mathrm{Attention}(Q \cdot W_Q^i, K \cdot W_K^i, V \cdot W_V^i)$, h is the number of heads, W_Q^i, W_K^i, and W_V^i are head-specific weights of queries, keys and values, W is the linear combination of outputs from all heads. The interpretable multi-head attention with a shared head ($\tilde{H}$) can be represented by:

$$\mathrm{IMultiHead}(Q, K, V) = \tilde{H} \cdot W, \tag{9}$$

$$\tilde{H} = \frac{1}{h} \sum_{i=1}^{h} \mathrm{Attention}(Q \cdot W_Q^i, K \cdot W_K^i, V \cdot W_V), \tag{10}$$

where W_V are the weights of V shared across all heads. A more detailed description of TFT and its major components can be found in Lim et al. [25].

3.2. TFT model with spatiotemporal inputs

It has been reported that TFT can produce more skillful solar irradiance and power forecasts than other machine/deep learning methods (e.g., MLP, LSTM, and XGBoost) [8]. However, these TFT models generally consider multiple types of time series data at a single location, for instance, historical irradiance/power measurements and future clear-sky irradiance estimations. These single-location based data generally do not contain any spatiotemporal information. Since the use of spatiotemporal inputs can generally lead to better solar forecasts than the inputs from a single station [5], a simple and effective method is used to integrate such spatiotemporal information into TFT based intra-day solar forecasting. In specific, multiple spatiotemporal satellite- and NWP-based data in a region are spatially averaged and then inputted to the TFT model for point irradiance forecasts (see Fig. 4). Actually, spatial averaging is widely used to improve the modeling performance of solar irradiance [47,48]. However, this has not yet been applied in TFT based models for solar forecasting. To explore the size effect of spatiotemporal inputs on forecasting results, regional spatiotemporal data with different sizes (pixels) are used:

- 11 × 11: All data of the 11 × 11 surrounding locations are used to calculate the means of satellite-derived irradiance and NWP irradiance and cloud amount forecasts.
- 7 × 7: The average values are obtained using data of the surrounding 7 × 7 pixels.
- 3 × 3: The means are calculated in the same way of satellite- and NWP-based data, while the involved region is of 3 × 3 locations.

Since TFT generally outperforms a variety of machine/deep learning models in time series forecasting, and the aim of this study is to investigate the benefits from spatiotemporal NWP inputs, the forecasts based on the single-pixel inputs (i.e., data of 1 × 1 pixel) are used as the benchmark. Note that there is a tradeoff between the spatial size of spatiotemporal inputs and the modeling efficiency [49]. The benefit of spatial scales becomes marginal when the size exceeding a certain threshold while much more efforts in data processing are required [49]. Therefore, 11 × 11 grids of surrounding data are adopted following several studies on satellite-to-irradiance conversion [45], cloud detection [50], and intra-day solar forecasting [6,51].

As shown in Fig. 4, both satellite-derived irradiance and lagged NWP products are used as inputs at the encoder side, while NWP forecasts are the only inputs at the decoder side. The point irradiance

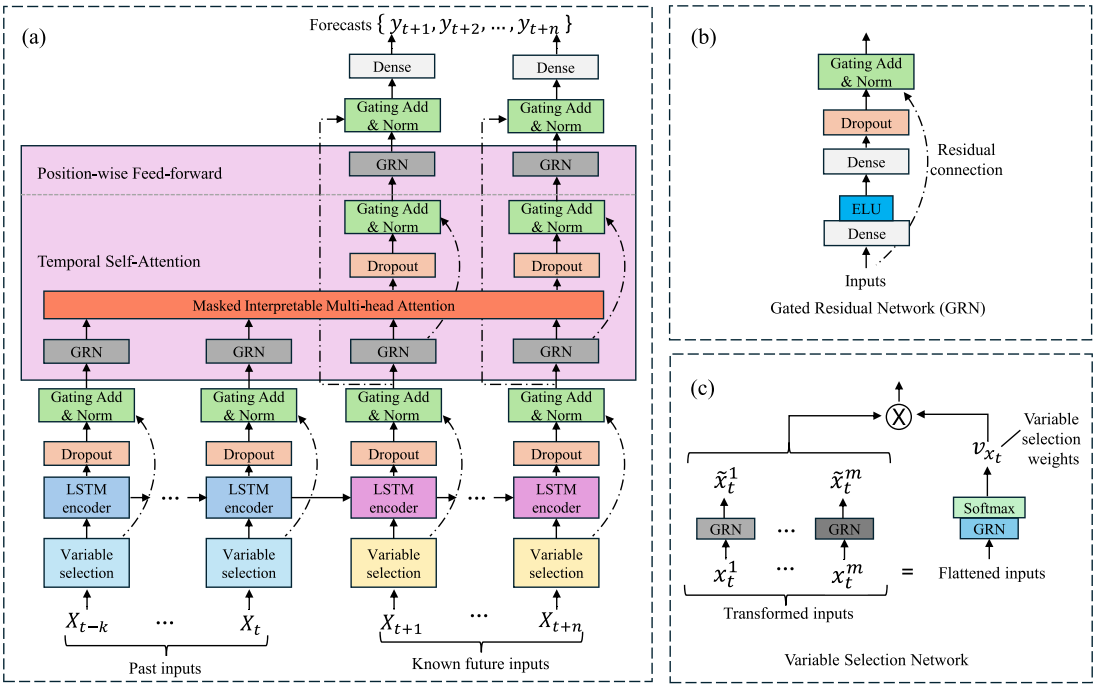


Fig. 3. The structure of (a) temporal fusion transformer (TFT) used for mulit-horizon solar forecasting, where (b) and (c) present the details of the gated residual network (GRN) and variable selection network, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

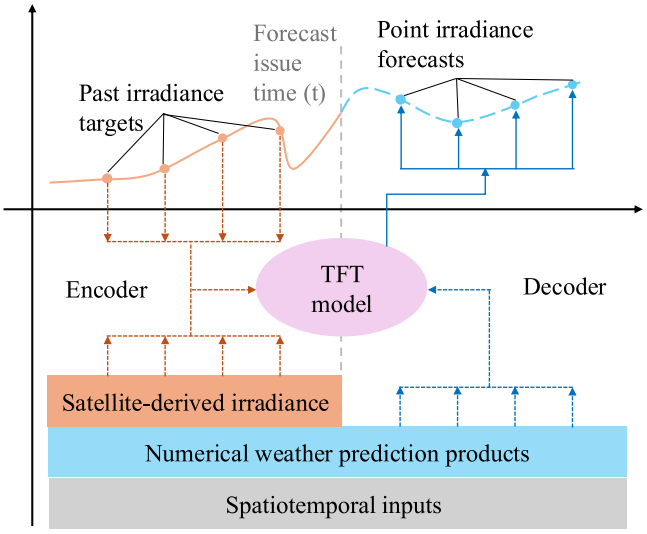


Fig. 4. The flowchart of multi-horizon solar forecasting with multiple spatiotemporal inputs.

forecasts of GHI are produced for next 4 hours, and the forecasting problem can be formulated as:

$$\hat{I}_{t_0+1}, \hat{I}_{t_0+2}, \ldots, \hat{I}_{t_0+4} = \mathbb{F}(X_{t_0-k}, X_{t_0-k+1}, \ldots, X_{t_0}, \tilde{X}_{t_0+1}, \ldots, \tilde{X}_{t_0+n}), \quad (11)$$

where $\hat{I}$ is the CSI forecast of GHI, $\mathbb{F}$ denotes the forecasting model, X and $\tilde{X}$ represent past and future-known inputs, respectively. t_0 represents the forecast issue time, k (equals 6) is the number of used past time steps, and n is the number of time steps predicted, which is 4 in this work.

