

Research paper

Improved satellite-based intra-day solar forecasting with a chain of deep learning models

Shanlin Chen ^{a,b}, Chengxi Li ^c, Roland Stull ^b, Mengying Li ^{a,*}

^a Department of Mechanical Engineering & Research Centre on Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

^b Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, Canada

^c Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region



ARTICLE INFO

Keywords:

Solar irradiance forecasting
Spectral satellite data
Satellite-derived irradiance
Deep learning

ABSTRACT

Satellite data and satellite-derived irradiance products have been extensively used in solar forecasting to better capture the spatio-temporal variations of solar irradiance. However, the potential advantages of using satellite-derived irradiance and its improvements in solar forecasting have not been thoroughly explored. This work proposes a deep learning model chain with two models, one for deriving more accurate spatial global horizontal irradiance (GHI) estimates from satellite data, and the other for subsequently producing intra-day GHI forecasts using the derived spatial GHI. To evaluate the efficacy of the proposed method, GHI forecasts using different inputs are compared, namely, spectral satellite images (SAT), GHI estimates of the national solar radiation database (NSRDB), and satellite-derived GHI using deep learning (SAT-DL). The results show that satellite-derived irradiance products (NSRDB and SAT-DL) generally outperform SAT. The improved GHI estimates of SAT-DL yield forecasts with lower normalized root mean square error (nRMSE), higher forecast skill, better ramp forecasts and forecast distributions, when compared with NSRDB for the cases studied. However, forecasting under frequent cloudy conditions is found to have enlarged nRMSE and compromised performance in ramp analysis, and forecasts are biased under high- and low-irradiance conditions. Despite these challenges, the deep learning model chain approach provides a novel framework for satellite-based solar forecasting that can yield more accurate forecasts than the benchmark deep learning methods, which is beneficial to a wide range of stakeholders in the solar energy sector.

1. Introduction

Renewables are expected to account for over 90% of the global electricity expansion in the next years due to energy security concerns and climate ambitions. The total capacity of solar photovoltaic (PV) is set to surpass coal and become the world's largest power source over 2022–2027 [1]. Despite the aggressive acceleration in the installed capacity, the operation and integration of solar power still face challenges because of its intermittency and uncertainty [2,3]. To address the issues associated with solar variability, solar forecasting could be one of the solutions [4,5]. Indeed, solar forecasting could be cost-effective, with the aim to support system management and scheduling to meet the changing demand and thereby mitigate the variability of solar power by providing predictions up to a few days ahead [5]. Generally, solar forecasting refers to both solar irradiance forecasting and solar power forecasting [6]. As indicated by the names, the former is focused on irradiance quantities such as global horizontal irradiance (GHI) or direct normal irradiance (DNI), while the latter is related

to the power output of PV or concentrated solar power systems. It is paramount to have accurate irradiance forecasts to get solar power forecasts with high quality [7], using either a data-driven approach or a physical model chain [5,8].

In solar forecasting, one of the salient features of solar irradiance is the two-frequency pattern (i.e., yearly and diurnal cycles). Therefore, a clear-sky model that estimates the ground-level irradiance under cloud-free conditions is usually required to remove the seasonality [9,10]. The spatio-temporal nature, as another important feature of solar irradiance, should also be considered in solar forecasting models by integrating available spatio-temporal information [5]. In fact, solar forecasts based on spatio-temporal inputs have demonstrated to be more beneficial than forecasts based solely on local measurements [11, 12]. This performance enhancement stems from the improved representation of cloud dynamics, as clouds are the primary contributors to solar variability [13]. Three main methods can be used to obtain the spatio-temporal data in solar forecasting [5], namely, sky images

* Corresponding author.

E-mail address: mengying.li@polyu.edu.hk (M. Li).

Nomenclature**Abbreviations**

ABI	Advanced Baseline Imager
BiLSTM	Bidirectional long-short-term memory
BON	Bondville
CNN	Convolutional neural network
CSI	Clear-sky index
DHI	Diffuse horizontal irradiance
DNI	Direct normal irradiance
DRA	Desert Rock
FPK	Fort Peck
FRI	False ramp index
GAN	Generative adversarial network
GHI	Global horizontal irradiance
GHIcs	Clear-sky global horizontal irradiance
GOES	Geostationary Operational Environmental Satellites
GRU	Gated recurrent unit
GWN	Goodwin Creek
LSTM	Long-short-term memory
MBE	Mean bias error
nMBE	Normalized mean bias error
nRMSE	Normalized root mean squared error
NS	Irradiance forecasts using NSRDB
NSRDB	National Solar Radiation Database
NWP	Numerical weather prediction
PSU	Pennsylvania State University
PV	Photovoltaic
QC	Quality control
RDI	Ramp detection index
RMI	Ramp magnitude forecast index
RMSE	Root mean squared error
SAT	Spectral satellite images
SAT-DL	Satellite-derived irradiance using deep learning
SDL	Irradiance forecasts using SAT-DL
SP	Smart persistence
SURFRAD	Surface Radiation Budget Network
SXF	Sioux Falls
TBL	Table Mountain
UTC	Coordinated Universal Time

Notations

\bar{o}	Mean of observations
\mathbf{x}	Spatio-temporal input of satellite data
o	Degree
Δt	Forecast horizon
\hat{f}	CSI forecast
λ	Wavelength
\mathbb{E}	Deep learning satellite-to-irradiance model
\mathbb{F}	Deep learning solar irradiance forecasting model
\mathbb{P}	Physical satellite-to-irradiance model
θ_z	Solar zenith angle
\tilde{L}	Normalized value at each pixel of satellite images
f	Forecast

L	The intensity of a pixel in the satellite image
N	Number of data points
N_{FNR}	Number of false predicted ramps
N_{hit}	Number of detected ramps
N_{miss}	Number of missed ramps
N_r	Number of ramps
N_{TNR}	Number of true predicted no-ramps
o	Observation
$offset$	Add offset
R^2	Coefficient of determination
$radiance$	Radiance received by the satellite sensor
raw	Packed-scaled value of radiance
$scale$	Scale factor
t	Time

Subscripts

f	Forecast of interest
p	Perfect forecasts
r	Reference forecast
t	Time

Superscripts

b	Spectral band
-----	---------------

for intra-hour forecasting [4], satellite measurements for intra-day forecasting [14], and numerical weather prediction (NWP) data for day-ahead forecasting [5].

The utilization of remote-sensing data is one of the advanced aspects of intra-day solar forecasting [5]. Geostationary satellites, such as Geostationary Operational Environmental Satellites (GOES), Meteosat, and Himawari, collectively offer a coverage of all areas within latitudes of $\pm 60^\circ$. Modern geostationary satellites provide much finer data in both spatial and temporal resolutions with more spectral bands. For example, the GOES-16 satellite, with its 16 spectral channels, can provide data at a temporal resolution of 5-min, and the spatial resolution ranges from 0.5- to 2-km. The advancement in remote sensing technology has substantially contributed to the development of satellite-derived irradiance products, with the spatio-temporal resolution improved to 2-km and 5-min [15,16]. To this end, spatio-temporal satellite images and satellite-derived irradiance products have become essential data sources for solar forecasting [5,17,18].

1.1. Related work

Given the inherent spatio-temporal nature of solar irradiance, it is more likely that high-accuracy forecasts can be produced by incorporating spatio-temporal information [5]. On this account, satellite data and satellite-derived irradiance products have been extensively used in solar forecasting, either as the entire exclusive input or as a part of the inclusive exogenous inputs. Some insights regarding the use of satellite data and/or satellite-derived products for intra-day solar forecasting are summarized as follows.

Satellite data used in solar forecasting can be from one single channel [19,20], two visible channels [14], or multiple visible and infrared channels [21,22]. Satellite data of visible channels are primarily used to derived cloud index maps [14,19], which are used to predict future cloud information. The methods for cloud motion prediction can be physical (e.g., optical flow [14]) and data-driven (e.g., deep learning [23]); the former is based on cloud advection and extrapolation, while the latter is to forecast the cloud index maps with deep learning methods. The irradiance forecasts can then be obtained from the cloud

forecasts using the Heliosat method [14,24]. The usage of data from multiple spectral channels is to account for the modulation effect of clouds on solar radiation. In such cases, multiple spectral satellite data can be used by machine/deep learning models [21,25] to predict the irradiance directly or by a physical classification method [22] to determine the cloud type and subsequently generate irradiance forecasts.

Forecasting methods can be semi-empirical [14], deep learning based [12,21], or a hybridization of both [26]. The semi-empirical method mainly focuses on the determination of cloud dynamics using cloud motion analysis based on optical flow [14,24], and the clear-sky irradiance is quantified by a clear-sky model [14]. Deep learning can be applied to detect cloud motion and directly produce forecasts using data of multiple satellite channels [21] or satellite-derived irradiance over the target region [12]. The hybridized method combines deep learning and optical flow to extract feature maps of clouds using infrared satellite images [26].

As an important input for intra-day solar forecasting, the used satellite-derived irradiance products can be based on semi-empirical models [27] or physical models [28,29]. Similar to satellite data, satellite-derived irradiance can also provide spatio-temporal information in solar forecasting [27,28]. The semi-empirical satellite-to-irradiance method relies on the Heliosat model that empirically determines the cloud attenuation with historical satellite measurements [27]; In contrast, the physical method applies radiative transfer models to retrieve solar irradiance from satellite-derived atmospheric properties, exemplified by the National Solar Radiation Database (NSRDB) [28]. The use of satellite-derived irradiance products can avoid the complexities associated with estimating cloud properties from various satellite spectral channels [12]. Therefore, satellite-derived irradiance maps (e.g., NSRDB) can be used to predict future solar irradiance maps and PV power for an entire region [30].

Satellite data or satellite-derived irradiance products can be used as the single exogenous input [14,29] or a part of the inclusive inputs [27,31]. For example, satellite or satellite-derived data can be used as the only input to derive and predict future cloud field and thus solar irradiance [14,29], where solar power can be obtained using the irradiance-to-power conversion [30,32]. To further improve the forecasting performance, some other types of data including on-site measurements of solar irradiance or power [26,27], meteorological information [18], NWP products [27], and sky images [31] are used as additional inputs for solar forecasting.

Solar forecasts can be deterministic [21] or probabilistic [29]. So far, studies on solar forecasting are more focused on deterministic forecasts that offer single 'best-guess' values, either using satellite data [21] or satellite-derived irradiance products [12,27]. On the other hand, probabilistic forecasts that quantify the uncertainty in the forecasting, provide valuable information for solar energy projects and therefore have attracted more attention in the field [5]. Both satellite images and satellite-derived irradiance can be used to produce probabilistic forecasts [29,31].

1.2. Motivation and contributions

Satellite data and satellite-derived irradiance are widely used as inputs for intra-day solar forecasting with deep learning. However, when it comes to the selection of spatio-temporal inputs (i.e., satellite data or satellite-derived irradiance) for intra-day solar forecasting, there has been no clear determination regarding which type of spatio-temporal data has the potential to produce more accurate forecasts. For instance, Nielsen et al. [23] claimed that satellite data is advantageous compared to satellite-derived irradiance products, while Pérez et al. [12] suggested that satellite-derived irradiance products are more beneficial than satellite data from multiple spectral bands. Therefore, there has been a dearth of study to investigate what spatio-temporal inputs are more beneficial for intra-day solar forecasting.

