Energy and AI 24 (2026) 100718

Contents lists available at ScienceDirect

Energy and AI

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

The role of spatio-temporal satellite-derived products and numerical weather forecasts in data-driven intra-day solar forecasting

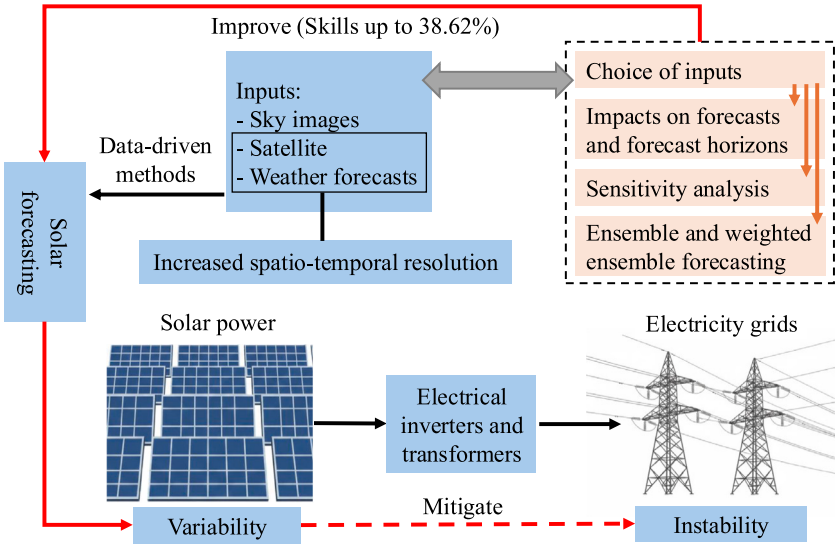
Shanlin Chen [a], Tao Jing [b], Dazhi Yang [c], Yuntian Chen [d], Mengying Li [b,1], Yinghao Chu [a,1,*]

[a] Department of Systems Engineering, City University of Hong Kong, Kowloon, Hong Kong Special Administrative Region of China
[b] Department of Mechanical Engineering & Research Institute for Smart Energy, The Hong Kong Polytechnic University, Kowloon, Hong Kong Special Administrative Region of China
[c] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China
[d] Eastern Institute for Advanced Study, Eastern Institute of Technology, Ningbo, Zhejiang, China

HIGHLIGHTS

- High-resolution NWP forecasts are introduced for intra-day GHI and DNI forecasting.
- Satellite- and NWP-based data are useful for solar forecasting at different horizons.
- Significance of inputs selection in data-driven solar forecasting is investigated.
- Ensemble methods considering the importance of used inputs are more effective.

GRAPHICAL ABSTRACT



ARTICLE INFO

Dataset link: https://gml.noaa.gov/aftp/data/radiation/surfrad/, https://noaa-goes16.s3.amazonaws.com/index.html, https://nsrdb.nrel.gov/data-viewer, https://noaa-hrrr-bdp-pds.s3.amazonaws.com/index.html

Keywords:
Solar forecasting
Satellite-derived irradiance
Numerical weather prediction
Deep learning

ABSTRACT

Accurate intra-day forecasting of global horizontal irradiance (GHI) and direct normal irradiance (DNI) is critical for the efficient operation of solar energy systems. While satellite-based data have traditionally been used as the most relevant exogenous spatio-temporal input for such forecasts, numerical weather prediction (NWP) models with increasing resolution now offer promising benefits. This study evaluates the effectiveness of spatio-temporal satellite- and NWP-based inputs (i.e., irradiance and cloud cover forecasts) and their impacts on intra-day GHI and DNI forecasts using deep learning methods from three perspectives: single type of input, input combinations, and ensemble forecasting. Results demonstrate that the combined use of satellite- and NWP-based inputs generally shows better performance than their individual use, achieving skill scores up to 36.35% and 30.87% for 4-hour-ahead GHI and DNI forecasts, respectively. Among the three spatio-temporal

* Corresponding author.
E-mail addresses: mengying.li@polyu.edu.hk (M. Li), yinghchu@cityu.edu.hk (Y. Chu).
[1] The two authors have the same contribution to this study.

https://doi.org/10.1016/j.egyai.2026.100718
Received 15 December 2025; Received in revised form 12 February 2026; Accepted 12 March 2026
Available online 14 March 2026

Ensemble forecasting

inputs, NWP irradiance exerts a greater overall influence, while satellite-derived products contribute more to short-term predictions (less than 2 h ahead), NWP cloud cover is more influential for longer forecast horizons. Based on impacts of satellite- and NWP-based inputs on the forecast, two weighted ensemble strategies are proposed to further enhance the forecasting performance, where skill scores can reach 38.62% for GHI and 34.61% for DNI in 4-hour-ahead forecasts. With the growing capacity addition of solar energy systems, this work highlights the advantages of integrating spatio-temporal satellite-derived products and NWP forecasts in intra-day solar forecasting with deep learning. This offers practical benefits to a wide range of stakeholders in the solar energy community for real-world applications.

Nomenclature

Abbreviations

3D-CNN	3-dimensional convolutional neural network
BON	Bondville
C	HRRR cloud forecasts
CNN	Convolutional neural network
ConvLSTM	Convolutional long-short-term memory
CSP	Concentrated solar power
DHI	Diffuse horizontal irradiance
DNI	Direct normal irradiance
DRA	Desert Rock
FPK	Fort Peck
GHI	Global horizontal irradiance
GOES	Geostationary Operational Environment Satellite
GWN	Goodwin Creek
HRRR	High-Resolution Rapid Refresh
I	HRRR irradiance forecasts
LSTM	Long-short-term memory
MBE	Mean bias error
MERRA-2	Modern Era Retrospective analysis for Research and Applications, version 2
nMBE	Normalized mean bias 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
SDI	Satellite-derived irradiance
SP	Smart persistence
SURFRAD	Surface Radiation Budget Network
SXF	Sioux Falls
TBL	Table Mountain
UTC	Coordinated Universal Time

Notations

∘	Degree
C	HRRR cloud forecasts
f	Forecast
I	HRRR irradiance forecasts
N	Number of data points
o	Observation
t	Time
DSWRF	Incoming shortwave radiation in HRRR
SDI	Satellite-derived irradiance
TCDC	Total cloud cover in HRRR
VBDSF	Incoming direct radiation in HRRR

Subscripts

f	Forecast of interest
r	Reference forecast
t	Time

1. Introduction

Solar energy, with worldwide potential, is one of the main renewable energy sources to meet the growing need for clean power and to mitigate climate change [1]. Solar radiation, including global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI), is known as the "fuel" for all solar energy technologies, such as photovoltaic (PV) and concentrating solar power (CSP), both of which are essential technologies on the path toward carbon neutrality [2]. Therefore, knowledge of the future reliability of solar radiation is crucial for the operation of solar energy systems, including scheduling, maintenance, and grid integration [3]. The penetration of solar power into the power grid will continue to increase, driven by reduced costs and supportive policies [1]. However, the intermittent nature of solar radiation greatly hinders the application and expansion of solar energy systems [4]. To facilitate the efficient operation of solar energy projects and support their integration, solar forecasting is a cost-effective method to alleviate the variability of solar power [5,6].