Since X and $\tilde{X}$ have different types of inputs, the parameters in their variable selection networks are different (this is why the colors for variable selection in Fig. 3 are not the same). As the irradiance forecasts of interest are deterministic, the loss function is not quantile but based on the mean absolute error (MAE). The used optimizer is Ranger [52], and the hyperparameters for the forecasting model are optimized based on ranges defined in Table 3. Data from January, February, March, July, August, and September are selected for training, while data in April and October are for validation, and the rest months are used for testing. Note that the TFT model for solar forecasting at different stations are trained separately, therefore the hyperparameters might be different.

Table 3
Hyperparameters for the transformer based solar forecasting model.

Hyperparameter	Values
Optimizer	Ranger
Loss function	MAE
Batch size	32
Learning rate	[1e−5, 1e−2]
Gradient clipping range	[0.01, 1]
Hidden size	[16, 64]
Hidden continuous size	[16, 64]
Attention head	[2, 6]
Dropout	[0.3, 0.6]
Reduce on plateau patience	5
Early stopping patience	10

3.3. Evaluation metrics

Although the CSI of GHI is the forecasting target, evaluations of error are based on the irradiance [W/m^2]. Therefore, CSI forecasts are converted back to GHI by multiplying the clear-sky irradiance of REST2 at the predicted time stamps. A variety of evaluation metrics are used to assess the overall forecasting accuracy [53], including root mean squared error (RMSE), normalized RMSE (nRMSE), MAE, normalized MAE (nMAE) and coefficient of determination (R^2), as defined by the following equations:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum(f_i - o_i)^2} \quad (12)$$

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{N}\sum(f_i - o_i)^2}}{\frac{1}{N}\sum o_i} \tag{13}$$

$$\text{MAE} = \frac{1}{N}\sum |f_i - o_i| \tag{14}$$

$$\text{nMAE} = \frac{\frac{1}{N}\sum |f_i - o_i|}{\frac{1}{N}\sum o_i} \tag{15}$$

$$R^2 = 1 - \frac{\sum(f_i - o_i)^2}{\sum(o_i - \bar{o}_i)^2}, \tag{16}$$

where f_i and o_i are the pair of GHI forecast and ground measurement, $\bar{o}_i$ is the mean of the measurement, N is the number of total data points compared.

The forecast skill quantifies the relative improvement of a forecasting model over the benchmark [3,17]. The forecast skill can be calculated based on RMSE [40], defined as follows:

$$\text{Forecast skill} = \left(1 - \frac{\text{RMSE}_f}{\text{RMSE}_r}\right) \times 100\% \tag{17}$$

where RMSE_f is the forecast error of the evaluated model, and RMSE_r is based on the benchmark model, which uses single-location data of 1×1 as inputs (see Section 3.2).

4. Results and discussion

In this section, solar irradiance forecasts based on different sizes of spatiotemporal inputs are compared and evaluated. The effect of spatiotemporal inputs with different size on the forecasting results and the benefit of spatiotemporal inputs are presented and discussed in Section 4.1. The forecasting performance under different sky conditions are compared in Section 4.2, and the interpretability analysis of the forecasting model is performed in Section 4.3. Section 4.4 presents a comparative analysis and investigates the effectiveness of using spatiotemporal inputs for irradiance forecasting with other methods.

4.1. The benefit of spatiotemporal inputs

To evaluate the benefit of spatiotemporal information for solar irradiance forecasting using TFT, this section presents the forecasting results of GHI based on different sizes of used inputs. As shown in Fig. 5, with extended time horizons, the forecasts are associated with increased nRMSEs. This indicates that forecasts for longer time horizons are typically more difficult to model and therefore have higher uncertainties. Compared with GHI forecasts generated by the single-pixel data (1×1), forecasts based on spatiotemporal inputs generally have lower nRMSE values, demonstrating the benefit of using a greater amount of spatiotemporal information as the input. Meanwhile, the size of used spatiotemporal inputs also has an impact on the forecasting results. When comparing the forecasts produced using different sizes of spatiotemporal inputs, the general trend is that forecasts produced using larger spatial sizes of inputs typically show lower nRMSE errors. This could be attributed to the increased amount of information from the surrounding area. Therefore, even using the simple spatial averaging method, the spatiotemporal information can be used with TFT to yield GHI forecasts with lower discrepancies. However, it should be noted that the increased amount of spatiotemporal information does not necessarily lead to better forecasting results. For example, GHI forecasts based on spatially averaged inputs at DRA do not show significant improvements than those generated from single-pixel data (see Fig. 5(b)), which is possibly due to the highest occurrence of clear skies at that location [6]. When the sky is free of clouds, spatial averaging actually does not show any differences. Moreover, at BON and PSU (see Fig. 5(a) and (e)), the larger size of spatiotemporal inputs even results in slightly higher errors in the forecasts. The possible reason could be the drawback of spatial averaging and the inherent uncertainty of NWP forecasts (which is discussed in Section 4.2).

Table 4
The forecast skill [%] over the benchmark for solar irradiance forecasts up to 4-hour ahead using different sizes of spatiotemporal inputs, at all the SURFRAD stations.

Station	Inputs	Forecast horizon			
		1-hour	2-hour	3-hour	4-hour
BON	3×3	23.42	16.74	11.78	11.13
	7×7	22.24	14.85	9.31	8.11
	11×11	20.43	15.36	10.84	9.14
DRA	3×3	9.05	3.73	1.53	1.80
	7×7	8.50	6.78	5.63	6.05
	11×11	10.24	5.76	3.30	3.31
FPK	3×3	12.28	9.67	7.01	3.97
	7×7	15.31	13.23	11.16	9.60
	11×11	17.61	16.11	14.17	12.24
GWN	3×3	2.82	4.43	4.82	4.70
	7×7	3.88	5.96	6.04	7.26
	11×11	6.91	8.61	9.58	10.04
PSU	3×3	12.19	10.68	11.06	10.27
	7×7	14.98	8.24	7.69	6.13
	11×11	12.47	7.94	8.19	7.26
SXF	3×3	5.63	3.53	3.56	3.34
	7×7	5.09	3.99	4.64	5.54
	11×11	13.54	10.44	9.91	11.19
TBL	3×3	8.51	6.02	3.32	2.33
	7×7	9.62	5.98	4.03	4.18
	11×11	10.26	7.69	5.74	5.72

The comparison of nMAE for 1–4-hour ahead GHI forecasts at all SURFRAD stations is presented in Fig. 6. Although nMAE and nRMSE describe different aspects of the GHI forecast, they generally show a similar trend. Specifically, forecasts for longer horizons are typically associated with higher errors. The use of spatially averaged spatiotemporal data can be used to generate better results in most cases, and the increased size of surrounding area can further reduce the GHI forecast errors. The comparable performance at DRA and out-of-trend nMAE variations at BON and PSU can also be observed. However, it is necessary to mention that the increased amount of spatiotemporal information usually require more efforts in data collection and pre-processing, while the improvement in resulted forecasts could be marginal. This tradeoff should be considered when choosing the size of spatiotemporal data for solar forecasting applications. In other words, the cost on data storage and computing should be compared with the economic benefit gained from the improved forecasts [54].

To further quantitatively evaluate the advantage of using spatiotemporal inputs with TFT in solar forecasting, forecast skills over the benchmark (forecasts based on single-pixel (1×1) data) are calculated. As shown in Table 4, although there are some site-specific variations (e.g., at BON and PSU), the use of larger size of spatiotemporal inputs can generally produce more skillful forecasts. This indicates that the spatiotemporal data at a bigger surrounding region can offer more informative inputs for solar forecasting applications with TFT. Meanwhile, it can be observed that forecast skills generally decrease with the extended horizons, the reason is that forecasting at longer horizons is more challenging. Another cause of this might be the accumulated uncertainties in NWP forecasts at a relatively longer term (3–4-hour ahead in this case), since satellite-derived irradiance only have larger impacts on shorter forecast horizons. Therefore, one possible way to further improve the skill for longer term forecasts is to use the calibrated NWP forecasts.