Furthermore, in our recent study [16], spectral satellite images from GOES-16 and deep learning algorithms were applied to estimate both GHI and DNI with a spatio-temporal resolution of 2-km and 5-min at ground level. The results show substantial improvements in accuracy when compared with NSRDB [33], which is considered as the state-of-the-art in satellite-based solar irradiance estimation. Therefore, another research question arises as to whether the satellite-derived irradiance products with improved accuracy (i.e., satellite-derived irradiance by deep learning (SAT-DL)) could lead to more accurate solar forecasts. To address this question, a deep learning model chain method is proposed to improve intra-day solar forecasting using spectral satellite data. The deep learning model chain consists of two deep learning models: one is to obtain spatial GHI estimates with improved accuracy from spectral satellite images, and the other is to subsequently produce forecasts based on the improved spatial GHI estimates. The deep learning model chain provides a novel framework for satellite-based solar forecasting compared to the existing forecasting methods in the literature. The efficacy of the proposed method is evaluated by comparisons with an end-to-end deep learning model and a hybrid physical-deep learning model. The major contributions of this work are summarized as follows:

- The usage of spectral satellite data and satellite-derived irradiance products in deep learning-based intra-day solar forecasting are compared, at the spatio-temporal resolution of 2-km and 5-min. This comparison provides valuable insights into the selection of satellite-based spatio-temporal inputs for solar irradiance forecasting.
- The advantages of satellite-derived irradiance products with improved accuracy for solar irradiance forecasting are investigated and quantified. While initially developed for more accurate solar irradiance estimation at a single location, the deep learning method based on spectral satellite data [16] is extended for regional applications.
- A deep learning model chain is proposed for intra-day solar irradiance forecasting using spectral satellite data. This model chain comprises two models: one improves spatial GHI estimates from spectral satellite measurements, and the other subsequently generates GHI forecasts based on these improved estimates. The deep learning model chain is benchmarked against an end-to-end deep learning model that uses spectral satellite data, and a hybrid physical-deep learning model that employs spatial NSRDB estimates.
- The deep learning model chain approach provides a novel framework for satellite-based solar irradiance forecasting. This framework has the potential to incorporate more sophisticated deep learning architectures to further improve the accuracy of both satellite-derived irradiance products and intra-day solar irradiance forecasts.

The rest of this work is structured as follows: Section 2 describes the data and the methods, where data acquisition and pre-processing procedure are presented in Section 2.1, the utilization of the SAT-DL product is detailed in Section 2.2, the forecasting setup is presented in Section 2.3, and the performance evaluation methods for forecasts is presented in Section 2.4. The forecast results are quantitatively and qualitatively evaluated and compared in Section 3, and the implications are discussed in Section 4. Finally, Section 5 summarizes the key findings of this study.

2. Data and methods

This section describes the data and methods utilized for solar irradiance forecasting. As shown in the methodology flowchart (see Fig. 1), lagged spectral satellite images and two satellite-derived irradiance products are used as three alternative inputs for intra-day solar irradiance forecasting from 15 min to 180 min ahead. The raw satellite data are from several selected spectral bands of GOES-16, while datasets

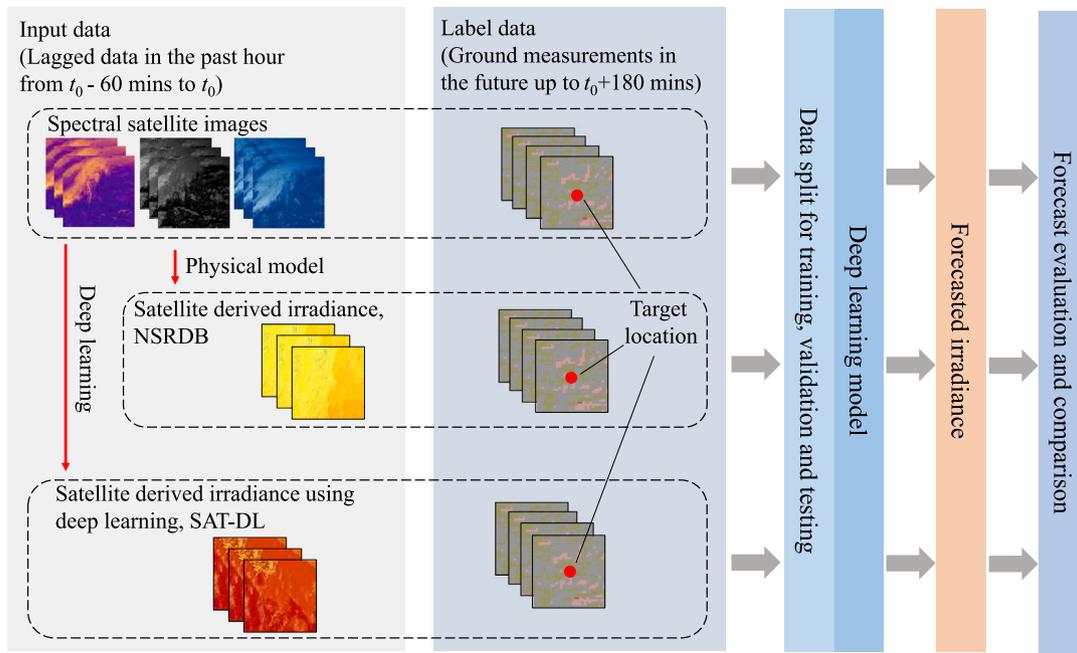


Fig. 1. The methodology flowchart of satellite-based solar irradiance forecasting with different methods.

Table 1

A summary of the publicly available data used in this work.

	Data type	Description	Time period (year)
SURFRAD ^a	Irradiance measurements	On-site measurements of solar irradiance	2019, 2020
GOES-16 ^b	Satellite measured radiance	Radiance of eight selected spectral bands	2019, 2020
NSRDB ^c	Derived ground-level irradiance	Satellite-derived irradiance with a physical model	2020

^a Available at <https://gml.noaa.gov/grad/surfrad/>, can be downloaded with the SolarData [35].

^b Available at <https://registry.opendata.aws/noaa-goes/>, can be download with the GOES-2-go [36].

^c Available at <https://nsrdb.nrel.gov/>, can be downloaded with the SolarData [35].

of NSRDB and SAT-DL are satellite-derived irradiance products from raw satellite measurements. The main difference between NSRDB and SAT-DL is that the former applies a physical solar model (radiative transfer model) to derive irradiance, while the latter employs a deep learning method. The labels are ground measurements (i.e., GHI) at the location centered in the domain of the satellite images. The inputs and corresponding labels are divided into subsets of training, validation, and testing to train and evaluate deep learning-based forecast model.

The accuracy of irradiance predictions using the three alternative inputs is evaluated against ground measurements. Note that SAT-DL is obtained from spectral satellite images with deep learning, therefore, using spatial SAT-DL irradiance estimates as inputs for another deep learning model forms a deep learning model chain. The end-to-end deep learning model using raw satellite measurements and the hybrid physical-deep learning model based on NSRDB are used as benchmarks. Detailed descriptions of data pre-processing, the deep learning model chain, and the development and optimization of the forecasting model are presented in the following subsections.

2.1. Publicly available data

Publicly available data used in this study include ground-level irradiance measurements, spectral satellite images, and satellite-derived irradiance products. Ground-level measurements are from the Surface Radiation Budget Network (SURFRAD) stations [34]. Satellite images are from GOES-16, but only a subset of spectral bands are selected [16]. Satellite-derived irradiance products are from NSRDB [33]. A brief summary of the publicly available data used is presented in Table 1.

2.1.1. Solar irradiance measurements from SURFRAD

SURFRAD is a network of seven stations located in five different climatological zones across the contiguous United States. As one of the radiation networks with the highest data quality in the world, SURFRAD has supported widespread applications and research since its establishment in 1995 [34]. Data from all seven stations are used, namely, Bondville (BON), Desert Rock (DRA), Fort Peck (FPK), Goodwin Creek (GWN), Pennsylvania State University (PSU), Sioux Falls (SXF), and Table Mountain (TBL). Since SURFRAD data have been extensively utilized and described in the literature and our previous works [10,15,37], the details will not be reproduced here.

Raw measurements, including GHI, DNI, diffuse horizontal irradiance (DHI), and solar zenith angle in 2019 and 2020 at all seven stations are downloaded. Although GHI is the forecasting target, other data of DNI, DHI, and solar zenith angle are required for quality control (QC). The pre-processing of SURFRAD data includes: (i) QC to remove unreasonable data points, (ii) data aggregation, and (iii) normalization, described as follows:

- QC: Several QC procedures are considered, including the extremely rare limit test and the three-component closure test [35]. These QC steps are based on the theoretical aspects detailed by Long and Shi [38].
- Data aggregation: After QC, the 1-min averaged GHI measurements are aggregated to 5-min intervals and indexed in Coordinated Universal Time (UTC). This is to be compatible with satellite data in the temporal resolution of 5-min. Due to the high air-mass effect at solar mornings and evenings, all GHI measurements for a solar zenith angle of 85° or greater are discarded [37].

Table 2
The detailed information of the selected ABI spectral bands of GOES-16 [40].

Band	λ [μm]	Center λ [μm]	Resolution [km]	Type	Valid range	Scale factor	Add offset
1	0.45–0.49	0.47	1	Visible	0–1022	0.8121	–25.9366
3	0.846–0.885	0.865	1	Near-Infrared	0–1022	0.3769	–20.2899
4	1.371–1.386	1.378	2	Near-Infrared	0–2046	0.0707	–4.5224
5	1.58–1.64	1.61	1	Near-Infrared	0–1022	0.0958	–3.0596
6	2.225–2.275	2.25	2	Near-Infrared	0–1022	0.0301	–0.9610
7	3.80–4.00	3.90	2	Infrared	0–16382	0.0016	–0.0376
9	6.75–7.15	6.95	2	Infrared	0–2046	0.0225	–0.8236
11	8.30–8.70	8.50	2	Infrared	0–4094	0.0334	–1.3022

- Normalization: It is suggested to remove the double-seasonal pattern of solar irradiance prior to forecasting [5]. On this point, the 5-min GHI measurements are normalized using clear-sky GHI (GHICs) estimations in NSRDB to generate clear-sky indexes (CSI = GHI/GHICs), which are used as labels in the forecasting model.

2.1.2. Spectral radiance data from GOES-16

GOES-16 is one of the GOES-R series geostationary satellites located at the operational longitude of 75.2°W during the investigated period of this study. The Advanced Baseline Imager (ABI) has 16 spectral bands (two visible, four near-infrared, and ten infrared channels), which monitor the Earth with much finer temporal and spatial resolutions compared with its predecessors. Among the available spectral bands, only a subset is selected as shown in Table 2. This is because some bands are highly correlated [16], and using only the representative satellite bands can improve the learning efficiency by eliminating redundant features [16]. 5-min GOES-16 images of eight selected spectral bands in 2019 and 2020 are extracted and georeferenced to the target SURFRAD stations. These images are sized at 11 × 11 pixels for 2019 and 21 × 21 pixels for 2020, respectively. The pre-processing of GOES-16 data includes radiance conversion and normalization:

- Radiance conversion: This process converts the packed-scaled value into radiance received by each band of GOES-16. The radiance for each band is calculated using the scale factor and add offset (see Table 2):

$$\text{radiance} = \text{raw} * \text{scale} + \text{offset}, \quad (1)$$

where *raw* is the packed-scaled value, *scale* is the scale factor, and *offset* is the add offset.

- Normalization: To remove diurnal effects and facilitate model training, the spectral radiance data of each band is further normalized by the band range (i.e., the minimum and maximum spectral responses) and solar zenith angle. This is adopted from [39] to weight each band equally:

$$\tilde{L}_{ij}^b|_t = 1 - \frac{L_{ij}^b|_t - L_{\min}^b}{\cos(\theta_z|_t)(L_{\max}^b - L_{\min}^b)}, \quad (2)$$

where $\tilde{L}_{ij}^b|_t$ is the normalized value at each pixel in band *b* at time *t*, $L_{ij}^b|_t$ is the measured radiance of that pixel, $\theta_z|_t$ is the solar zenith angle at time *t*, L_{\min}^b and L_{\max}^b are the minimum and maximum spectral responses of band *b*, respectively.

Note that the spectral bands have different spatial resolutions at the sub-satellite point as shown in Table 2. To ensure the consistency among selected bands, the bands with better resolutions are re-scaled as 2-km. As a multi-band passive imaging radiometer, the radiance received by ABI cannot theoretically be less than zero. Therefore, all negative values are discarded. The end timestamp of each scan in UTC time is used to index the image after rounding to the next nearest 5-min interval. This is done to be compatible with GHI measurements and to facilitate real-time applications.