Solar forecasting refers to both solar irradiance forecasting and solar power forecasting [7,8]. These two topics are inseparable, as solar irradiance is the primary factor that affects the power output of any solar energy system. In fact, it is necessary to have accurate irradiance forecasts to produce high-quality PV and CSP power forecasts [2,7], as end-to-end data-driven solar power forecasting models are usually constrained by the limited availability of high-quality power data [9]. Solar forecasts with various forecast horizons can be broadly classified as intra-hour, intra-day, and day-ahead forecasting [1]. These forecasts are essential for accommodating various regulations and requirements for the control and operation of power systems, such as economic dispatch, real-time regulation, and intra-day and day-ahead unit commitment [1,10].

Solar forecasts produced using spatio-temporal inputs are generally more advantageous than those based solely on single-location measurements [11], due to the spatio-temporal nature of ground-level solar irradiance [1]. The spatio-temporal inputs for solar forecasting at different time horizons can generally be obtained from locally sensed data, satellite measurements, and numerical weather prediction (NWP) forecasts [1]. Locally sensed data, such as sky images, are particularly informative for intra-hour forecasting [5], where image processing is applied to extract cloud motions for the near future [12]. Satellite images and satellite-derived products that provide atmospheric information over a larger scale are suitable for intra-day forecasting [3,13]. Since NWP models can simulate long-term atmospheric dynamics, irradiance and cloud predictions from NWP are typically applied

for day-ahead forecasting [1]. In all of these scenarios, data-driven methods, such as deep learning, have been extensively applied for both deterministic and probabilistic solar forecasting applications [6,14], especially when physical models struggle to capture and analyze spatio-temporal cloud information [15]. For instance, in satellite-based intra-day GHI forecasting using cloud motion vectors, the cloud field is only subject to advection without considering the formation and dissipation of clouds [16], whereas data-driven methods (e.g., deep learning) that can learn complex atmospheric patterns are not constrained by this limitation and consequently demonstrate superior performance [17, 18].

NWP models can simulate long-term atmospheric dynamics using initial state information and adequate computational resources [4]. However, insufficient information in the initial conditions and the complexity of physical atmospheric processes diminish the accuracy of NWP forecasts [19], especially for solar forecasting at shorter horizons [20]. Therefore, NWP outputs are generally considered more suitable for long-term solar forecasting [1], when data-driven methods cannot capture the dynamics and drastic changes of clouds over a few hours [21]. It is widely accepted that NWP irradiance forecasts should be post-processed for bias refinement due to the model-led systematic errors of NWP models [22]. This step has been deemed essential in the irradiance-to-power conversion process [23]. In addition to post-processing NWP irradiance forecasts, many data-driven methods have been proposed for intra-day solar forecasting using a variety of other NWP outputs, such as temperature, cloud cover, and wind, to further improve forecasting accuracy [24,25]. However, the benefit of spatio-temporal information for solar forecasting is also applicable to NWP outputs [26]. These single-grid NWP data (near the location of interest) might not account for the atmospheric dynamics, especially the clouds, in the surrounding area; thus, they could lead to inferior performance [26,27].

With increased resolutions of satellite-derived products and NWP forecasts [1], there are more possibilities to combine different types of spatio-temporal inputs for improved solar forecasts using deep learning. For instance, sky images and high-resolution (i.e., 5-min) satellite products can be jointly used for improved intra-hour solar forecasting [28]; similarly, it could be beneficial to incorporate satellite-based data and hourly updated NWP models for more skillful intra-day solar forecasts [1,3]. Studies have shown that the combined use of satellite- and NWP-based data can lead to improved intra-day solar forecasts. For instance, although satellite- and NWP-based inputs show different impacts on the forecast, the fusion of satellite- and NWP-based data with the transformer approach [11] generally leads to better results; similarly, the blending of their individual forecasts can also result in improved performance [29]. Nonetheless, the inclusive use of satellite- and NWP-based data is mainly applied for intra-day GHI forecasting, and the spatio-temporal resolution of NWP forecasts used is relatively coarser [30,31]. On the other hand, there are only a few studies on the application of NWP for forecasting DNI [32,33]. These DNI forecasts are improved based on NWP data, whereas satellite-based data are not used.

As an essential component of solar irradiance, DNI is directly related to CSP and is also required in modeling PV power output (though there is some uncertainty, DNI can be obtained from GHI separation [34]). This indicates the need for direct and accurate DNI forecasts to improve the irradiance-to-power conversion process. However, the integration of both satellite- and NWP-based data for DNI forecasting remains limited, especially in the context of intra-day forecasting. Theoretically, all spatio-temporal information that can reflect the future atmospheric status should be used collectively in the solar forecasting model [1]. This is particularly feasible for satellite-derived and NWP-based data due to their wide spatial coverage and improved temporal resolution [11]. Moreover, the use of multiple spatio-temporal data sources enables forecast combination [7], or ensemble forecasting, which can further improve performance [35]. However, ensemble forecasting is usually an arithmetic mean without the selection of members [7]. Although there are some benefits in forecast combination, the impacts of various inputs on the forecast that could support the design of ensemble forecasting methods are often less studied. In this regard, this work aims to investigate the effectiveness of using integrated satellite- and NWP-based spatio-temporal inputs for both intra-day GHI and DNI forecasting with data-driven methods, studying the impacts of satellite and NWP inputs on the forecasts (and forecast horizons) and their influences on forecast combinations. The major contributions are summarized as follows:

- Given that an hourly updated NWP model with finer spatio-temporal resolution (i.e., 1-h–3-km), such as the High-Resolution Rapid Refresh (HRRR) model [36], is operationally available, this work aims to investigate the effectiveness of such NWP forecasts in improving the accuracy of both intra-day GHI and DNI forecasting. This study on GHI and DNI forecasting using satellite- and NWP-based inputs provides substantial benefits for solar power systems, particularly for CSP applications.
- Considering their distinct characteristics, the impacts of spatio-temporal satellite- and NWP-based inputs on forecasting performance are evaluated separately and in combination using data-driven methods. In contrast to conventional approaches that primarily focus on refining model architecture, this work demonstrates that forecasts can also be improved through the careful curation of inputs.
- The importance of each type of spatio-temporal input on intra-day GHI and DNI forecasting is evaluated using sensitivity analysis conducted within the framework of ensemble forecasting. Based on the sensitivity analysis, new strategies for weighted ensemble solar forecasting, considering the features of the inputs used, are proposed for more skillful intra-day GHI and DNI forecasts.