4.2. Comparison under different sky conditions

Since the seven stations in SURFRAD are located in a variety of climate zones, three of them, namely, DRA, PSU, and SXF, are selected

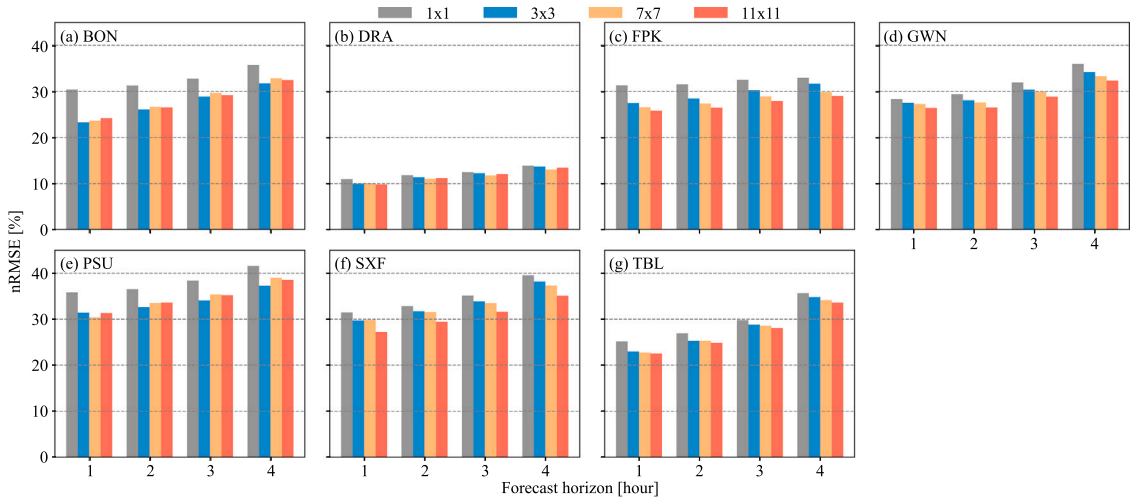


Fig. 5. Comparison of nRMSE [%] for 1–4-hour solar irradiance forecasts at seven SURFRAD stations, using different sizes of spatiotemporal inputs. The number in the legend indicates the size (width × length) of used spatiotemporal data.

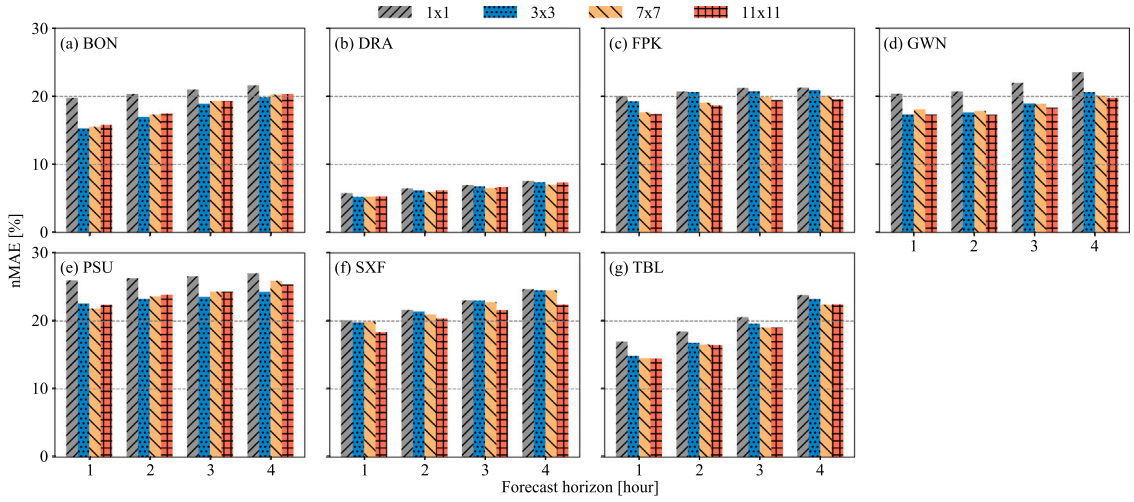


Fig. 6. Comparison of nMAE [%] for 1–4-hour solar irradiance forecasts at seven SURFRAD stations, using different sizes of spatiotemporal inputs. The number in the legend indicates the size (width × length) of used spatiotemporal data.

to be representatives of different sky conditions (i.e., DRA for the most clear, PSU for the most cloudy, and SXF for partly cloudy skies) [6]. The scatter plots of GHI forecast-measurement pairs up to 4-hour ahead at the three selected stations are shown in Fig. 7, which visually represents the joint distribution. It can be seen that the density of GHI forecast-measurement pairs is higher along with the identity line than elsewhere. This indicates that the 1–4-hour ahead GHI forecasts are correctly time-aligned with measurements at the three selected stations under different sky conditions. GHI forecasts are found to have larger errors for the most cloudy skies, as evidenced by the more sparsely distributed forecast-measurement points and the lower values of R^2 at PSU. This means that NWP forecasts also face challenges in correctly modeling complex cloud dynamics, and therefore TFT with such inputs could not produce GHI forecasts as accurate as those at DRA and SXF. The R^2 generally increases with the forecast horizon, suggesting that GHI forecasts can capture the pattern of measurements more effectively at longer time scales. However, the concurrent increase of nRMSE and nMAE indicates a growing difference between predictions and measurements. Nevertheless, the forecasts based on spatiotemporal inputs (11 × 11) generally show higher R^2 than those generated with single-pixel (1 × 1) data. Therefore, the use of spatially averaged spatiotemporal information is more beneficial for the overall solar forecasts under different cloudy conditions.

To gain a more detailed understanding of the solar irradiance forecasting performance under different cloudy conditions, the GHI forecasts are further evaluated in terms of clear, cloudy, and partly cloudy periods at the DRA, PSU, and SXF stations. Periods are classified based on 1-minute on-site irradiance measurements using the Bright-Sun clear-sky detection model [55,56]: a clear hour means all 1-minute instances are detected as clear in the hour, while a cloudy hour indicates all the 1-minute data are identified as cloudy, and the rest hours are labeled as partly cloudy.

The comparisons of 1–4-hour ahead GHI forecasts in clear, cloudy, and partly cloudy periods at DRA, PSU, and SXF are presented in Figs. 8 and 9 for nRMSE and nMAE, respectively. The overall observation is that GHI forecasts in cloudy periods generally have higher values of nRMSE and nMAE than those in clear and partly cloudy periods, while the discrepancies of GHI forecasts in clear periods are comparatively lower. As PSU has more frequent cloudy skies, the forecasts in cloudy periods are also found to have larger errors than those of DRA and SXF, possibly due to the compromised performance of NWP and machine learning in modeling cloud dynamics in cloudy conditions. This indicates that clouds are the main cause of uncertainties for solar irradiance forecasts, and more advanced methods should be developed to learn the nonlinear relationship between clouds and solar irradiance.