2.1.3. Satellite-derived irradiance from NSRDB

The NSRDB is a widely accessed and publicly available database that provides satellite-derived solar irradiance over the United States and a growing number of international locations [33,41]. Data in NSRDB includes broadband solar radiation of GHI, DHI, DNI, and their clear-sky expectations. Meanwhile, there are also many other auxiliary variables, such as solar zenith angle, cloud type, and meteorological data. The NSRDB is serially complete with more than 20 years of coverage, providing data in 30-min temporal and 4 km spatial resolutions. Starting in 2018, the temporal and spatial resolutions of NSRDB are further improved to 5-min and 2-km, respectively [41]. The NSRDB is produced using a physical solar model, which is a two-step model to compute solar radiation from satellite data (e.g., GOES-16) and products of a number of other associations [33]. The validation of 5-min irradiance data of NSRDB against SURFRAD measurements shows that the new 5-min NSRDB has higher discrepancies due to its higher temporal resolution [42]. Nonetheless, it is still a remarkable milestone in solar irradiance modeling and resource assessment.

The NSRDB GHI values and their clear-sky expectations in 2020 are downloaded for 11 × 11 locations surrounding each SURFRAD station. The temporal resolution is 5-min and spatial resolution is 2-km. Clear-sky GHI estimations in NSRDB are calculated using the REST2 model [43], which has been repeatedly identified as one of the clear-sky models with high-performance [9,44]. The pre-processing of NSRDB mainly includes normalization. Similar to the removal of double-seasonal effects of irradiance measurements, the GHI estimates are normalized by the GHICs to generate the CSI, which is used as one of the alternative inputs. The data at the region of these 121 surrounding locations can provide spatio-temporal information for solar forecasting at the target stations.

2.2. Satellite-derived irradiance from SAT-DL

In our previous study [16], high spatio-temporal resolution spectral satellite images of GOES-16 were used to estimate ground-level GHI and DNI via deep learning. The results were verified against measurements at SURFRAD stations and showed better performance in estimating both GHI and DNI compared with NSRDB. Three main steps were involved: (i) mapping spectral satellite images of representative bands with ground observations, (ii) training and optimizing the deep learning model, and (iii) obtaining irradiance estimates from new satellite images. The deep learning model used in [16] employs convolutional neural networks (CNNs) [45], the attention mechanism [46] and fully-connected dense layers.

The original deep learning model in [16] was developed for ground irradiance estimates at a single location, which is centered in the domain of satellite images with 11 × 11 pixels. As shown in Fig. 2(a), the target is one of the SURFRAD stations (e.g., TBL), and the satellite images cover the same region with the surrounding locations. The target station can be anywhere as long as there are on-site irradiance measurements available. Following the same methodology, the target is expanded from one station to the 11 × 11 surrounding area with 121 locations. As shown in Fig. 2(b), selected spectral satellite images of GOES-16 (see Table 2) with the size of 21 × 21 pixels are used to obtain the GHI estimates for the whole region (11 × 11 pixels) via the

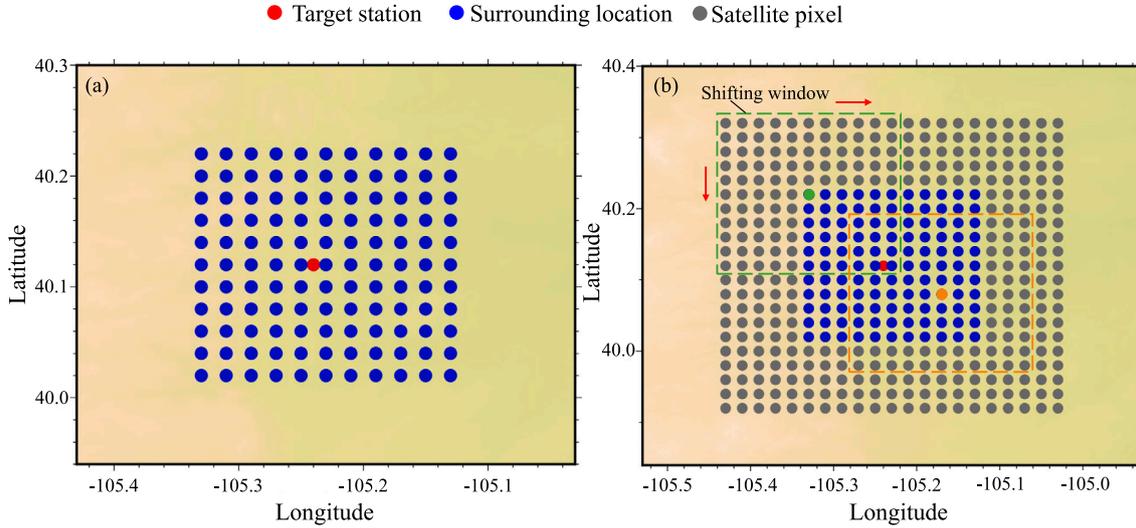


Fig. 2. An illustration of regional solar irradiance estimations for the Table Mountain (TBL) station using spectral satellite images. (a) The target station and 11×11 pixel grid of satellite images for single-station solar irradiance estimation. (b) The target station with 121 surrounding locations and the domain of used spectral satellite images for regional solar irradiance estimation.

pre-trained deep learning model. Specifically, datasets in 2019 are used to develop the pre-trained deep learning model for GHI estimation at a single SURFRAD station using spectral satellite images (11×11 pixels) and ground measurements. The pre-trained deep learning model is then applied to obtain GHI estimates in 2020 for the whole surrounding region (121 locations) by shifting an 11×11 -pixel window over the larger satellite images (21×21 pixels) as illustrated in Fig. 2(b). Note that the model's output is CSI of GHI normalized by the REST2 clear-sky model (available in NSRDB). The whole 11×11 matrices of CSI values are used as inputs for GHI forecasting at the target station. Therefore, CSI outputs are not converted back to GHI estimates at this step.

Solar forecasting based on the SAT-DL irradiance thus forms a deep learning model chain, i.e., spectral satellite data are converted to GHI estimates by one deep learning model [16], and the obtained GHI estimates are used subsequently to produce GHI forecasts using another deep learning model. Satellite images and ground measurements in 2019 are used to train and validate the deep learning model for GHI estimates (SAT-DL), while SAT-DL GHI estimates in 2020 are used as inputs for solar irradiance forecasting.

The method to generate GHI estimates for a single location is efficient to implement. Using the pre-trained deep learning model to obtain GHI estimates over a large region could be more time-consuming; therefore, larger areas around the target station are not considered. Also, the pre-trained deep learning model is developed only using ground data at the target stations, thus the accuracy might decrease with the increased distance, especially for locations that are too far away. Considering that the surrounding terrain and climate feature would not differ significantly for a certain location, it is reasonable to assume that GHI estimates of SAT-DL have lower discrepancies than those of NSRDB in the area around the SURFRAD stations. In a more general case, one could change the size of the surrounding area to investigate the effects on GHI estimates and forecasting, but this is not considered in this study.

2.3. Forecasting method

Deep learning is widely used in solar forecasting and resource assessment applications using satellite images and related products. The deep learning forecasting model applied in this work employs CNNs, the attention mechanism, and fully-connected dense layers based on Tensorflow [47]. Compared with the deep learning irradiance estimation model in [16], the main differences are the sources of inputs

and the number of outputs. As illustrated in Fig. 3, three different datasets are used as inputs to generate CSI forecasts for lead times ranging from 15 to 180 min. As mentioned, the deep learning model chain can obtain more accurate spatial GHI estimates from satellite data and subsequently produce GHI forecasts using the spatial GHI estimates of SAT-DL. The end-to-end deep learning model only employs spectral satellite data as inputs, while the hybrid physical-deep learning model applies GHI estimates of NSRDB as inputs. The inputs are lagged spectral satellite images or satellite-derived GHI matrices in the past hour with the spatio-temporal resolution of 2-km and 5-min, while the outputs are 5-min averaged CSI forecasts for various forecast horizons. For instance, the 5-min average means that the CSI forecast at $t = 12:00$ is the averaged value over times 11:58, 11:59, 12:00, 12:01, and 12:02 [42].

Note that the input sizes of satellite-derived GHI matrices and satellite images are different, the satellite-derived GHI matrices are in the size of 11×11 surrounding locations as shown in Fig. 2(a), while the spectral satellite images are in the size of 21×21 pixels as illustrated in Fig. 2(b). This is to perform a fair comparison among satellite images and derived GHI products, since GHI estimates of SAT-DL are based on the image size of 21×21 .

Deep learning can be used to produce multiple CSI forecasts (a multiple-output model) with forecast horizons (Δt) up to 180 min (i.e., 15, 30, 45, 60, 90, 120, 150, and 180-min). The end-to-end deep learning model can be expressed as:

$$\hat{I}_{t_0+15}, \hat{I}_{t_0+30}, \dots, \hat{I}_{t_0+180} = \mathbb{F}(\mathbf{x}_{t_0}, \mathbf{x}_{t_0-5}, \dots, \mathbf{x}_{t_0-60}), \quad (3)$$

where \hat{I} denotes the CSI forecast, which can be converted back to GHI by multiplying clear-sky GHI of REST2 at time $t + \Delta t$; \mathbb{F} is the forecasting model; \mathbf{x} represents the spatio-temporal spectral satellite measurements.

The hybrid physical-deep learning model is:

$$\hat{I}_{t_0+15}, \hat{I}_{t_0+30}, \dots, \hat{I}_{t_0+180} = \mathbb{P}(\mathbb{P}(\mathbf{x}_{t_0}, \mathbf{x}_{t_0-5}, \dots, \mathbf{x}_{t_0-60})), \quad (4)$$

where \mathbb{P} is the physical model used to convert spectral satellite measurements \mathbf{x} to spatial GHI estimations of NSRDB.

The deep learning model chain is then formulated as:

$$\hat{I}_{t_0+15}, \hat{I}_{t_0+30}, \dots, \hat{I}_{t_0+180} = \mathbb{F}(\mathbb{E}(\mathbf{x}_{t_0}, \mathbf{x}_{t_0-5}, \dots, \mathbf{x}_{t_0-60})), \quad (5)$$

where \mathbb{E} is the deep learning model to derive spatial GHI estimates of SAT-DL from spectral satellite measurements \mathbf{x} .

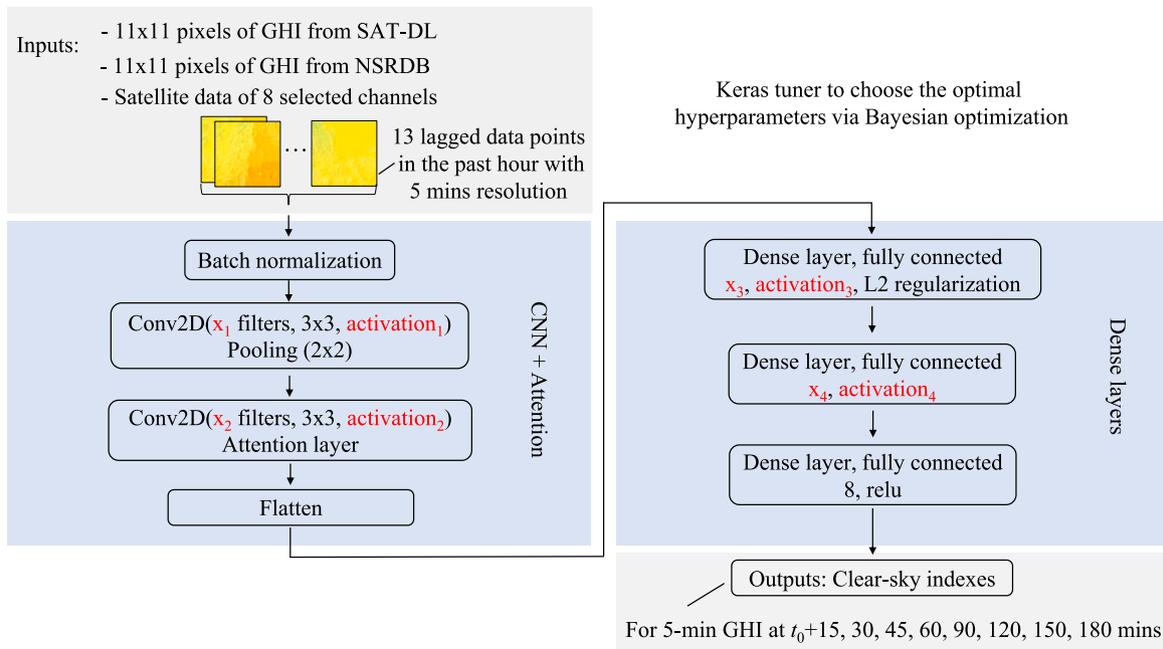


Fig. 3. The structure of deep learning models for multiple CSI forecasts using spectral satellite images or satellite-derived GHI products. The hyperparameters are optimized using the KerasTuner, as shown in Table 3.