The rest of this work is structured as follows. Section 2 describes the data and methods used in this study, including data pre-processing, the forecasting method, and performance evaluation metrics. Forecasting results are evaluated and compared in Section 3, while the discussion is presented in Section 4. Finally, the key findings of this study are summarized in Section 5.

2. Data and methods

Data used in this study include on-site measurements, satellite-derived irradiance products, and NWP forecasts. The data-driven method employs various spatio-temporal inputs and their combinations to produce different deterministic irradiance predictions, which are used to generate ensemble forecasts for enhanced performance. The improved forecast is essential for mitigating solar variability and supporting grid integration.

2.1. Ground-level measurements

Radiometric measurements from the Surface Radiation Budget Network (SURFRAD) [37] are used as both the target for model training and the ground truth for performance evaluations. SURFRAD has been providing high-quality on-site measurements for a variety of applications and research related to climate, weather, and energy [38]. There is a network of seven stations in SURFRAD, including Bondville (BON), Desert Rock (DRA), Fort Peck (FPK), Goodwin Creek (GWN), Penn. State Uni. (PSU), Sioux Falls (SXF), and Table Mountain (TBL), located in five different climatological zones across the contiguous United States as shown in Fig. 1. More details on SURFRAD are available in [7,39].

In-situ measurements of GHI, DNI, DHI, and the calculated solar zenith angle in the year 2020 are retrieved for all the stations. Although GHI and DNI are the variables of interest, data of DHI and solar zenith angle are required in the quality control (QC) process. The following steps are performed in pre-processing SURFRAD data:

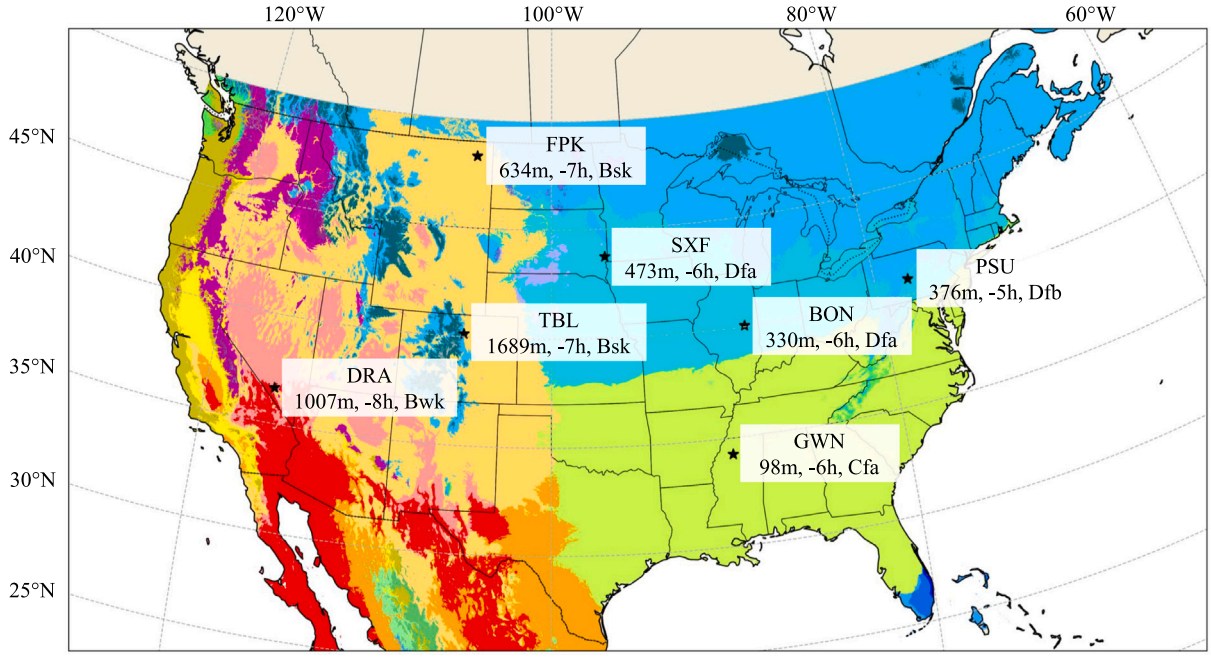


Fig. 1. Geographic distribution of the seven SURFRAD stations, with information in brackets being (altitude [m], time difference from UTC [hours], climate zones).

- QC: The extremely rare limit test and three-component closure test [40] are used to filter out data points that do not meet the QC tests from the raw 1-min SURFRAD measurements. Meanwhile, irradiance measurements for a solar zenith angle of 85° or greater are discarded due to the high airmass effect during solar mornings and evenings [41]. Therefore, nighttime data points are not used in model training and testing.
- Data aggregation: After QC, the 1-min GHI and DNI measurements are aggregated to a 1-h resolution and indexed in Coordinated Universal Time (UTC). This is done to ensure compatibility with the NWP data, which has a temporal resolution of 1-h. Note that the aggregation should use the "center" scheme as suggested in [7,39].
- Normalization: The normalization is performed to remove the double-seasonal pattern (i.e., yearly and daily) of solar irradiance. This is a common practice in solar forecasting [1,42]. The 1-h GHI and DNI measurements are normalized using clear-sky GHI and clear-sky DNI estimations from REST2 [43] to generate clear-sky indexes for GHI and DNI, respectively. These clear-sky indexes are used as labels in the forecasting model. Guidelines for obtaining the REST2 clear-sky irradiance are provided in Section 2.2.

2.2. Satellite-derived irradiance products

Satellite-derived irradiance products have been widely used as inputs, either exclusively or in conjunction with other exogenous information, in data-driven solar forecasting. As an extensively accessed and publicly available database, the National Solar Radiation Database (NSRDB) provides satellite-derived irradiance for the United States, and the number of international locations is still growing [44,45]. The NSRDB employs a physical solar model to produce solar irradiance estimations with inputs from a number of sources, such as data from the Geostationary Operational Environmental Satellite (GOES) and the Modern Era Retrospective analysis for Research and Applications, version 2 (MERRA-2). The data in the NSRDB include GHI, DNI, and other auxiliary variables, with the spatio-temporal resolution improved to 2 km and 5 min. However, the 5-min NSRDB data show a high discrepancy when validated against the SURFRAD measurements [46,47]. Therefore, an improved satellite-to-irradiance method has been developed for both GHI and DNI estimations using multi-spectral satellite images from GOES-16 and deep learning [48]. The combination of remote sensing and deep learning demonstrates great potential in extracting cloud information for solar irradiance estimation. Compared to the NSRDB, the 5-min GHI and DNI estimations exhibit comparatively lower uncertainties [48].