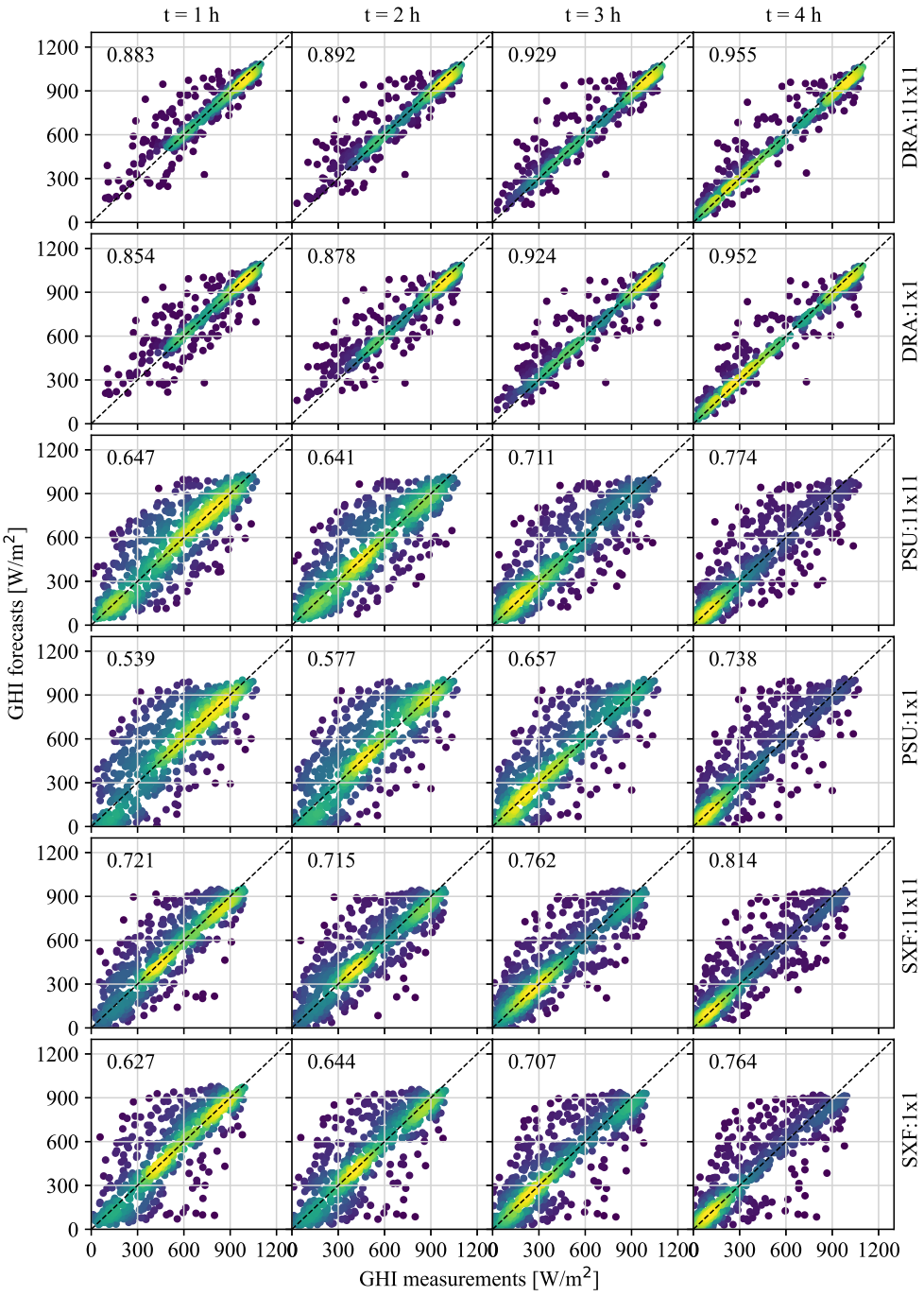


Fig. 7. Scatter plots of GHI forecast-measurement pairs at three selected SURFRAD stations: DRA, PSU, and SXF. The forecasts are based on two different sizes of spatiotemporal inputs: 11 × 11 and 1 × 1 (the baseline). The number in each subplot is the coefficient of determination (R^2) between the GHI forecast and measurement.

When comparing forecasts based on different inputs, the use of a larger size of spatiotemporal data can typically yield forecasts with lower nRMSE and nMAE in cloudy periods. However, these forecasts tend to show slightly higher errors in clear and partly clear periods (e.g., PSU) as shown in Figs. 8 and 9. The reason could be that the spatially averaged spatiotemporal data cannot reflect the true cloud condition at the target location in the center, especially for clear and partly cloudy periods. In other words, the clear sky at the location of interest could be represented as partly cloudy due to the presence of clouds at surrounding pixels, and the true cloud amount in partly cloudy periods might be under- or over-quantified by the surrounding cloud information. These misrepresentations of cloud amount introduced by the spatial averaging can lead to increased uncertainties, therefore the use of larger size of spatiotemporal data does not result in better forecasts for such conditions. A possible solution could be condition-based solar forecasting, where the inputs are selected based on different sky conditions.

4.3. Interpretability analysis

Machine learning models, particularly complex deep neural networks, are usually operated as "black boxes" in various applications. Therefore, it is necessary to perform interpretability analysis to ensure the decision-making process of such models is transparent and

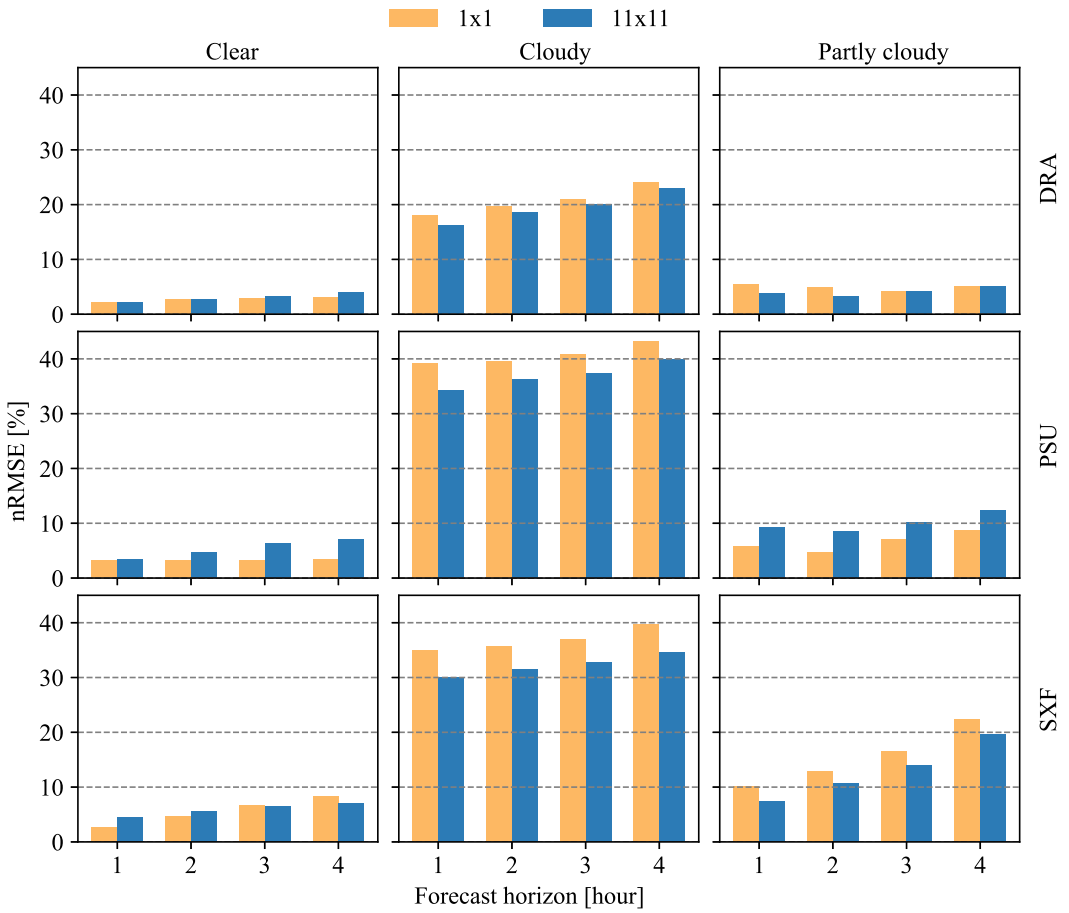


Fig. 8. Comparison of nRMSE [%] for 1–4-hour solar irradiance forecasts at three selected SURFRAD stations (DRA, PSU, and SXF), under clear, cloudy, and partly cloudy sky conditions, using different sizes of spatiotemporal inputs.

understandable. This is particularly important in the integration of solar energy into power systems. One of the most popular methods in interpretability analysis is to calculate the contribution of each feature to the forecasts [34]. On this point, the feature importance of solar forecasting with TFT is analyzed in the following two aspects: (1) the importance based on the variable selection network in the TFT model, and (2) the importance obtained from the feature ablation study.