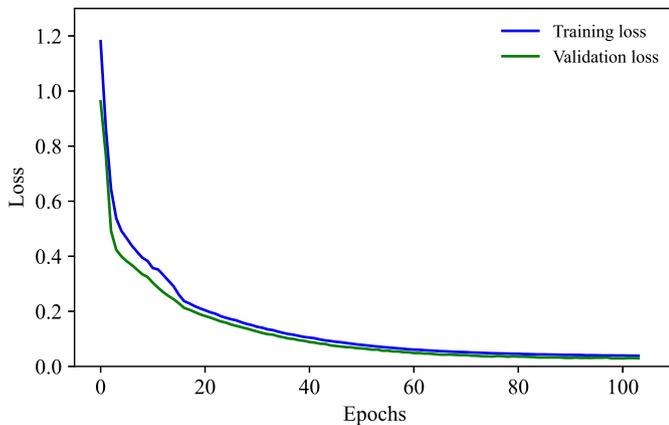


Fig. 4. The loss curves of the training and validation processes for GHI forecasting based on satellite-derived irradiance using deep learning (SAT-DL) at the BON station.

Although the structure of the deep learning model \mathbb{F} is similar in different forecasting scenarios, the hyperparameters are different for each spatio-temporal input. Forecasting models for each SURFRAD station with different inputs are trained separately, and the optimal hyperparameters are obtained using Bayesian optimization via the KerasTuner [48] as shown in Table 3. To avoid overfitting and improve the generalization ability of the forecasting models, L2 regularization and Early Stopping techniques are applied as shown in Fig. 3 and Table 3. Fig. 4 presents the loss curves during the training and validation processes for GHI forecasting using SAT-DL estimates at the BON station. Note that the validation loss is less than the training loss due to the L2 regularization. Since the forecasting model is developed using data from year 2020, to better represent the yearly variability, data in March, June, September, and December are utilized for testing, while the rest of the datasets in 2020 are used as training and validation subsets (in which 20% of the data is used for validation).

A diverse range of deep learning architectures, such as the long-short-term memory (LSTM) network, transformers, and generative adversarial networks (GANs), are suitable for solar forecasting with

Table 3

Hyperparameters shown in Fig. 3 for Bayesian optimization using the KerasTuner.

Hyperparameter	Values
Optimizer	Adam
Loss function	Huber
Learning rate	[1e-5, 1e-4, 2e-4, 5e-4, 1e-3, 2e-3, 1e-2]
x_1	range(min = 16, max = 128, step = 2)
activation ₁	[relu, gelu, selu, tanh]
x_2	range(min = 8, max = 64, step = 2)
activation ₂	[relu, gelu, selu, tanh]
x_3	range(min = 8, max = 64, step = 2)
activation ₃	[relu, gelu, selu, tanh]
x_4	range(min = 8, max = 64, step = 2)
activation ₄	[relu, gelu, selu, linear]
Early Stopping	With the patience equal to 5

spatio-temporal inputs [49]. More sophisticated deep learning networks may indeed capture spatio-temporal features more effectively, potentially enhancing the forecast accuracy. However, the focus of this study is not to develop an advanced deep learning model architecture but to evaluate the effectiveness of various spatio-temporal inputs for intra-day solar forecasting. Therefore, CNNs have been selected for their well-established ability to process spatial data. This choice aims to maintain a consistent and efficient computational framework that facilitates a fair comparison of spatio-temporal inputs, thereby isolating the input variable as the primary subject of investigation. Nevertheless, the performance of the proposed deep learning model chain is further evaluated through comparisons with several other deep learning methods for satellite-based solar forecasting, including CNN-LSTM, LSTM, bidirectional LSTM (BiLSTM), and gated recurrent unit (GRU). Note that the proposed deep learning model chain is also designed to establish a foundational framework for satellite-based solar forecasting. This approach allows for the possibility of integrating more advanced deep learning algorithms (such as GAN) into subsequent research to potentially improve the forecasting accuracy further [21,50].

2.4. Performance evaluation of forecasts

As mentioned, the model outputs are CSI forecasts, which are converted back to irradiance by multiplying the clear-sky irradiance at the

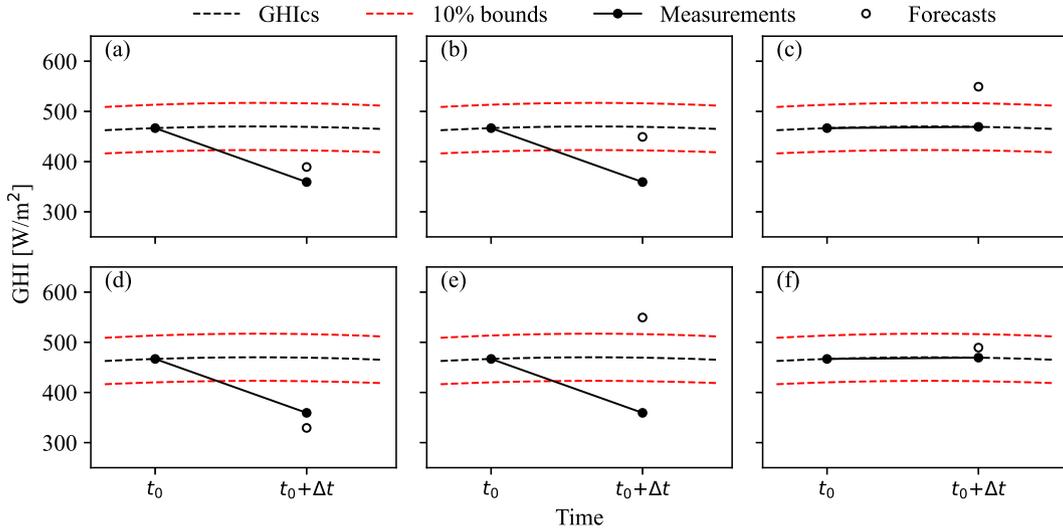


Fig. 5. Some possible outputs of ramp forecasts. GHICs denotes the clear-sky irradiance, 10% bounds define the upper and lower bounds on the threshold, Δt is the forecast horizon.

Source: Modified based on Fig. 2 in [51].

predicted time stamps. Consequently, the final error evaluations and visualizations are expressed in terms of irradiance [W/m^2]. The evaluation metrics employed include the root mean squared error (RMSE) and mean bias error (MBE), along with their normalized forms (nRMSE and nMBE), to assess the overall forecasting accuracy. Additionally, the coefficient of determination (R^2) is utilized to determine how accurately the model reproduces the observed data. The evaluation metrics are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum (f_i - o_i)^2}, \quad (6)$$

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{N} \sum (f_i - o_i)^2}}{\bar{o}_i}, \quad (7)$$

$$\text{MBE} = \frac{1}{N} \sum (f_i - o_i), \quad (8)$$

$$\text{nMBE} = \frac{\sum (f_i - o_i)}{\sum o_i}, \quad (9)$$

$$R^2 = 1 - \frac{\sum (f_i - o_i)^2}{\sum (o_i - \bar{o}_i)^2}, \quad (10)$$

where f_i and o_i are the pairs of irradiance forecasts and ground observations (i.e., GHI), \bar{o}_i is the mean of the observations, N is the total number of data points compared.

The forecast skill is used to provide a relative measure of improvement for a model's prediction over that of a reference model, which can be calculated based on RMSE:

$$\text{FS} = \left(1 - \frac{\text{RMSE}_f}{\text{RMSE}_r}\right) \times 100\%, \quad (11)$$

where FS is the forecast skill, RMSE_f is based on forecasts of the evaluated model, and RMSE_r is calculated using forecasts of a reference model. In this work, smart persistence is used as the reference model, which assumes the CSI at the current time t remains unchanged over the forecast horizon, as defined by:

$$\text{CSI}_{t+\Delta t} = \text{CSI}_t, \quad (12)$$

where Δt is the forecast horizon. The GHI forecast is then obtained by multiplying the clear-sky irradiance at $t + \Delta t$.

Solar irradiance exhibits huge variability with the rapid change of clouds in the sky. The ramp analysis is used to evaluate the model's performance in capturing the ramp events in irradiance, which can

better support the integration of solar energy. A ramp event is defined by the irradiance change in the time interval of $[t, t + \Delta t]$ that exceeds a threshold (e.g., 10% of current clear-sky irradiance [51]). Some possible outputs of ramp forecasts are shown in Fig. 5. Following the definitions in [51], three metrics are defined in the ramp analysis, namely, ramp detection index (RDI), false ramp index (FRI), and ramp magnitude forecast index (RMI) as defined by:

$$\text{RDI} = \frac{N_{\text{hit}}}{N_{\text{hit}} + N_{\text{miss}}}, \quad (13)$$

$$\text{FRI} = \frac{N_{\text{FRP}}}{N_{\text{FRP}} + N_{\text{TNR}}}, \quad (14)$$

$$\text{RMI} = 1 - \sqrt{\frac{\sum_{i=1}^{N_r} (o_{i+\Delta t} - f_{i+\Delta t})^2}{\sum_{i=1}^{N_r} (o_{i+\Delta t} - o_i)^2}}, \quad (15)$$

where N_{hit} means the number of cases when ramps are detected (cases (a) and (d) in Fig. 5), N_{miss} means the number of cases when ramps are missed (cases (b) and (e) in Fig. 5), N_{FRP} means the number of cases when ramp is forecasted but there is no ramp (Fig. 5(c)), N_{TNR} means the number of cases when ramp is not forecasted and is actually not presented (Fig. 5(f)), N_r is the number of ramp events, o is the observation and f is the forecast. More details are available in [51].

3. Results

In this section, the performance of GHI forecasts using the developed deep learning model is evaluated and compared with different inputs, namely, spectral satellite images (the end-to-end deep learning model), GHI estimates from NSRDB (the hybrid physical-deep learning model), and GHI estimates of SAT-DL (the deep learning model chain). Three categories of evaluation are adopted: (i) forecasting accuracy in terms of nRMSE and nMBE is presented in Section 3.1, along with the comparison of R^2 and forecast skill, (ii) the ramp analysis of irradiance forecasts including ramp detection and magnitude forecasts, is elaborated in Section 3.2, and (iii) distribution-oriented approaches for verifying irradiance forecasts are detailed in Section 3.3. Meanwhile, Section 3.4 details the comparison with several other deep learning methods to further evaluate the performance of the proposed deep learning model chain.