The satellite-to-irradiance method, based on multi-spectral satellite images from GOES-16 and deep learning, is further expanded to produce regional GHI estimates, which can subsequently be used as inputs for solar forecasting with improved performance [13]. Since the reduced uncertainties in spatial satellite-to-irradiance estimations can further enhance forecasting accuracy [13], the satellite-derived GHI and DNI irradiance for 2020 used in this study are obtained following the same methodology outlined in [13] (details are presented in Appendix A for easier reference). Note that the spatial irradiance estimates obtained by the pre-trained deep learning model are in the form of clear-sky indexes, which are used as spatio-temporal inputs for data-driven intra-day solar forecasting at the target location (beneath the center pixel). Therefore, these clear-sky indexes are not converted back to irradiance values at this stage. Due to some technical issues and communication failures during the transfer of satellite measurements, there are some missing points in the spatial irradiance estimates. These gaps are filled using NSRDB irradiance values normalized by the clear-sky irradiance computed with the REST2 model [44]. It should be noted that REST2 clear-sky irradiance data for both GHI and DNI are already included in the NSRDB. For more details on the REST2 model and its required atmospheric inputs, such as MERRA-2, readers are referred to [43,49]. The 5-min clear-sky indexes are then aggregated to a 1-h resolution using the "center" scheme, as mentioned in Section 2.1.

2.3. NWP forecasts

With hourly updated NWP models available, such as the HRRR model [36], it is both possible and beneficial to use high-resolution spatio-temporal NWP forecasts for intra-day solar forecasting [1,7]. As an hourly updated NWP model, the HRRR produces 0–18-h-ahead forecasts, where horizon 0 indicates reanalysis data for the initial

conditions. Furthermore, there are four runs (i.e., UTC 00, 06, 12, 18) to extend HRRR forecasts to 48-h ahead [39]. The spatial resolution of HRRR forecasts is 3 km. The forecast variables related to solar irradiance in HRRR include GHI, DNI, and cloud amount, as clouds are the primary factor attenuating ground-level solar irradiance. Other HRRR variables, such as temperature, wind, and precipitation, are not applied, as they typically have larger effects on the power output of PV systems than solar irradiance [50]. In HRRR, GHI is represented by incoming shortwave radiation (the variable name is *DSWRF*), DNI is known as incoming direct radiation (the variable name is *VBDSF*), and the cloud amount is represented by total cloud cover (the variable name is *TCDC*). In the context of intra-day solar forecasting, the 1–12-h HRRR forecasts in 2020 of GHI [W/m^2], DNI [W/m^2], and cloud amount [%] issued at UTC 00 and UTC 12 are downloaded for the 11 × 11 locations surrounding the target station. Note that there are gaps in the HRRR data, which are filled with linear interpolations.

2.4. Forecasting method

In this study, 1–4-h-ahead clear-sky indexes for GHI and DNI, with a temporal resolution of 1 h, are generated using deep learning models that leverage various types of hourly spatio-temporal inputs and their combinations, as illustrated in Fig. 2(a). The employed spatio-temporal inputs are:

- Satellite-derived irradiance (SDI) products: 11 × 11 clear-sky indexes with the target location in the center. The hourly data frames are SDI_{t_0-4}, SDI_{t_0-3}, SDI_{t_0-2}, SDI_{t_0-1}, where t_0 indicates the current time. It should be noted that the irradiance values are aggregated in a 1-h resolution using the "center" scheme, as described in Section 2.1 (e.g., data points from 11:31 to 12:30 are aggregated and indexed at 12:00). Therefore, SDI_{t_0} is not operationally available for use as input.
- Raw NWP irradiance forecasts (I): 11 × 11 data points with the target location in the center. These NWP irradiance forecasts are normalized as NWP clear-sky indexes using the clear-sky irradiance of REST2. The hourly inputs used include I_{t_0-4}, I_{t_0-3}, ⋯, I_{t_0+6} (11 time stamps in total). Note that NWP irradiance forecasts used as inputs for data-driven solar forecasting are not post-processed.
- NWP cloud forecasts (C): 11 × 11 data points with the target location in the center and at 11 hourly time stamps C_{t_0-4}, C_{t_0-3}, ⋯, C_{t_0+6}.

The use of surrounding spatio-temporal satellite-derived products and NWP forecasts, rather than single-location information, can improve solar forecasting accuracy [1]. It should be noted that a tradeoff exists between the spatial size of inputs (complexity) and modeling accuracy: a smaller surrounding area may limit model performance, while a larger one increases computational costs [51]. Therefore, considering that the improvement resulting from an increased spatial region could be marginal [11], the size of 11 × 11 surrounding pixels is empirically adopted in this work, following the practices of several studies on cloud detection [52], satellite-to-irradiance conversion [48], and satellite-based intra-day solar forecasting [13,53]. A total of seven groups of inputs (three single types and four combined types) are used to generate intra-day forecasts, as illustrated in Fig. 2(b) and (c). A deep learning framework comprising convolutional neural networks (CNNs), an attention mechanism, and fully connected dense layers [13] is employed to perform intra-day GHI and DNI forecasting (1–4-h ahead). Details on hyperparameter tuning and defined ranges are described in Appendix B. To account for seasonal variations, data from March, June, September, and December are used for testing, while the remaining data from 2020 are split into training (80%) and validation (20%) subsets [48].

Given the variety of deep learning architectures available for solar forecasting with spatio-temporal inputs, such as CNN-long short-term memory (LSTM), convolutional LSTM (ConvLSTM), and 3D-CNN, one might argue that more advanced deep learning models could better capture spatio-temporal features, potentially leading to improved forecasts. Indeed, ConvLSTM has been shown to slightly outperform CNNs [54,55]. However, more sophisticated models, such as ConvLSTM, typically demand greater computational resources and time. Furthermore, the primary focus of this study is to investigate the effectiveness of combining satellite- and NWP-based data for intra-day solar forecasting rather than developing a more advanced deep learning model. Considering the comparable performance of CNNs and their lower computational burden, particularly for ensemble forecasting, CNNs are adopted in this work. Nevertheless, more advanced deep learning algorithms, such as transformers and graph neural networks, could be explored in future research to further enhance forecasting performance.

Subsequently, a sensitivity analysis of the inputs is performed, as shown in Fig. 2(d), and ensemble forecasts are developed. The ensemble method is widely used in weather forecasting, where averaging forecasts produced by different methods has been shown to outperform individual ensemble members [7]. The sensitivity analysis is performed using averaged ensemble forecasts by excluding members based on the specific type of spatio-temporal input. For example, there are seven ensemble members with different inputs (i.e., SDI, I, C, SDI+I, SDI+C, I+C and SDI+I+C), when evaluating the importance of I (NWP irradiance) in ensemble forecasts, all members based entirely or partly on I (i.e., I, SDI+I, I+C and SDI+I+C) are excluded from the ensemble calculation. Based on the sensitivity analysis, two alternative strategies for producing weighted ensemble forecasts are proposed to further enhance forecasting performance. The underlying idea is to combine forecasts from different ensemble members, such as by averaging predictions generated by various models or inputs. More details are provided in Section 3.3.