As mentioned in Section 3.2, the variable selection networks for past and known-future inputs are not the same and the parameters are therefore different (see Fig. 3). This means that the NWP-based data used as past inputs (in the encoder) might have different impacts on the forecasting results compared with those used as known-future inputs (in the decoder). The overall feature importance of variables in the encoder and decoder of TFT-based solar forecasting for DRA, PSU, and SXF at all forecast horizons are presented in Fig. 10(a) and (b), respectively. Overall, it can be seen that satellite-derived GHI (GHI_sat) and total cloud amount (TCDC) have larger impacts in the encoder, while visible beam downward solar flux (VBDSF) and middle cloud amount (MCDC) are more influential in the decoder. However, there are also some site-specific variations as DRA, PSU, and SXF have different climate conditions. For instance, at DRA, GHI_sat shows the largest significance in the encoder, while the second most important feature is DNI_sat rather than TCDC (which is a comparatively more significant input in the encoder of forecasting models at PSU and SXT). The possible reason is that DRA has the most occurrence of clear skies, satellite-derived irradiance of DNI_sat (apart from GHI_sat) is therefore a more important input than the total cloud amount in the forecasting model. As for PSU and SXF with much cloudier sky conditions, the cloud information becomes relatively more influential than DNI_sat (which is directly blocked by the cloud). When it comes to the decoder part, VBDSF generally shows a higher influence than DSWRF, while impacts of cloud covers in different atmospheric layers vary obviously across different climate conditions at DRA, PSU, and SXF. This indicates that solar forecasting model should be developed and interpreted considering local weather characteristics at the station of interest. For the solar forecasting model with a higher generalizability, data from locations with different climate conditions should be used.

The influence of input features on forecasting results across different time horizons is further analyzed using feature ablation [34]. In specific, the forecasting model is firstly trained with all input features and the forecast error is obtained on the test set; then, each feature is removed sequentially and the forecasting model is retrained using the remaining features; finally, the forecasting error is calculated and compared. The relative difference between these two errors can be used to quantify the importance of the excluded feature. Note that the calculation of ablation-based feature importance defined above can be tedious and time-consuming, particularly with a large amount of features. Therefore, features are grouped with similar physical characteristics in this work, namely, satellite-derived irradiance, NWP irradiance, and NWP cloud amount.

The importance of different groups of features for TFT-based solar forecasting at DRA, PSU, and SXF are presented in Table 5. Generally, NWP irradiance shows larger influences on overall forecasting results across various forecast horizons and under different climate conditions, since they are irradiance predictions themselves. Meanwhile, there are also some variations for different locations and forecast horizons. All features tend to be more important for irradiance forecasts when more frequent cloudy sky presents, as the importance of features at PSU is typically higher than those of DRA and SXF. The exclusion of features in TFT-based solar forecasting at DRA generally does not lead

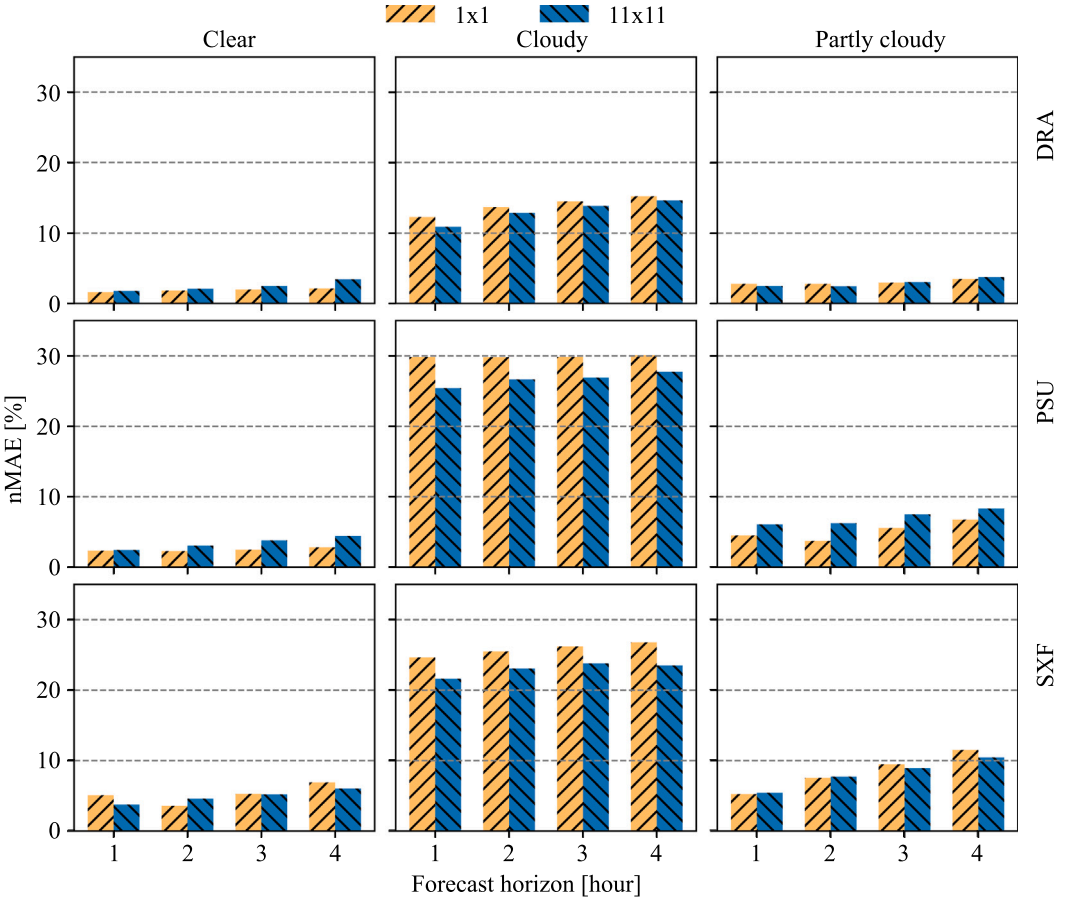


Fig. 9. Comparison of nMAE [%] for 1–4-hour solar irradiance forecasts at three selected SURFRAD stations (DRA, PSU, and SXF), under clear, cloudy, and partly cloudy sky conditions, using different sizes of spatiotemporal inputs.