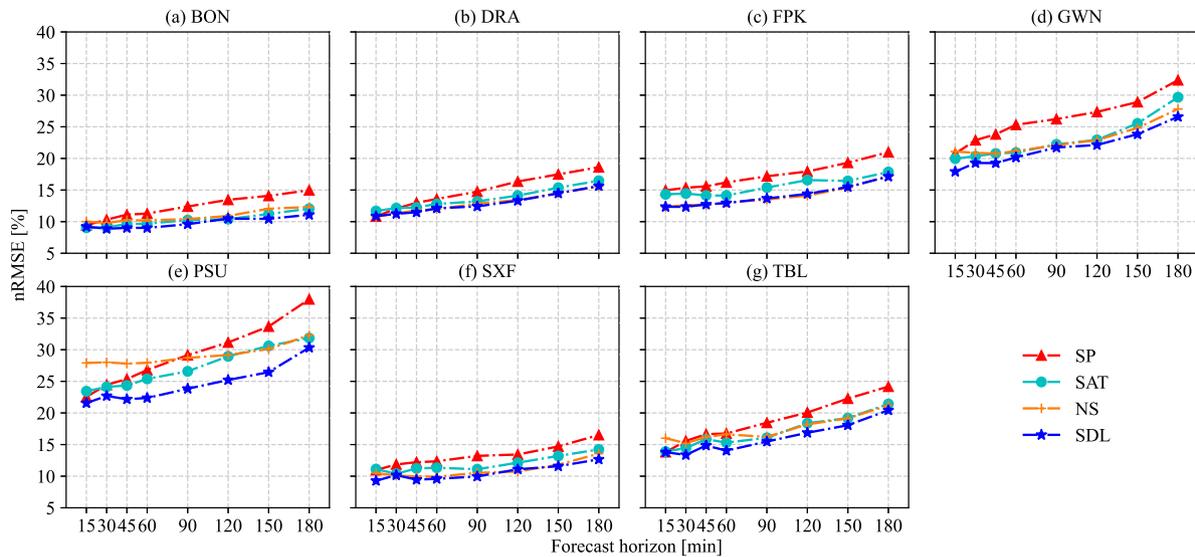


Fig. 6. The nRMSE [%] of GHI forecasts up to 180 min at all the SURFRAD stations with different methods: smart persistence (SP), spectral satellite images (SAT), NSRDB GHI estimates (NS), GHI estimates of SAT-DL (SDL).

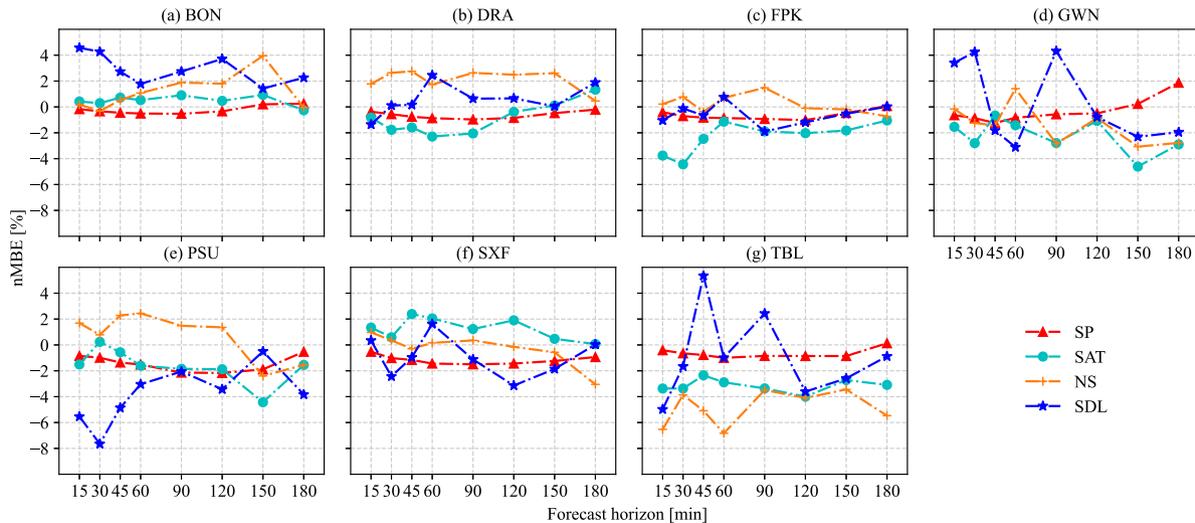


Fig. 7. Same as Fig. 6, but for the comparison for nMBE [%] of GHI forecasts.

3.1. Forecast evaluation in terms of statistical metrics

The results of intra-day GHI forecasts using different spatio-temporal inputs at all SURFRAD stations are presented in Fig. 6 (for nRMSE), Fig. 7 (for nMBE), and Fig. 8 (for R^2). It is shown in Fig. 6 that the forecasting nRMSE generally increases with the extension of the forecast horizon regardless of used inputs. Deep learning models generally outperform the smart persistence model (SP) in terms of nRMSE, especially for longer forecast horizons (e.g., those longer than 60 min). Furthermore, using satellite-derived GHI products (NS and SDL) typically yield better forecasts than using raw satellite images (SAT) in most cases, even though the region of raw satellite images is much larger (see Fig. 2). When comparing the forecast results based on satellite-derived irradiance products, using GHI estimates of SAT-DL typically produce forecasts with lower nRMSE than using those from NSRDB, as reflected by the comparison between NS and SDL in Fig. 6 for all SURFRAD stations.

Although satellite-derived products generally yield better GHI forecasts (in terms of nRMSE) than the forecasts based on raw satellite images, there are some site-specific differences across the SURFRAD

stations. For instance, at PSU, the nRMSE values of GHI forecasts in all forecast horizons are larger than at other stations, and SAT are more accurate than NS in most conditions. However, SDL still yields the best results at PSU. When comparing the nMBE of GHI forecasts (see Fig. 7), apart from some site-specific divergences, there are also some differences across the forecast horizon, but no obvious trends are observed. In general, satellite-irradiance-based forecasts (NS and SDL) tend to have larger biases.

Among all the SURFRAD stations, GHI forecasts at GWN and PSU have comparatively larger nRMSE (see Fig. 6). Similarly, it is shown in Fig. 8 that the R^2 values of GHI forecasts at GWN and PSU are relatively lower, regardless of the forecasting methods used. The R^2 values tend to increase with the extending forecast horizon while there are some fluctuations, especially at GWN, PSU, and TBL. When comparing between the forecasting methods/inputs, GHI forecasts using satellite-derived irradiance products (NS and SDL) generally have higher R^2 values than the forecasts using raw satellite images (SAT). Furthermore, SDL shows superior overall performance when compared to both NS and SAT.

GHI forecast skills relative to the smart persistence model with different inputs are further compared. As presented in Fig. 9, forecast

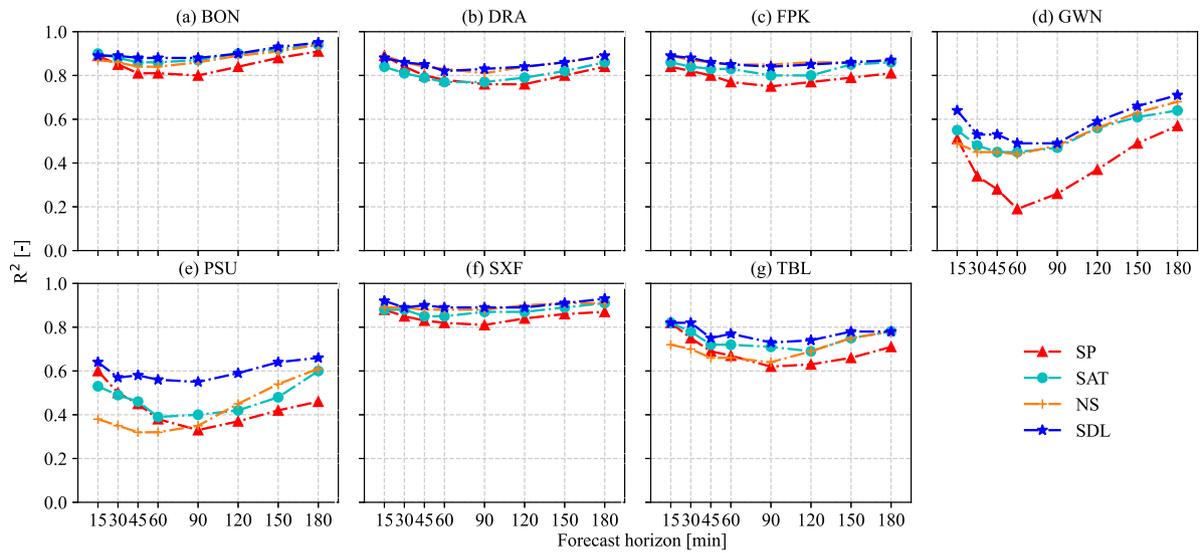


Fig. 8. Same as Fig. 6, but for the comparison for R^2 of GHI forecasts.

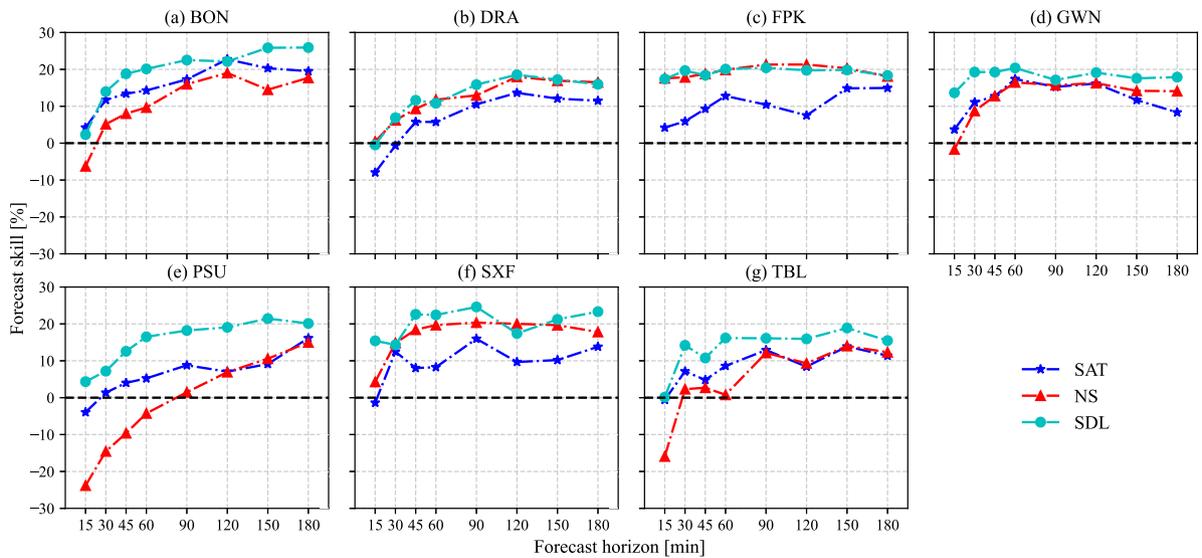


Fig. 9. Comparison of GHI forecast skills using different inputs across the forecast horizons up to 180 min at all the SURFRAD stations. A larger skill indicates a more accurate forecast.

skill generally improves with the extension of forecast horizon no matter which input is used. The increase of forecast skill is more obvious for shorter forecast horizons (i.e., less than 60 min); while in longer forecast horizons (e.g., longer than 60 min), fluctuations and even decreases in skill can be observed. When comparing the forecast skill when different inputs are used, SDL generally outperforms NS and SAT at all the SURFRAD stations. It is observed that SDL also has higher robustness compared to NS, as evidenced by the inferior performance of NS at BON, TBL, and especially PSU (see Fig. 9).

3.2. Forecast evaluation in terms of ramp metrics

Since solar variability introduces great difficulties to the grid integration of solar power systems, especially when sudden fluctuations present, it is also beneficial to perform ramp analysis of irradiance forecasts to evaluate the model's ability in predicting the variability. As defined in Section 2.4, three indexes are used in ramp analysis, namely, RDI to identify the successful rate of ramp detection, FRI to evaluate the false ramp events, and RMI to quantify the magnitude of ramp predictions.

Table 4

The amount of clear and cloudy periods (in percentage) in 2020 at all the SURFRAD stations. The clear and cloudy conditions are labeled based on 1-min QC data and the Bright-Sun clear-sky detection model [52].

	BON	DRA	FPK	GWN	PSU	SXF	TBL
Clear periods [%]	27.83	49.88	22.89	27.44	14.64	28.50	29.38
Cloudy periods [%]	72.17	50.12	77.11	72.56	85.36	71.50	70.62

Table 4 shows the percentage of clear and cloud periods in 2020 at all the SURDRAD stations. It is shown that DRA has the most occurrence of clear conditions, PSU has the most occurrence of cloudy conditions, while the remaining sites show similar distributions of clear and cloudy periods. It has also been observed in many studies that DRA has a higher occurrence of clear periods, while PSU shows the most frequent cloudy skies [37,39,53]. Therefore, the ramp analysis is conducted exclusively at DRA, PSU, and SXF, which have been chosen to represent a variety of cloudy conditions among the SURFRAD stations, as illustrated in Fig. 10.