2.5. Performance evaluation metrics

The evaluations and visualizations are conducted based on irradiance values [W/m^2], with nighttime data excluded. Forecasts of clear-sky indexes are converted back to GHI and DNI by multiplying them with the REST2 clear-sky irradiance at the corresponding predicted timestamps. Statistical metrics of root mean squared error (RMSE), mean bias error (MBE), their normalized counterparts (nRMSE and nMBE), and the coefficient of determination (R^2) are used to assess the overall forecasting performance [3,5], as defined by the following equations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum(f_i - o_i)^2}, \tag{1}$$

$$\mathrm{nRMSE} = \sqrt{\frac{1}{N}\sum(f_i - o_i)^2}/\bar{o}_i, \tag{2}$$

$$\mathrm{MBE} = \frac{1}{N}\sum(f_i - o_i), \tag{3}$$

$$\mathrm{nMBE} = \frac{1}{N}\sum(f_i - o_i)/\bar{o}_i, \tag{4}$$

$$R^2 = 1 - \frac{\sum(f_i - o_i)^2}{\sum(o_i - \bar{o}_i)^2}, \tag{5}$$

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

Furthermore, the forecast skill score is also used to assess the relative improvement of a forecasting model over a reference model [3,5],

$$\mathrm{FS} = \left(1 - \frac{\mathrm{RMSE}_f}{\mathrm{RMSE}_r}\right) \times 100\%, \tag{6}$$

where FS represents the forecast skill, RMSE_f and RMSE_r are the errors of the evaluated forecasting model and the reference forecasting model,

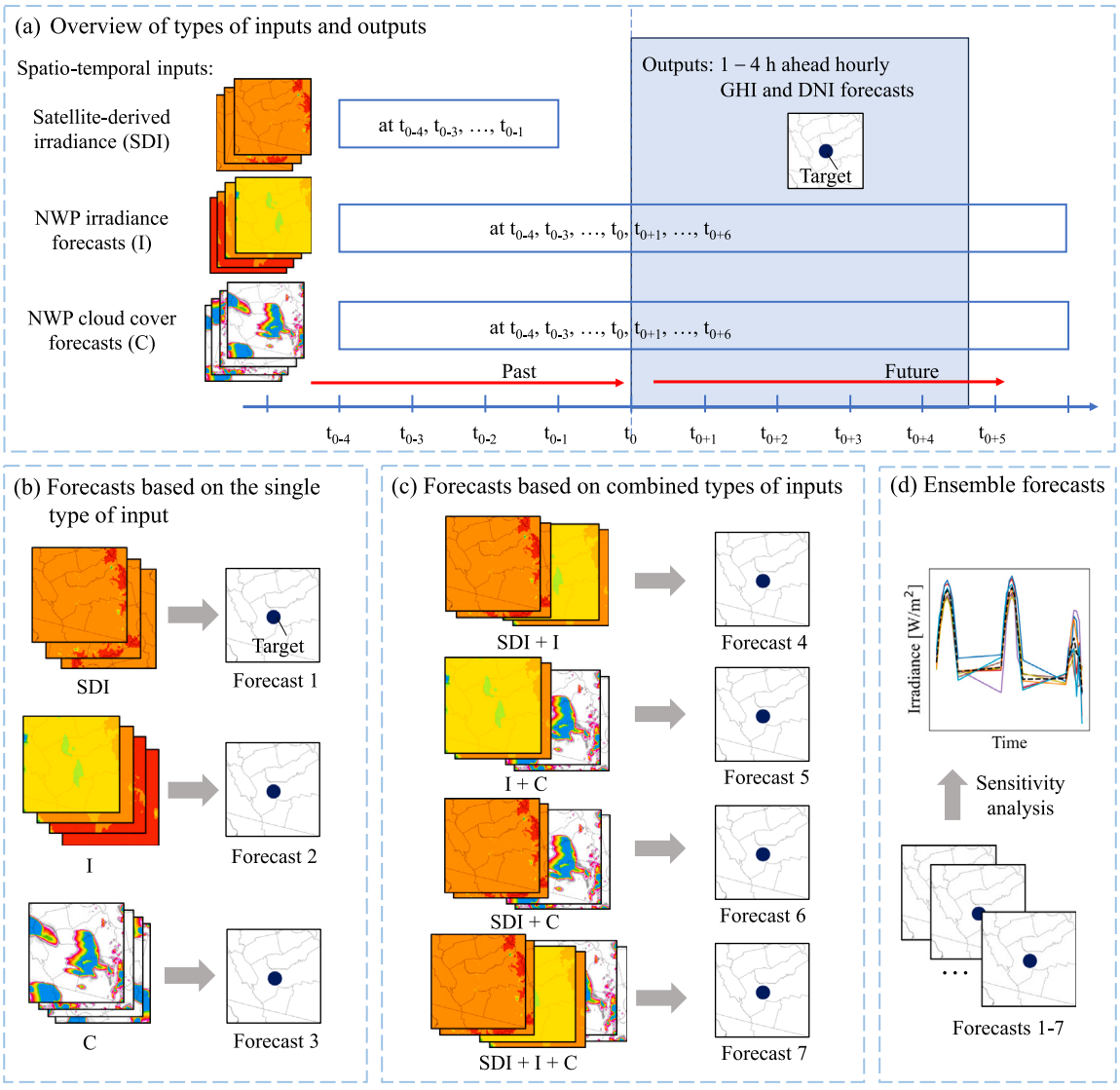


Fig. 2. The illustration of data-driven intra-day solar irradiance forecasting with different spatio-temporal inputs and their combinations. (a) Overview of the proposed types of spatio-temporal inputs and outputs, (b) forecasts based on the single type of input, (c) forecasts based on combined types of inputs, and (d) ensemble forecasts based on sensitivity analysis.

respectively. The referenced smart persistence model assumes that the clear-sky index remains constant over the forecast horizon. Note that since the hourly averaged irradiance measurements under the "center" scheme are not available at t_0, the clear-sky index at t_{0-1} is used in the smart persistence model.

3. Results

3.1. Forecasts based on the single type of spatio-temporal input

This section presents data-driven GHI and DNI forecasting results using a single type of spatio-temporal input: satellite-derived irradiance (denoted as SDI), NWP irradiance (denoted as I), and NWP total cloud cover (denoted as C). Additionally, GHI and DNI forecasting results from the NWP model and the smart persistence forecast model are included for comparison. As illustrated in Fig. 3, the nRMSE generally increases as the forecast horizon extends. Among the five scenarios, raw NWP irradiance predictions show the largest error for both GHI and DNI forecasting. When comparing the types of spatio-temporal inputs used, forecasts generated based on NWP irradiance typically show comparatively lower nRMSE for both 1–4-h-ahead GHI and DNI predictions. SDI is beneficial for forecasting at a shorter horizon (i.e., for the 1-h-ahead forecast), while the discrepancy becomes larger as the forecast horizon extends. For longer forecast horizons, spatio-temporal NWP forecasts (both I and C) are generally more beneficial than SDI as inputs for intra-day GHI and DNI forecasting. Compared to GHI forecasts, DNI forecasts show similar trends but usually exhibit higher uncertainties, as shown in Fig. 3. It can be seen that DNI forecasts are also associated with greater biases, while raw NWP irradiance forecasts exhibit the highest bias, often resulting in over-predictions for both GHI and DNI. For data-driven forecasting based on different types of spatio-temporal input, the biases vary across stations. Both GHI and DNI tend to be under-predicted, with some site-specific divergences. In general, the use of NWP cloud cover leads to relatively larger biases compared to the use of NWP irradiance or satellite-derived irradiance products.