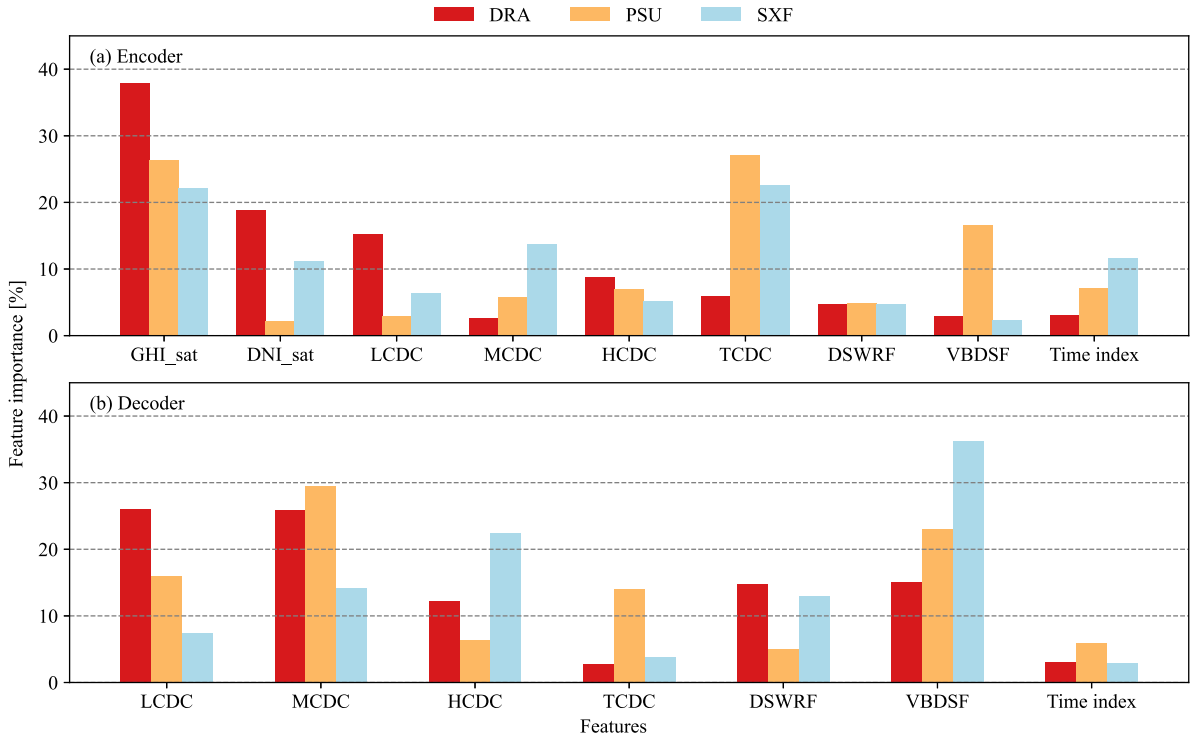


Fig. 10. The importance [%] of used features in the (a) encoder and (b) decoder of TFT based solar forecasting. The description of feature names is available in Table 2.

Table 5
The feature importance [%] ($(nRMSE_{-n}/nRMSE_{All}-1)\times 100\%$, where n denotes removed features) for groups of inputs in the feature ablation study.

Station	Inputs	nRMSE (Feature importance[a]) [%]			
		1-hour	2-hour	3-hour	4-hour
DRA	All	9.82 (–)	11.12 (–)	12.02 (–)	13.43 (–)
	- SDI[b]	10.57 (7.64)	11.53 (3.69)	12.56 (4.49)	14.16 (5.44)
	- NWPI[c]	10.43 (6.21)	11.93 (7.28)	13.32 (10.82)	15.61 (16.23)
	- NWPC[d]	9.92 (1.02)	11.16 (0.36)	12.17 (1.25)	13.69 (1.94)
PSU	All	31.31 (–)	33.62 (–)	35.21 (–)	38.55 (–)
	- SDI	34.43 (9.96)	35.05 (4.25)	36.25 (2.95)	38.87 (0.83)
	- NWPI	38.22 (22.07)	38.13 (13.41)	38.32 (8.83)	40.99 (6.33)
	- NWPC	33.30 (6.36)	34.87 (3.72)	36.67 (4.15)	39.72 (3.04)
SXF	All	27.20 (–)	29.04 (–)	31.63 (–)	35.09 (–)
	- SDI	31.93 (17.39)	32.37 (9.95)	33.75 (6.70)	37.02 (5.50)
	- NWPI	30.14 (10.81)	32.79 (11.38)	35.77 (13.09)	40.03 (14.08)
	- NWPC	29.28 (7.65)	31.02 (5.37)	33.34 (5.41)	37.44 (6.70)

[a] The importance is calculated by comparing the nRMSE of forecasts produced by all features (ALL) and the nRMSE of forecasts based on inputs excluding certain features.
[b] "- SDI" means that satellite-derived irradiance of GHI_{sat} and DNI_{sat} are removed.
[c] "- NWPI" denotes that NWP irradiance of DSWRF and VBDSF are not used.
[d] "- NWPC" excludes NWP cloud cover of HCDC, MCDC, LCDC, and TCDC.

to the high increase of nRMSEs, as DRA has the lowest presence of cloudy skies. Another interesting observation is that satellite-derived irradiance is likely to have larger impacts than NWP irradiance on 1-hour ahead forecasts at DRA (7.64% vs. 6.21%) and SXF (17.39% vs. 10.81%) with less occurrence of cloudy skies, while NWP irradiance becomes more influential at PSU with more clouds. This indicates satellite-derived irradiance can provide more useful information for the forecasting model than NWP models, which become more informative for longer forecast horizons. The importance of NWP cloud amount also varies with climate conditions and forecast horizons. It can be observed that NWP cloud cover shows lowest importance for most clear sky conditions at DRA, and for 1–2-hour ahead forecasts at PSU and SXF with more cloudy skies. As the forecast horizon extends, NWP cloud amount becomes more important than satellite-derived irradiance, since the former is a direct prediction of the future atmospheric conditions. This again suggests that solar forecasting models should be developed, tested, and interpreted under various climate zones and forecast horizons to increase the generalizability.

4.4. The comparative analysis with other methods

Since the use of spatiotemporal inputs generally leads to improved solar irradiance forecasting with TFT, it would be beneficial to test the effectiveness of spatiotemporal inputs for other methods. This section is to present a comparative analysis on irradiance forecasting using several other models with different sizes of inputs. Three benchmark methods, namely, gated recurrent unit (GRU), TimesNet [57], and temporal convolution network (TCN) [58] are used for the comparative analysis. As shown in Table 6, at the three locations of DRA, PSU, and SXF, TFT generally produces forecasts with lower nRMSEs and nMAEs across all forecast horizons compared to the other methods of GRU, TimesNet, and TCN, regardless of the input size. In fact, TFT has shown a superior performance in time series forecasting to a variety of methods [8,25]. It can also be observed in Table 6 that using a larger size of input can lead to improved forecasts no matter which forecasting method is used. This indicates that the use of spatiotemporal inputs, even through the simple spatial averaging, can provide more information for improved solar irradiance forecasting with data-driven methods.

These improved irradiance forecasts can be used to obtain PV power outputs simultaneously, for example, using the physical irradiance-to-power model chain. This can better support the integration of solar energy into the power system. In specific, these 1–4-hour ahead PV power forecasts are necessary for the intra-day unit commitment and economic dispatch [5], which can also provide the latest information for the day-ahead schedules. In this case, further optimizations of the coordination between solar, other renewables, and traditional power generation can be achieved [59]. For more information on solar forecasting and grid integration, the reader is referred to [15,60]. Although the irradiance forecasting models are trained and tested at SURFRAD stations, the methodology can be applied at any other locations and tailored with site-specific data availability.

5. Conclusion

With the growing penetration of solar energy in power systems, accurate solar forecasting has emerged as a cost-effective method to mitigate solar variability and facilitate its integration. As one of the most widely applied deep learning methods, transformer-based approaches (i.e., TFT) show great potential in both solar irradiance and power output forecasting. However, the inputs for TFT are generally multiple one-dimensional time series data, and NWP products are seldom used. To further improve both the performance and transparency of TFT for solar forecasting, this work presents a simple yet effective method to integrate two-dimensional spatiotemporal satellite- and NWP-based inputs for more skillful solar irradiance forecasts.