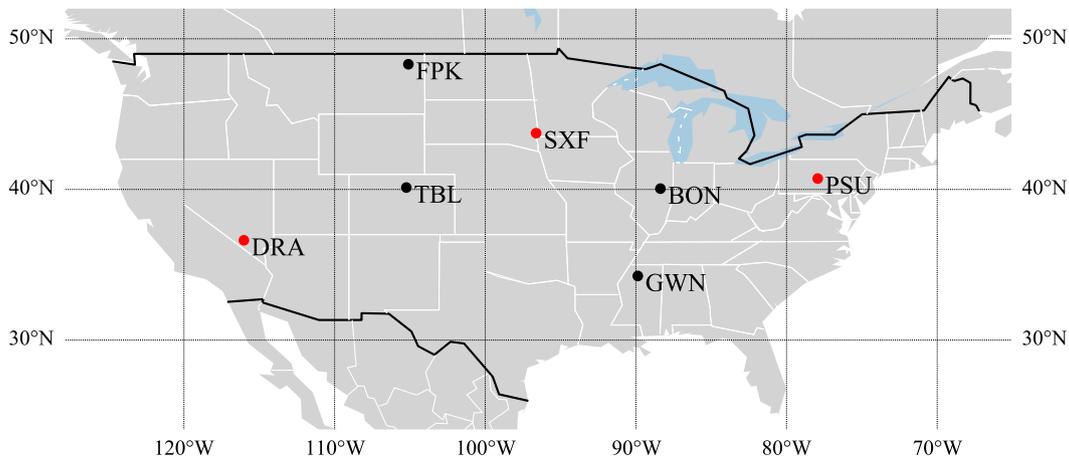


Fig. 10. The selected representative stations of SURFRAD. DRA has the most clear periods, PSU has the most cloudy skies, and SXF is selected as it is in the geographical ‘middle’ of the rest stations.

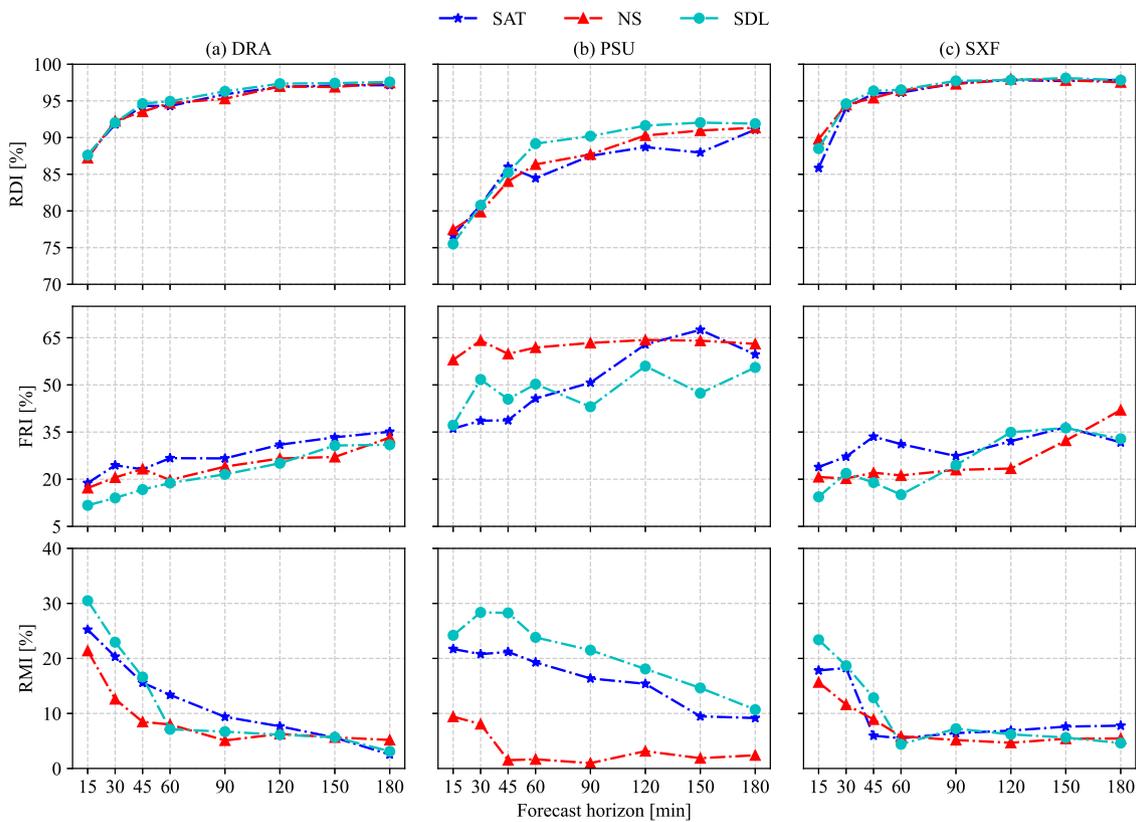


Fig. 11. The ramp analysis for GHI forecasts using RDI, FRI and RMI at (a) DRA, (b) PSU and (c) SXF. For RDI and RMI, values closer to 100% are better. For FRI, smaller is better.

As shown in Fig. 11, both RDI and FRI typically increase while the RMI decreases with the extended forecast horizons. When cloudy conditions become more frequent, the forecasting method generally exhibits increased uncertainty in predicting ramp events. Specifically, the RDI at PSU is lower than DRA and SXF, while the FRI at PSU is comparatively larger. As for ramp magnitude forecasts, the RMI at PSU is relatively higher than DRA and SXF, especially for longer forecast horizons (except the use of NSRDB GHI as inputs, which will be discussed in Section 4).

When comparing the inputs used for ramp forecasts, none shows dominant performance over the others (see Fig. 11). All inputs exhibit similar performance for ramp detection at DRA and SXF, while SDL

typically has a higher RDI at PSU, especially for longer forecast horizons. Regarding the false ramps detection, DRA does not present huge difference across the forecast horizons, while SXF shows some fluctuations, but the general performance is still comparable. In contrast, at PSU, NS tends to generate a relatively higher rate of false ramps across all forecast horizons. Similarly, the ramp magnitude forecasts at DRA and SXF do not differ greatly, while at PSU, NS could not produce equivalent results to those of the other two types of inputs.

3.3. Distribution-oriented forecast evaluation

As suggested by Yang et al. [53], evaluation based solely on accuracy is not always inclusive. Therefore, other aspects of forecasting

quality, for instance, distribution-oriented approaches, should be investigated. In this section, the GHI forecasts based on three different inputs (SAT, NS, and SDL) are verified against observations using different distribution-oriented approaches, including joint, marginal, and conditional distributions. The joint distribution enables the identification of obvious outliers and unnatural patterns, while the verification of marginal and conditional distributions are equivalent. More details about joint, marginal, and conditional distributions can be found in [53, 54].

The joint and marginal distributions of GHI forecasts based on three different inputs (i.e., SAT, NS, and SDL) against the measurements at three SURFRAD stations (i.e., DRA, PSU, and SXF) over different forecast horizons are shown in Fig. 12. Compared with PSU, the joint distributions at DRA and SXF have higher probabilities around the identity line, no matter which inputs are used. In addition to the higher probability density under the diagonal, GHI forecasts at PSU also show sparser distributions compared with those at DRA and SXF, which indicates that larger divergences are presented (see Fig. 6). The histograms in Fig. 12 represent marginal distributions (with forecasts on the right and measurements on the top). If the forecasts had no errors, the two marginal distributions would be the same. Fig. 12 shows different extents of discrepancies observed for various scenarios regarding different locations and forecast horizons. When comparing the inputs used for GHI forecasts, predictions based on SDL are distributed closer to the identity line than those based on SAT and NS, especially at PSU. Moreover, at PSU, NS tends to produce larger forecast errors, as evidenced by the more scattered density distributions. The conditional distribution of GHI forecasts given observations at DRA, PSU, and SXF are shown in Fig. 13, where forecasts are based on three different inputs (i.e., SAT, NS, and SDL). Forecasts are considered synchronized if the centroids of the distributions align with the identity line [53]. It can be seen in Fig. 13 that the distributions of GHI forecasts at DRA and SXF are more centered than those at PSU no matter which input is used. Generally, the forecasts are associated with positive bias in the low-irradiance range and tend to produce under-predictions for high-irradiance conditions. It is more obvious to notice the over-prediction (for low-irradiance conditions) and under-prediction (for high-irradiance conditions) at PSU. When comparing the forecasts based on different inputs, the results at DRA and SXF are similar, the centroids of GHI forecasts from SDL are slightly closer to the identity line compared with those from SAT and NS. However, at PSU, forecasts based on NS show more divergences than those from SDL, with wider forecast distributions for given measurements and the sparser centroids dispersed around the identity line.

3.4. Comparison with other deep learning models

To further evaluate the performance of the deep learning model chain for satellite-based solar forecasting, a comparative analysis with other deep learning methods using the same inputs is conducted. The compared methods include LSTM, BiLSTM, CNN-LSTM, and GRU, and the used inputs are spectral satellite images. The involved SURFRAD stations are DRA, PSU, and SXF to represent different cloudy conditions (see Section 3.2). As presented in Fig. 14, SDL and CNN-LSTM, where CNNs are adopted to process the spatio-temporal inputs, generally outperform LSTM, BiLSTM, and GRU. SDL and CNN-LSTM usually higher forecast skills at all the three stations, while the R^2 values of SDL and CNN-LSTM are higher at PSU, and comparable R^2 values are observed at DRA and SXF. Among the compared methods, SDL shows relatively better performance under different sky conditions at DRA, PSU, and SXF. For clear (at DRA) and low cloudy (at SXF) sky conditions, SDL typically has higher forecast skills with some fluctuations, and the divergences of R^2 are not significant. When there are more clouds (at PSU) in the sky, the effectiveness of the deep learning model chain becomes more obvious, with comparatively higher forecast skill and R^2 across all the forecast horizons.

4. Discussion

Based on the evaluation of GHI forecasts in terms of statistical metrics (Section 3.1), NS and SDL generally show lower forecast discrepancies than SAT, as shown in Fig. 6. However, deriving irradiance products from spectral satellite images, whether using physical models or deep learning methods, requires extra computation time, which should be considered in operational forecasting applications. Given that GHI forecasts from SDL have comparatively lower errors of nRMSE (Fig. 6) and higher R^2 values (Fig. 8) than those from NS, the proposed deep learning model chain, which maps satellite data to spatial irradiance and then to irradiance forecasts, is effective in improving intra-day GHI forecasts. Regarding to the nMBE, GHI forecasts based on NS and SDL have larger biases when compared to those from SAT, as shown in Fig. 7. This could be due to that the loss function used during the training process does not adequately account for the bias. Despite the enlarged bias, satellite-derived irradiance products are still valuable inputs for solar irradiance forecasting, since the bias could be further corrected and refined using post-processing methods [55].

GHI forecasts from SDL have higher forecast skill and higher robustness compared to those from SAT and NS (see Fig. 9). According to the assumption made in Section 2.2 that spatial GHI estimates of SAT-DL exhibit lower discrepancy than those of NSRDB in the surrounding region of a target station, one implication here is that using surrounding spatial information (i.e., GHI estimates) with lower uncertainty could improve the forecasting performance at the target station. That said, deep learning itself can be a feature extraction tool, and using the extracted features (SAT-DL GHI estimates) in another deep learning model could produce better solar irradiance forecasts than the end-to-end model (directly from satellite data to irradiance forecasts) and the hybrid physical-deep learning model (based on NSRDB). Since better forecasts can be produced using spatial GHI estimates with reduced uncertainty, it is beneficial to improve the accuracy and spatio-temporal resolutions of satellite-derived irradiance products.