The forecast skill scores for both GHI and DNI forecasts at 1–4-h ahead with different types of spatio-temporal input are presented in Table 1. Generally, the forecast skill increases with the forecast horizon, regardless of the input used. Data-driven forecasts using NWP irradiance usually demonstrate comparatively higher forecast skills than those using satellite-derived irradiance and NWP cloud cover, with a few exceptions where forecasts based on satellite-derived irradiance

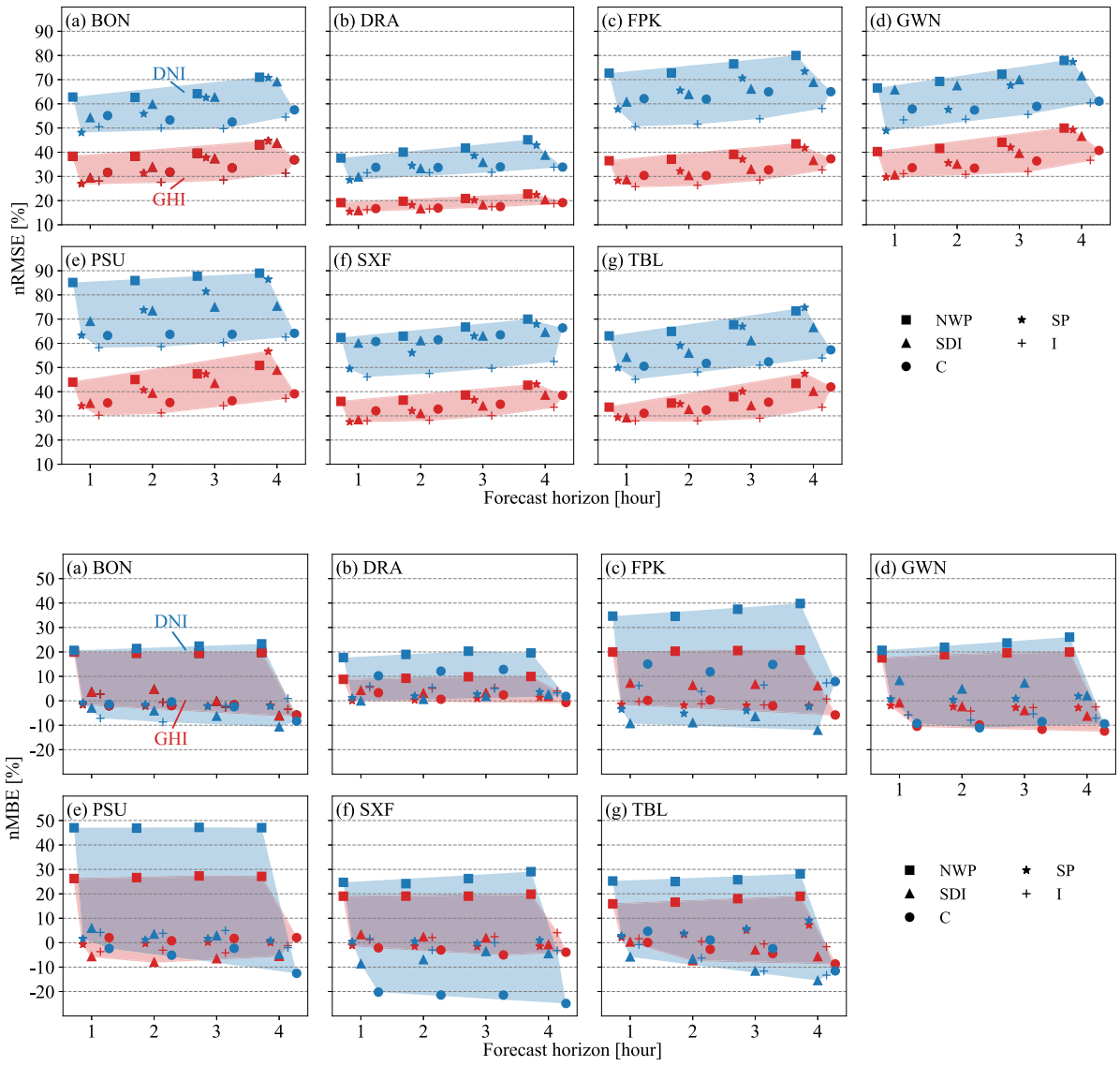


Fig. 3. The nRMSE and nMBE [%] for GHI (red) and DNI (blue) forecasts using single type of spatio-temporal input across the forecast horizon up to 4 h ahead at seven SURFRAD stations. "NWP" means the raw NWP irradiance forecast, "SP" denotes the smart persistence model, "SDI" refers data-driven forecasts produced using satellite-derived irradiance products, "I" is the data-driven forecasts generated using NWP irradiance, and, "C" represents the data-driven forecasts using NWP cloud cover. The shaded area covers all data points for better visualization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

exhibit better performance at the forecast horizon of 1 h ahead (i.e., at DRA and GWN). Negative skill scores can be observed for most 1-h-ahead forecasts, especially those generated using satellite-derived irradiance or NWP cloud cover. When comparing forecasts using satellite-derived irradiance and NWP cloud cover, satellite-derived irradiance tends to yield better results in the 1-h ahead forecast horizon, while NWP cloud cover is generally more advantageous for forecasts in 3–4-h ahead scenarios. Another interesting observation is that 1-h-ahead GHI and DNI forecasts using satellite-derived irradiance sometimes show even higher forecast skills than those using NWP irradiance. Nonetheless, NWP irradiance demonstrates better overall performance for both intra-day GHI and DNI forecasting when used as the sole type of spatio-temporal input in data-driven models.