Compared with the single-location data, using spatiotemporal information can produce better solar forecasts with TFT, even with a simple spatial-averaging method. This is attributed to the increased amount of information from the surrounding area, which can be used to track the cloud movement. Although the increased spatial size of spatiotemporal data can usually lead to better forecasts, more efforts are required in data pre-processing while the improvement of forecasting performance could be marginal under some climate conditions. This tradeoff between the size of spatiotemporal data and their benefits for solar forecasting are critical for real-time application and therefore should be considered. It should be noted that the increased size of surrounding spatiotemporal information could introduce some uncertainties in solar forecasting under clear and partly cloudy sky conditions. The reason might be that the spatially averaged surrounding information could not accurately reflect the true cloud condition at the target location in the center under these conditions. Nevertheless, the combined use of NWP inputs as both past and future-known covariates with satellite data can improve solar forecasting performance with TFT, as evidenced by the interpretability analysis. In general, NWP irradiance products show a larger impact on the overall forecasts, and satellite-derived irradiance tends to have a higher influence on shorter term (1-hour) forecasts. Although the interpretability analysis of solar forecasting models exhibits some variations under different climate conditions and forecast horizons, it serves as an effective tool for developing, testing, and interpreting data-driven solar forecasting models at locations with different climate features.

Despite the improved forecasting performance and interpretability, the proposed method has several limitations: (1) the method to incorporate two-dimensional spatiotemporal information is quite simple, which may lead to misrepresentations of true cloud conditions, more advanced techniques that can efficient extract spatial features should be applied; (2) results show that irradiance forecast errors are slightly higher under clear and partly cloudy periods, where condition-based solar forecasting that takes sky conditions (i.e., clear, partly cloudy, and cloudy) into account could be an effective solution; and (3) the method is only developed and applied at some specific locations, solar forecasting models with enhanced generalizability should be developed using data from various climate zones in the future research.

Table 6
Summary of nRMSE [%] and nMAE [%] for 1–4-hour solar irradiance forecasts with different methods (i.e., gated recurrent unit (GRU), TimesNet, temporal convolution network (TCN), and temporal fusion transformer (TFT)) and sizes of inputs.

Station	Method	Inputs	nRMSE [%]				nMAE [%]			
			1-hour	2-hour	3-hour	4-hour	1-hour	2-hour	3-hour	4-hour
DRA	GRU	1 × 1	12.95	14.37	14.93	17.84	6.79	8.30	10.34	13.25
		11 × 11	11.90	13.35	14.95	17.66	6.93	8.18	9.95	13.29
	TimesNet	1 × 1	12.27	13.03	14.10	17.33	7.41	8.03	9.52	12.61
		11 × 11	13.67	13.99	14.31	16.61	8.32	8.45	9.51	11.82
	TCN	1 × 1	14.42	15.80	16.74	19.11	8.13	9.83	11.92	14.23
		11 × 11	12.90	14.02	15.76	19.08	7.65	8.85	11.03	13.57
	TFT	1 × 1	10.94	11.80	12.43	13.89	5.78	6.48	6.96	7.58
		11 × 11	9.82	11.12	12.02	13.43	5.30	6.20	6.66	7.37
PSU	GRU	1 × 1	38.54	44.12	47.17	49.91	26.53	30.69	34.28	35.62
		11 × 11	35.25	36.52	38.37	43.06	25.27	27.84	28.72	35.07
	TimesNet	1 × 1	34.80	38.70	42.97	47.38	25.88	28.74	32.39	36.71
		11 × 11	33.78	40.49	40.86	46.36	24.29	30.03	30.20	35.87
	TCN	1 × 1	35.38	42.69	45.50	46.55	25.15	29.32	32.43	33.90
		11 × 11	37.47	38.75	41.25	43.68	26.03	27.61	29.74	33.50
	TFT	1 × 1	35.77	37.05	38.35	41.57	25.97	26.29	26.56	26.99
		11 × 11	31.31	33.62	35.21	38.55	22.39	23.88	24.31	25.35
SXF	GRU	1 × 1	33.27	36.78	44.70	47.04	21.78	25.27	28.72	33.82
		11 × 11	30.22	40.18	41.82	47.77	20.43	26.32	28.55	36.76
	TimesNet	1 × 1	30.64	37.14	41.76	49.93	20.34	25.44	27.53	37.57
		11 × 11	27.81	33.36	38.29	47.12	18.93	22.93	24.96	34.96
	TCN	1 × 1	32.63	37.26	45.10	47.52	20.80	25.25	29.80	34.25
		11 × 11	29.38	36.94	38.69	48.52	19.94	25.02	27.05	36.08
	TFT	1 × 1	31.46	32.87	35.11	39.51	20.09	21.64	23.06	24.68
		11 × 11	27.20	29.44	31.63	35.09	18.36	20.36	22.79	22.42

CRediT authorship contribution statement

Shanlin Chen: Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Tao Jing:** Writing – original draft, Software, Methodology. **Mengying Li:** Writing – review & editing, Funding acquisition, Conceptualization. **Hiu Hung Lee:** Writing – review & editing, Supervision, Resources. **Ming Chun Lam:** Writing – review & editing, Data curation. **Siqi Bu:** Writing – review & editing, Supervision, Project administration.

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.

Acknowledgments

The authors gratefully acknowledge the substantial support from the Centre for Advances in Reliability and Safety (CAiRS) admitted under AIR@InnoHK Research Cluster, and the Innovation and Technology Commission (Project No. ITS/188/23) of the Hong Kong Special Administrative Region, China. Open access is supported by EPFL.

Data availability

Data will be made available on request.

References

[1] IEA. Electricity 2025 [online]. 2025, Available at: https://www.iea.org/reports/electricity-2025. (Accessed on 01 April 2025).
[2] Alkhayat G, Mehmood R. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. Energy AI 2021;4:100060.
[3] Chu Y, Wang Y, Yang D, Chen S, Li M. A review of distributed solar forecasting with remote sensing and deep learning. Renew Sustain Energy Rev 2024;198:114391.
[4] Kleissl J. Solar energy forecasting and resource assessment. Academic Press; 2013.
[5] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, et al. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. Renew Sustain Energy Rev 2022;161:112348.
[6] Chen S, Li C, Stull R, Li M. Improved satellite-based intra-day solar forecasting with a chain of deep learning models. Energy Convers Manage 2024;313:118598.
[7] Sengupta M, Habte A, Wilbert S, Gueymard C, Remund J. Best practices handbook for the collection and use of solar resource data for solar energy applications. Technical report, Golden, CO (United States): National Renewable Energy Lab.(NREL); 2021.
[8] Yu H, Chen S, Chu Y, Li M, Ding Y, Cui R, et al. Self-attention mechanism to enhance the generalizability of data-driven time-series prediction: A case study of intra-hour power forecasting of urban distributed photovoltaic systems. Appl Energy 2024;374:124007.
[9] Yang D, Wang W, Xia X. A concise overview on solar resource assessment and forecasting. Adv Atmospheric Sci 2022;1–13.
[10] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. Renew Sustain Energy Rev 2022;154:111768.
[11] Verdone A, Scardapane S, Panella M. Explainable spatio-temporal graph neural networks for multi-site photovoltaic energy production. Appl Energy 2024;353:122151.
[12] Hasnat MA, Asadi S, Alemazkoor N. A graph attention network framework for generalized-horizon multi-plant solar power generation forecasting using heterogeneous data. Renew Energy 2025;122520.
[13] Mayer MJ, Yang D. Optimal place to apply post-processing in the deterministic photovoltaic power forecasting workflow. Appl Energy 2024;371:123681.
[14] Markovics D, Mayer MJ. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew Sustain Energy Rev 2022;161:112364.
[15] Yang D, Kleissl J. Summarizing ensemble NWP forecasts for grid operators: Consistency, elicitability, and economic value. Int J Forecast 2023;39(4):1640–54.
[16] Sweeney C, Bessa RJ, Browell J, Pinson P. The future of forecasting for renewable energy. Wiley Interdiscip Rev: Energy Environ 2020;9(2):e365.
[17] Chu Y, Li M, Coimbra CFM, Feng D, Wang H. Intra-Hour Irradiance Forecasting Techniques for Solar Power Integration: A Review. IScience 2021;103136.
[18] Zhang Q, Zeng Z, Zhou Q, Ding X, Dong Z, Tian M, et al. Multimodal bidirectional MAMBA framework for ultra-short-term photovoltaic power forecasting using sky images. Renew Energy 2025;124361.
[19] Ruan Z, Wu L, Shi H, Ni M, Zhang M, Zhu W, et al. Spatiotemporal solar radiation forecasting driven by satellite-based and reanalysis data for distributed PV integration using fully-convolutional neural network. Renew Energy 2025;124556.
[20] Stull RB. Practical meteorology: an algebra-based survey of atmospheric science. University of British Columbia; 2015.