In terms of ramp analysis, as presented in Fig. 11, the increased detection of ramps and false ramps means that the forecasts of ramp events for longer time horizons are relatively easier, but are also associated with a higher likelihood of false ramp forecasts. This is because the data-driven methods produce forecasting only referring past observations, the uncertainty would increase for longer forecast horizons. The integration of NWP products as future-known inputs could be helpful. The decrease in ramp magnitude forecasts indicates that it is possible to detect ramp events for longer time horizons, but the accurate forecast of ramp magnitude is much more challenging. The compromised RDI performance at PSU indicates that the ramp forecasts for locations with frequent cloudy skies may have larger unpredictability. As for the higher RMI at PSU, the reason could be the greater attenuation of irradiance by more frequent clouds, which leads to a lower irradiance magnitude. Therefore, the ramp magnitude forecasts at PSU perform better than those at DRA and SXF.

When it comes to distribution-oriented evaluation, the distributions of forecasts at DRA and SXF are more closely aligned along the identity line (see Figs. 12 and 13), which means that the GHI forecasts are more calibrated at DRA and SXF than those at PSU. When comparing the results at each location, the distributions of SDL forecasts are more centered than those of SAT and NS, especially at PSU. The conditional distributions in Fig. 13 reveal that the GHI forecasts for more frequent cloudy skies (at PSU) are more challenging compared to moderate and low cloudy conditions (at SXF and DRA) regardless of the input used. Nevertheless, the deep learning model chain can produce better overall forecasts than SAT, NS, and a bunch of other deep learning methods including CNN-LSTM, LSTM, BiLSTM, and GRU, especially when more clouds are present.

It is found that the forecasting methods have compromised performance when more frequent clouds are present. For instance, at PSU, higher errors of nRMSE, inferior performance in ramp forecasts, and

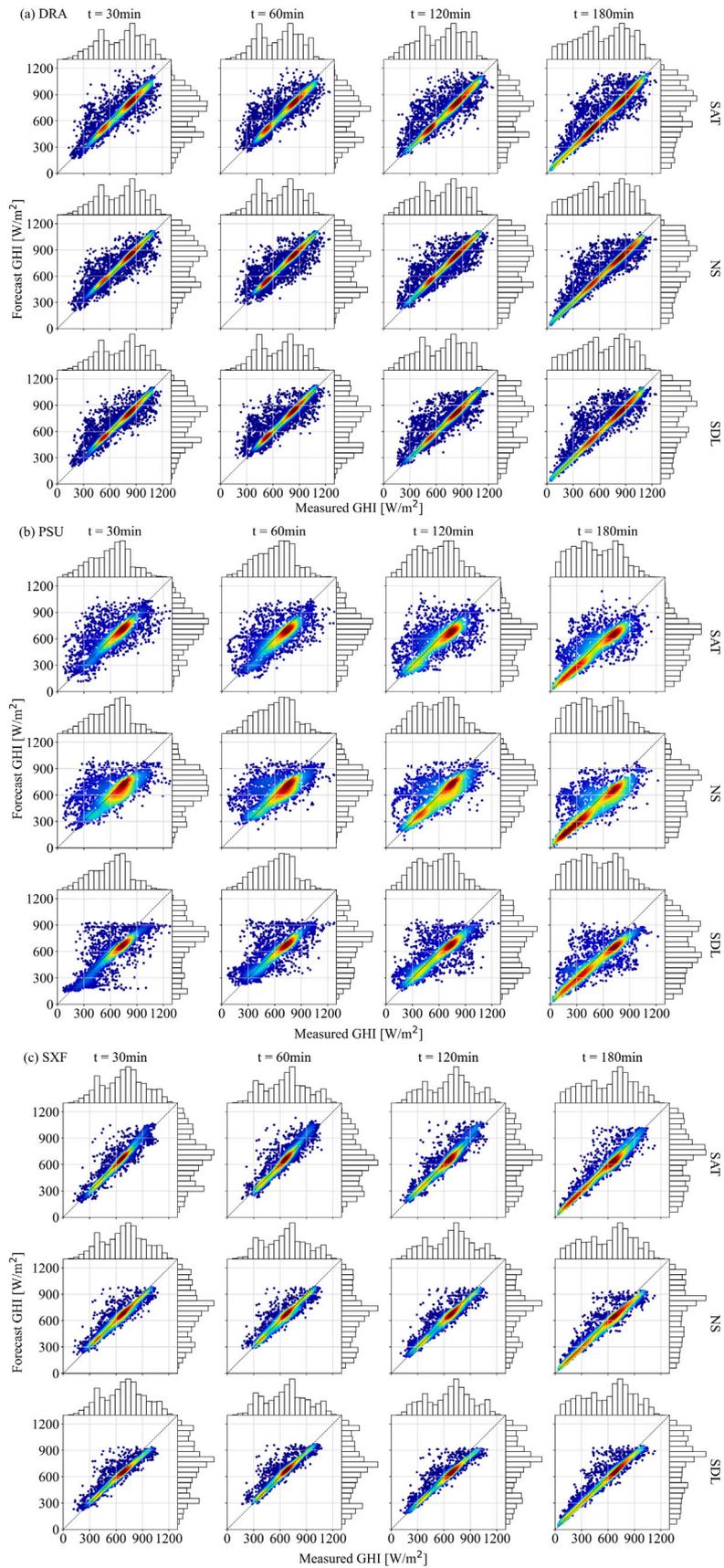


Fig. 12. Joint and marginal distributions of GHI forecasts and measurements at (a) DRA, (b) PSU, and (c) SXF. GHI forecasts are based on raw satellite data (SAT), satellite-derived GHI estimates of NSRDB (NS) and SAT-DL (SDL).

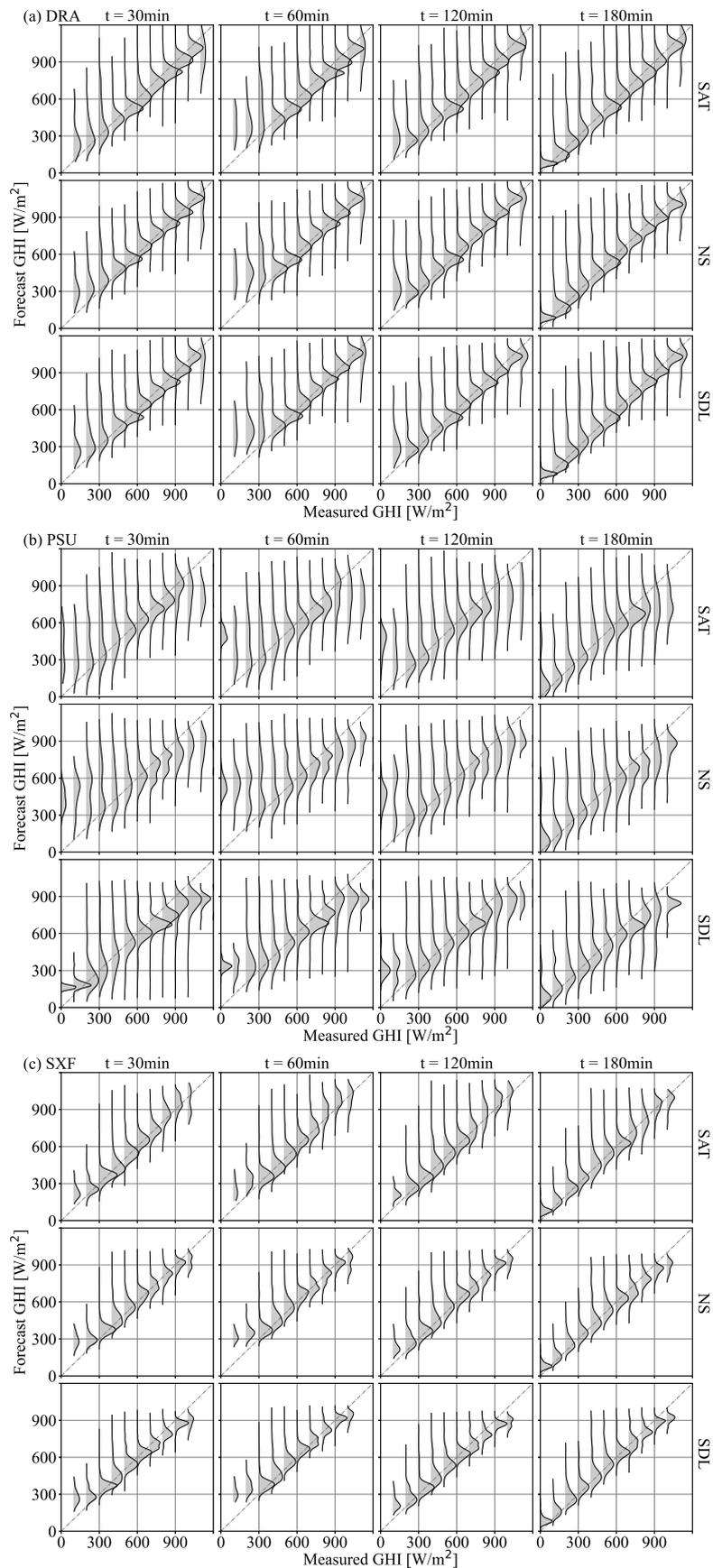


Fig. 13. Same as Fig. 12, but for conditional distributions of GHI forecasts, given the measurements.

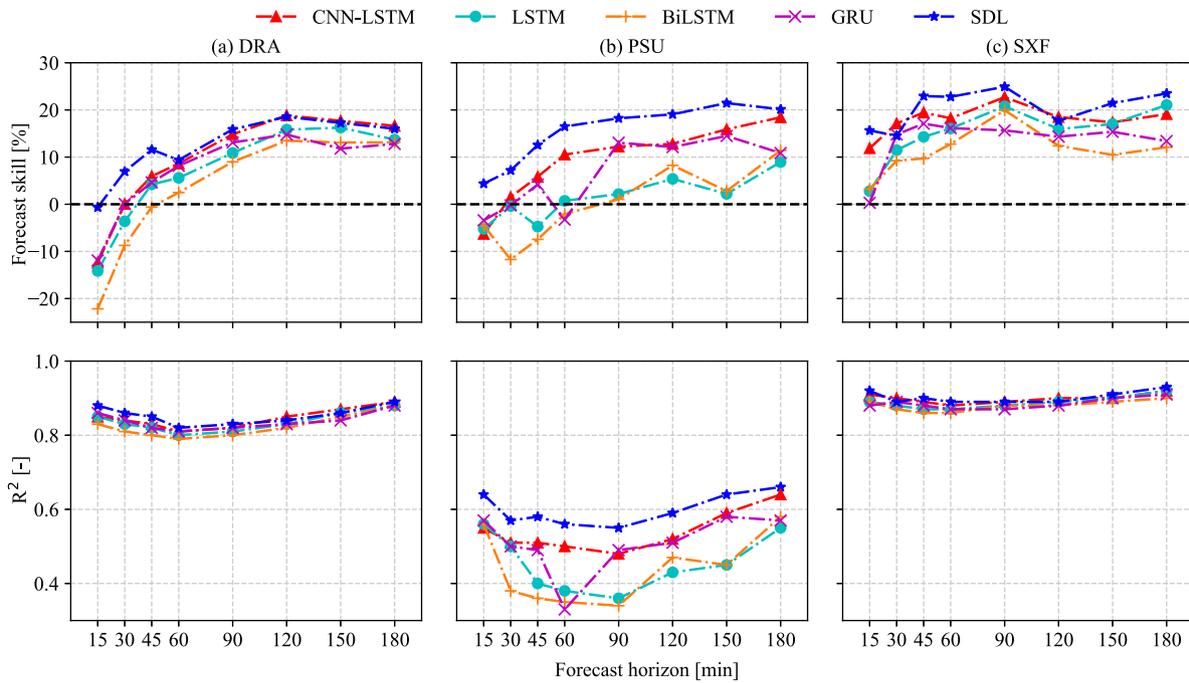


Fig. 14. The comparison of forecast skill [%] and R^2 for GHI forecasts with different deep learning methods at (a) DRA, (b) PSU, and (c) SXF. Higher forecast skill and higher R^2 indicate better results.

sparser joint and conditional distributions can be observed. This is related to the difficulties in cloud detection and prediction, as clouds are present in the three-dimensional atmosphere, while geostationary satellites can only provide two-dimensional information. It is suggested to integrate locally-sensed data, such as sky images [31], to provide additional information. Condition-based forecasting, which classifies clear, partially clear, and cloudy situations for model development, might be helpful to address the issues related to cloudy skies. Meanwhile, spatio-temporal satellite-derived irradiance products can also be used to track cloud movements [29]. Therefore, deriving more accurate irradiance estimates from satellite data also serves the objective, as evidenced by the improved performance of the proposed deep learning model chain for GHI estimates and forecasts. However, more research is still needed in modeling cloud movements in a three-dimensional perspective, since the ability to predict the amount, optical depth, movements, and locations of clouds is indispensable in improving solar forecasting [5].