3.2. Forecasts based on combined types of spatio-temporal inputs

Due to the increased resolution and update frequency of NWP models, it is feasible to integrate such high-resolution spatio-temporal NWP predictions with satellite-derived irradiance to generate intra-day solar forecasts. In this section, four combinations of three types of input data (i.e., SDI, I, and C) are evaluated, namely SDI+I, I+C, SDI+C, and SDI+I+C. The nRMSE and nMBE values for 1–4-h-ahead GHI and DNI forecasts using different input combinations are presented in Fig. 4. For comparison, forecasts from the smart persistence model and those using the best single type of input are also included. Similar to the results presented in Section 3.1, the nRMSEs of GHI and DNI forecasts increase with the extended forecast horizon, regardless of which input combination is used. The error of DNI forecasts is generally higher than that of GHI forecasts across different stations and forecast horizons. Although there are some site-specific variations, one general trend for both GHI and DNI forecasts is that the inclusive use of SDI, I, and C (SDI+I+C) could potentially generate better results. Among all other input combinations (excluding SDI+I+C), SDI+C usually generates forecasts with higher uncertainties, sometimes even exceeding those produced using the single type input (NWP irradiance, denoted as I in Fig. 4), while SDI+I could be used to produce forecasts with lower nRMSE than SDI+C. Forecasts based on the combination of I+C tend to

Table 1
The forecast skill score [%] over smart persistence for both GHI and DNI forecasts up to 4 h ahead using different types of inputs: satellite-derived irradiance (SDI), raw NWP irradiance (I), and NWP cloud cover (C), at all SURFRAD stations.

Station	Method	GHI				DNI			
		1 h	2 h	3 h	4 h	1 h	2 h	3 h	4 h
BON	SDI	−9.2	−8.0	0.0	2.2	−12.8	−7.0	−0.3	2.4
	I	**−4.3**	**12.0**	**24.6**	**29.8**	**−5.1**	**10.5**	**20.5**	**23.0**
	C	−17.4	−1.4	11.3	17.5	−14.5	4.6	16.1	18.8
DRA	SDI	**−2.6**	8.5	9.6	9.3	**−4.5**	3.5	7.1	9.7
	I	−5.3	**8.9**	**13.6**	**16.2**	−10.6	**8.4**	**17.7**	**21.2**
	C	−8.0	7.0	13.2	14.6	−18.5	2.3	11.9	21.2
FPK	SDI	−1.3	5.6	11.1	12.5	−5.2	2.6	6.4	6.2
	I	**8.9**	**18.2**	**23.1**	**21.8**	**12.4**	**21.3**	**23.8**	**20.9**
	C	−7.5	5.8	11.8	10.8	−7.6	5.5	7.9	11.4
GWN	SDI	**−3.0**	1.4	5.8	5.5	−34.6	−17.2	−3.5	7.5
	I	−4.8	**13.4**	**23.7**	**25.6**	**−9.1**	**6.9**	**17.8**	**21.9**
	C	−13.1	6.0	13.3	17.4	−18.4	0.3	12.9	21.0
PSU	SDI	−3.3	3.2	8.1	13.7	−9.1	0.4	8.0	12.8
	I	**11.2**	**23.2**	**27.7**	**34.3**	**8.0**	**20.5**	**25.9**	**27.6**
	C	−3.9	12.7	23.3	31.0	0.1	13.6	21.8	25.8
SXF	SDI	−3.3	3.4	6.8	10.5	−21.4	−9.0	−0.1	5.0
	I	**−1.8**	**12.0**	**17.8**	**22.1**	**6.9**	**15.2**	**21.1**	**22.8**
	C	−16.7	−2.3	4.9	10.7	−22.7	−9.6	−0.8	2.3
TBL	SDI	0.6	6.5	14.7	15.4	−8.7	5.3	8.8	11.1
	I	**4.8**	**19.9**	**27.6**	**29.3**	**9.6**	**18.5**	**23.9**	**27.9**
	C	−5.9	7.3	11.1	11.6	−1.3	12.4	21.8	23.4

Table 2
The forecast skill score [%] over smart persistence for both GHI and DNI forecasts up to 4 h ahead using different input combinations, where "SDI" means satellite-derived irradiance, "I" denotes NWP irradiance, and "C" represents NWP cloud cover, at all SURFRAD stations.

Station	Method	GHI				DNI			
		1 h	2 h	3 h	4 h	1 h	2 h	3 h	4 h
BON	SDI+I	2.9	15.7	26.8	31.2	3.3	14.5	20.0	24.9
	I+C	−1.5	13.8	**28.0**	**35.4**	0.4	**16.4**	**25.6**	**30.2**
	SDI+C	0.9	5.3	13.1	15.2	−3.3	7.7	12.9	15.4
	SDI+I+C	**6.9**	**15.8**	27.3	33.1	**4.8**	15.6	21.6	22.3
DRA	SDI+I	**6.5**	13.8	17.3	15.4	4.8	16.1	20.3	21.2
	I+C	−1.8	13.6	20.0	**21.8**	−5.8	9.6	15.8	18.5
	SDI+C	3.5	13.9	16.5	16.5	−1.3	13.0	15.0	18.2
	SDI+I+C	4.7	**18.1**	**21.0**	20.8	**11.6**	**16.6**	**22.7**	**24.1**
FPK	SDI+I	12.4	19.3	24.6	24.7	11.5	19.6	23.8	23.1
	I+C	10.8	**19.6**	23.5	21.5	11.1	21.6	**26.7**	**25.2**
	SDI+C	−0.8	10.1	16.7	14.5	7.5	15.2	18.2	19.2
	SDI+I+C	**15.2**	19.2	**27.8**	**24.9**	**16.8**	**23.3**	26.4	24.3
GWN	SDI+I	−4.9	8.0	15.9	16.8	−2.4	9.7	19.7	25.7
	I+C	−0.8	10.7	18.3	21.2	−9.2	8.4	17.7	25.5
	SDI+C	−5.1	10.6	16.4	18.6	−10.7	6.1	14.7	22.9
	SDI+I+C	**0.1**	**14.1**	**23.1**	**26.8**	**−1.8**	**11.7**	**20.7**	**26.6**
PSU	SDI+I	**14.6**	24.2	28.7	32.4	**14.3**	**21.1**	23.6	28.3
	I+C	10.9	**25.2**	28.9	35.6	11.3	21.0	25.1	**33.4**
	SDI+C	3.4	14.8	21.1	27.0	7.0	16.6	23.8	27.5
	SDI+I+C	11.3	23.3	**29.4**	**36.4**	13.0	19.9	**24.5**	30.9
SXF	SDI+I	3.9	15.6	18.2	21.8	**7.7**	**15.2**	21.2	**23.7**
	I+C	−2.8	9.9	16.4	20.4	5.9	13.2	17.9	18.9
	SDI+C	−2.0	5.2	9.5	13.5	−7.1	3.3	9.5	10.9
	SDI+I+C	**8.0**	**16.3**	**22.1**	**24.6**	7.4	15.0	**21.3**	23.6
TBL	SDI+I	**16.9**	**22.2**	26.7	27.6	8.4	19.8	26.6	30.1
	I+C	9.3	21.1	26.5	27.4	8.5	18.3	**27.6**	**31.8**
	SDI+C	3.3	12.5	15.6	16.0	6.8	17.7	23.3	22.5
	SDI+I+C	8.7	20.8	**28.6**	**30.8**	**12.3**	**19.8**	25.9	28.8

show lower nRMSE than those of SDI+I and SDI+C for longer forecast horizons (e.g., 3–4-h ahead), while performance is compromised for the 1-h-ahead forecast.