[21] Paletta Q, Terrén-Serrano G, Nie Y, Li B, Bieker J, Zhang W, et al. Advances in solar forecasting: Computer vision with deep learning. Adv Appl Energy 2023;100150.
[22] 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.
[23] Hamlyn D, Chaudhary S, Rahman T. Vision transformers for estimating irradiance using data scarce sky images. Energy AI 2025;100560.
[24] Quan H, Ge Y, Liu B, Zhang W, Srinivasan D. A deep-learning algorithm with two-stage training for solar forecast post-processing. Sol Energy 2024;273:112504.
[25] Lim B, Arık SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. Int J Forecast 2021;37(4):1748–64.
[26] Piantadosi G, Dutto S, Galli A, De Vito S, Sansone C, Di Francia G. Photovoltaic power forecasting: A Transformer based framework. Energy AI 2024;18:100444.
[27] Yang D, Wang W, Hong T. A historical weather forecast dataset from the European centre for medium-range weather forecasts (ECMWF) for energy forecasting. Sol Energy 2022;232:263–74.
[28] Pereira S, Canhoto P, Salgado R. Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data. Energy AI 2024;15:100314.
[29] Van Zyl C, Ye X, Naidoo R. Harnessing eXplainable artificial intelligence for feature selection in time series energy forecasting: A comparative analysis of Grad-CAM and SHAP. Appl Energy 2024;353:122079.
[30] Zhou H, Zheng P, Dong J, Liu J, Nakanishi Y. Interpretable feature selection and deep learning for short-term probabilistic PV power forecasting in buildings using local monitoring data. Appl Energy 2024;376:124271.
[31] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.
[32] Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. J Mach Learn Res 2019;20(177):1–81.
[33] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016, p. 1135–44.
[34] Liao W, Fang J, Ye L, Bak-Jensen B, Yang Z, Porte-Agel F. Can we trust explainable artificial intelligence in wind power forecasting? Appl Energy 2024;376:124273.
[35] Augustine JA, DeLuisi JJ, Long CN. SURFRAD–A national surface radiation budget network for atmospheric research. Bull Am Meteorol Soc 2000;81(10):2341–58.
[36] Reda I, Andreas A. Solar position algorithm for solar radiation applications. Sol Energy 2004;76(5):577–89.
[37] Yang D. SolarData: An R package for easy access of publicly available solar datasets. Sol Energy 2018;171:A3–12.
[38] Long CN, Shi Y. An automated quality assessment and control algorithm for surface radiation measurements. Open Atmospheric Sci J 2008;2(1).
[39] Chen S, Li M. Improved turbidity estimation from local meteorological data for solar resourcing and forecasting applications. Renew Energy 2022;189:259–72.
[40] Yang D, Wang W, Bright JM, Voyant C, Notton G, Zhang G, et al. Verifying operational intra-day solar forecasts from ECMWF and NOAA. Sol Energy 2022;236:743–55.
[41] Gueymard CA. REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation–Validation with a benchmark dataset. Sol Energy 2008;82(3):272–85.
[42] Benjamin SG, Weygandt SS, Brown JM, Hu M, Alexander CR, Smirnova TG, et al. A North American hourly assimilation and model forecast cycle: The rapid refresh. Mon Weather Rev 2016;144(4):1669–94.
[43] Sengupta M, Xie Y, Lopez A, Habte A, Maclaurin G, Shelby J. The National Solar Radiation Database (NSRDB). Renew Sustain Energy Rev 2018;89:51–60.
[44] Yang D. Validation of the 5-min irradiance from the National Solar Radiation Database (NSRDB). J Renew Sustain Energy 2021;13(1):016101.
[45] Chen S, Li C, Xie Y, Li M. Global and direct solar irradiance estimation using deep learning and selected spectral satellite images. Appl Energy 2023;352:121979.
[46] Ait Mouloud L, Kheldoun A, Deboucha A, Mekhilef S. Explainable forecasting of global horizontal irradiance over multiple time steps using temporal fusion transformer. J Renew Sustain Energy 2023;15(5).
[47] Lorenzo AT, Holmgren WF, Cronin AD. Irradiance forecasts based on an irradiance monitoring network, cloud motion, and spatial averaging. Sol Energy 2015;122:1158–69.
[48] Laguarda A, Alonso-Suárez R, Abal G. Improved estimation of hourly direct normal solar irradiation (DNI) using geostationary satellite visible channel images over moderate albedo areas. Sol Energy 2023;259:30–40.
[49] Jiang H, Lu N, Huang G, Yao L, Qin J, Liu H. Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data. Appl Energy 2020;270:115178.
[50] Matsunobu LM, Pedro HTC, Coimbra CFM. Cloud detection using convolutional neural networks on remote sensing images. Sol Energy 2021;230:1020–32.
[51] Liao Z, Coimbra CFM. Hybrid solar irradiance nowcasting and forecasting with the SCOPE method and convolutional neural networks. Renew Energy 2024;232:121055.
[52] Wright L. Ranger - a synergistic optimizer. 2019, https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer.
[53] Jing T, Chen S, Navarro-Alarcon D, Chu Y, Li M. SolarFusionNet: Enhanced solar irradiance forecasting via automated multi-modal feature selection and cross-modal fusion. IEEE Trans Sustain Energy 2025;16(2):761–73.
[54] Gandhi O, Zhang W, Kumar DS, Rodríguez-Gallegos CD, Yagli GM, Yang D, et al. The value of solar forecasts and the cost of their errors: A review. Renew Sustain Energy Rev 2024;189:113915.
[55] Bright JM, Sun X, Gueymard CA, Acord B, Wang P, Engerer NA. Bright-Sun: A globally applicable 1-min irradiance clear-sky detection model. Renew Sustain Energy Rev 2020;121:109706.
[56] Chen S, Liang Z, Dong P, Guo S, Li M. A transferable turbidity estimation method for estimating clear-sky solar irradiance. Renew Energy 2023;206:635–44.
[57] Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M. Timesnet: Temporal 2d-variation modeling for general time series analysis. 2022, arXiv preprint arXiv:2210.02186.
[58] Wang M, Rao C, Xiao X, Hu Z, Goh M. Efficient shrinkage temporal convolutional network model for photovoltaic power prediction. Energy 2024;297:131295.
[59] Makarov YV, Loutan C, Ma J, De Mello P. Operational impacts of wind generation on California power systems. IEEE Trans Power Syst 2009;24(2):1039–50.
[60] Yang D, Li W, Yagli GM, Srinivasan D. Operational solar forecasting for grid integration: Standards, challenges, and outlook. Sol Energy 2021;224:930–7.