5. Conclusions

Satellite data and satellite-derived irradiance products have been extensively used in solar forecasting, as the inclusion of spatio-temporal information is more likely to produce high-accuracy forecasts. However, the potential benefits of satellite-derived irradiance products and their improvements for intra-day solar forecasting have not been fully investigated in the literature. In this regard, this study compares the spatio-temporal inputs from spectral satellite measurements for intra-day solar irradiance forecasting with deep learning. Furthermore, a deep learning model chain for GHI forecasts is proposed to achieved improved forecasting performance. The model chain first converts satellite data to spatial GHI estimates via a deep learning model (SAT-DL), and then using the estimates as inputs to another deep learning model for intra-day solar irradiance forecasting. The effectiveness of the model chain in improving intra-day GHI forecasting is evaluated by comparing it with other methods, namely, the end-to-end deep learning model based on raw satellite images, the hybrid physical-deep learning model based on NSRDB, and other deep learning models of CNN-LSTM, LSTM, BiLSTM, and GRU.

Compared with using spectral satellite images, forecasting models based on spatial satellite-derived GHI estimates generally produce better forecasts (with lower nRMSE and higher R^2) in most cases. In addition, using SAT-DL GHI estimates show potentials for yielding better forecasts compared to using NSRDB GHI estimates. However, the forecasts based on satellite-derived GHI estimates tend to have larger biases. Despite the presence of larger bias, spatial GHI estimation products are still promising inputs for intra-day solar forecasting, as post-processing methods can be used for biases correction and refinement. When comparing the forecast skill, the proposed deep learning model chain generally have better performance and higher robustness. Given that spatial GHI estimates of SAT-DL have lower uncertainties than those of NSRDB in the nearby surrounding area, the improvement in GHI estimates could potentially lead to better GHI forecasts.

The results of ramp analysis and distribution-oriented verification show that there are larger forecasting uncertainties and biases for locations with cloudier skies. This is due to the higher unpredictability of the irradiance in such cloudy conditions. When there are more clouds, the forecasts are found to have compromised performance in the ramp detection. Meanwhile, GHI forecasts are also associated with larger biases for low- and high-irradiance conditions. This indicates that further efforts are required in detecting and predicting cloud movements. Nevertheless, the deep learning model chain shows better forecasting performance compared to the benchmark deep learning models.

In summary, the deep learning model chain shows great potential for improving GHI forecasts through the derivation of more accurate spatial GHI estimates from spectral satellite measurements. Compared with the benchmark deep learning models, the proposed deep learning model chain shows better performance, particularly under frequent cloudy conditions. However, there are still some limitations, such as the need of ground measurements and additional computational time to obtain GHI estimates via the pre-trained deep learning model. Considering that the physical radiative transfer model in NSRDB also requires extra time to produce GHI estimates for operational forecasting, the deep learning model chain provides a novel framework for satellite-based solar forecasting, where more advanced deep learning algorithms could be used as an alternative to obtain spatial GHI estimates to further improve intra-day solar forecasts.

CRedit authorship contribution statement

Shanlin Chen: Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Chengxi Li:** Writing – original draft, Methodology, Investigation. **Roland Stull:** Writing – review & editing, Supervision, Resources. **Mengying Li:** Writing – review & editing, Supervision, Conceptualization.

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.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT and GPT-4 in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgment

The authors gratefully acknowledge the substantial support from Hong Kong University Grants Committee (UGC) under project number 25213022.

References

- [1] IEA. Renewables 2022 [online]. 2022, Available at: <https://www.iea.org/reports/renewables-2022>. [Accessed 14 June 2023].
- [2] Vosoogh M, Kamyar M, Akbari A, et al. A novel modification approach based on MTLBO algorithm for optimal management of renewable micro-grids in power systems. *J Intell Fuzzy Systems* 2014;27(1):465–73.
- [3] Abbasi A, Abbasi S, Ansari J, Rahmani E. Effect of plug-in electric vehicles demand on the renewable micro-grids. *J Intell Fuzzy Systems* 2015;29(5):1957–66.
- [4] Chu Y, Li M, Coimbra CFM, Feng D, Wang H. Intra-hour irradiance forecasting techniques for solar power integration: A review. *iScience* 2021;103136.
- [5] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, Perez MJ, Perez R, Bright JM, Xia X, 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] Yang D, Wang W, Xia X. A concise overview on solar resource assessment and forecasting. *Adv Atmos Sci* 2022;1–13.
- [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] Wang W, Yang D, Huang N, Lyu C, Zhang G, Han X. Irradiance-to-power conversion based on physical model chain: An application on the optimal configuration of multi-energy microgrid in cold climate. *Renew Sustain Energy Rev* 2022;161:112356.
- [9] Yang D. Choice of clear-sky model in solar forecasting. *J Renew Sustain Energy* 2020;12(2):026101.
- [10] 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.
- [11] Van der Meer DW, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew Sustain Energy Rev* 2018;81:1484–512.
- [12] 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. *Sol Energy* 2021;218:652–60.
- [13] Li M, Chu Y, Pedro HTC, Coimbra CFM. Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term DNI forecasts. *Renew Energy* 2016;86:1362–71.
- [14] Kallio-Myers V, Riihelä A, Lahtinen P, Lindfors A. Global horizontal irradiance forecast for Finland based on geostationary weather satellite data. *Sol Energy* 2020;198:68–80.
- [15] Chen S, Liang Z, Guo S, Li M. Estimation of high-resolution solar irradiance data using optimized semi-empirical satellite method and GOES-16 imagery. *Sol Energy* 2022;241:404–15.
- [16] 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.
- [17] Yagli GM, Yang D, Gandhi O, Srinivasan D. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? *Appl Energy* 2020;259:114122.
- [18] Si Z, Yang M, Yu Y, Ding T. Photovoltaic power forecast based on satellite images considering effects of solar position. *Appl Energy* 2021;302:117514.
- [19] Harty TM, Holmgren WF, Lorenzo AT, Morzfeld M. Intra-hour cloud index forecasting with data assimilation. *Sol Energy* 2019;185:270–82.
- [20] Marchesoni-Acland F, Herrera A, Mozo F, Camiruaga I, Castro A, Alonso-Suárez R. Deep learning methods for intra-day cloudiness prediction using geostationary satellite images in a solar forecasting framework. *Sol Energy* 2023;262:111820.
- [21] 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 Intell* 2022;116:105493.
- [22] Nespoli A, Niccolai A, Ogliari E, Perego G, Collino E, Ronzio D. Machine Learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery. *Appl Energy* 2022;305:117834.
- [23] Nielsen AH, Iosifidis A, Karstoft H. IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. *Sol Energy* 2021;228:659–69.
- [24] Rodríguez-Benítez FJ, López-Cuesta M, Arbizu-Barrera C, Fernández-León MM, Pamos-Ureña MÁ, Tovar-Pescador J, Santos-Alamillos FJ, Pozo-Vázquez D. Assessment of new solar radiation nowcasting methods based on sky-camera and satellite imagery. *Appl Energy* 2021;292:116838.
- [25] Yang L, Gao X, Li Z, Jia D. Intra-day solar irradiation forecast using machine learning with satellite data. *Sustain Energy Grids Netw* 2023;36:101212.
- [26] Cheng L, Zang H, Wei Z, Ding T, Xu R, Sun G. Short-term solar power prediction learning directly from satellite images with regions of interest. *IEEE Trans Sustain Energy* 2021;13(1):629–39.
- [27] Aguiar LM, Pereira B, Laurent P, Díaz F, David M. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renew Energy* 2016;97:599–610.
- [28] Liu Y, Qin H, Zhang Z, Pei S, Wang C, Yu X, Jiang Z, Zhou J. Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. *Appl Energy* 2019;253:113596.
- [29] Yagli GM, Yang D, Srinivasan D. Ensemble solar forecasting and post-processing using dropout neural network and information from neighboring satellite pixels. *Renew Sustain Energy Rev* 2022;155:111909.
- [30] Wen H, Du Y, Chen X, Lim EG, Wen H, Yan K. A regional solar forecasting approach using generative adversarial networks with solar irradiance maps. *Renew Energy* 2023;216:119043.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] Yang D. SolarData: An R package for easy access of publicly available solar datasets. *Sol Energy* 2018;171:A3–12.
- [36] Blaylock BK. GOES-2-go: Download and display GOES-East and GOES-West data (Version 2022.07.15) [Computer software]. 2023, <https://github.com/blaylockbk/goes2go>.
- [37] Chen S, Li M. Improved turbidity estimation from local meteorological data for solar resourcing and forecasting applications. *Renew Energy* 2022;189:259–72.
- [38] Long CN, Shi Y. An automated quality assessment and control algorithm for surface radiation measurements. *Open Atmos Sci J* 2008;2(1).
- [39] Matsunobu LM, Pedro HTC, Coimbra CFM. Cloud detection using convolutional neural networks on remote sensing images. *Sol Energy* 2021;230:1020–32.
- [40] GOES R series product definition and users' guide (PUG) volume 3: Level 1B products. 2019, Available at: <https://www.goes-r.gov/users/docs/PUG-L1b-vol3.pdf>.
- [41] Buster G, Bannister M, Habte A, Hettinger D, Maclaurin G, Rossol M, Sengupta M, Xie Y. Physics-guided machine learning for improved accuracy of the National Solar Radiation Database. *Sol Energy* 2022;232:483–92.
- [42] Yang D. Validation of the 5-min irradiance from the National Solar Radiation Database (NSRDB). *J Renew Sustain Energy* 2021;13(1):016101.

- [43] 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.
- [44] Sun X, Bright JM, Gueymard CA, Acord B, Wang P, Engerer NA. Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis. *Renew Sustain Energy Rev* 2019;111:550–70.
- [45] Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT Press; 2016.
- [46] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30.
- [47] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, et al. *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*. 2016, arXiv preprint [arXiv:1603.04467](https://arxiv.org/abs/1603.04467).
- [48] O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al. *KerasTuner*. 2019, <https://github.com/keras-team/keras-tuner>.
- [49] Paletta Q, Terrén-Serrano G, Nie Y, Li B, Bieker J, Zhang W, Dubus L, Dev S, Feng C. Advances in solar forecasting: Computer vision with deep learning. *Adv Appl Energy* 2023;100150.
- [50] Paletta Q, Arbod G, Lasenby J. Benchmarking of deep learning irradiance forecasting models from sky images—An in-depth analysis. *Sol Energy* 2021;224:855–67.
- [51] Chu Y, Pedro HTC, Li M, Coimbra CFM. Real-time forecasting of solar irradiance ramps with smart image processing. *Sol Energy* 2015;114:91–104.
- [52] 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.
- [53] Yang D, Wang W, Bright JM, Voyant C, Notton G, Zhang G, Lyu C. Verifying operational intra-day solar forecasts from ECMWF and NOAA. *Sol Energy* 2022;236:743–55.
- [54] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, Blaga R, Boland J, Bright JM, Coimbra CFM, et al. Verification of deterministic solar forecasts. *Sol Energy* 2020;210:20–37.
- [55] Yang D, van der Meer D. Post-processing in solar forecasting: Ten overarching thinking tools. *Renew Sustain Energy Rev* 2021;140:110735.