The DNI forecast is generally accompanied by larger uncertainties than the GHI forecast, as shown in Fig. 4. When comparing the combined types of input, forecasts based on SDI+C exhibit larger biases for both GHI and DNI at most locations (e.g., BON, DRA, and SXF), with some site-specific variations. The bias of forecasts from SDI+C tends to increase with the forecast horizon. Forecasts based on SDI+I and I+C typically show comparatively lower biases than those of SDI+C; however, some higher nMBE values could also be observed (e.g., at FPK and TBL). Forecasts using SDI+I+C also exhibit some out-of-trend values of nMBE. The nMBE of DNI forecasts exhibits more fluctuations than that of GHI forecasts across all methods. However, no general

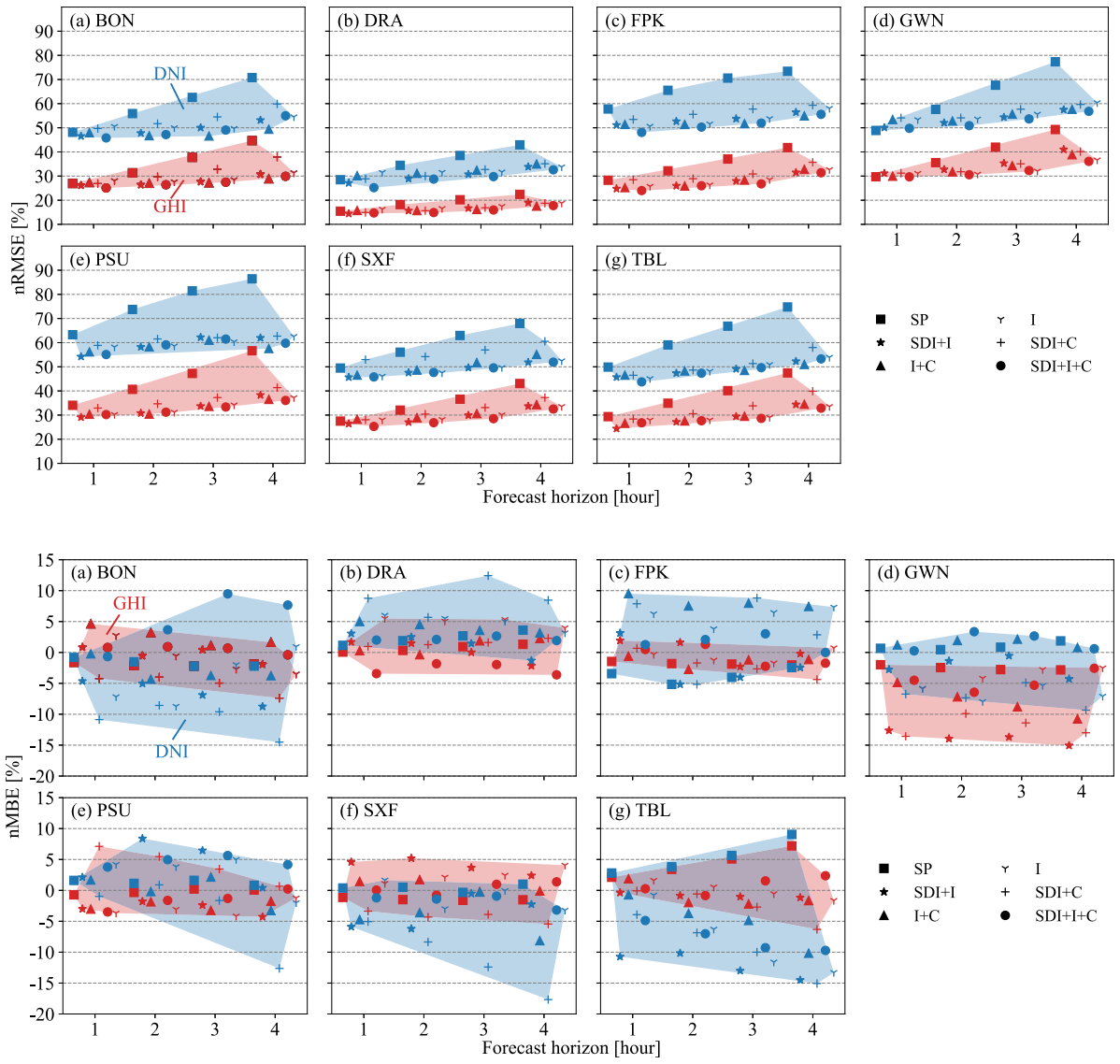


Fig. 4. The nRMSE and nMBE [%] for GHI (red) and DNI (blue) forecasts using combined types of spatio-temporal inputs across the forecast horizon up to 4 h ahead at seven SURFRAD stations. "SDI+I" refers data-driven forecasts using satellite-derived products and NWP irradiance as inputs; "SDI+C" refers to those using satellite-derived irradiance products and NWP cloud cover, and, "I+C" represents those using NWP irradiance and cloud cover, while "SDI+I+C" denotes those using satellite-derived products, NWP irradiance and cloud cover. Smart persistence (SP) and forecasts using the best single type input (I) are included as references. The shaded area covers all data points for better visualization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

trends could be observed regarding the bias associated with forecasts across the forecast horizon for all SURFRAD stations.

Table 2 presents the forecast skills of 1–4-h-ahead GHI and DNI forecasts using combined inputs at all SURFRAD stations. The forecast skill generally increases with the forecast horizon, regardless of which combinations of inputs are used. In most cases, the combination of SDI+I+C could produce more skillful 1–4-h-ahead GHI and DNI forecasts. However, there are some exceptions regarding climate zones (stations) and forecast horizons. As shown in Table 2, it is found that for shorter forecast horizons (i.e., 1–2-h ahead), the use of SDI can usually lead to improved forecasts, while this improvement may not extend to longer forecast horizons of 3–4-h ahead (see comparisons between I+C and SDI+I+C in Table 2). The addition of NWP cloud cover generally leads to higher forecast skills across the forecast horizon, with a few exceptions (e.g., the comparison of 1–2-h-ahead forecasts of SDI+I and SDI+I+C at PSU and TBL). Among the three types of spatio-temporal input, the integration of raw NWP irradiance shows more comparative benefits, as evidenced by the larger differences in forecast skills between SDI+C and SDI+I+C. The inclusion of satellite-derived irradiance can improve forecasting performance at shorter forecast horizons, while the addition of NWP cloud cover may introduce larger uncertainties; the combination of SDI+I+C could generally achieve overall higher skills in both GHI and DNI forecasting for all horizons.

3.3. Sensitivity analysis and ensemble forecasting

Overall, GHI and DNI forecasts using NWP irradiance achieve higher forecast skills when a single type of spatio-temporal input is used with deep learning (see Section 3.1). The combination of satellite-derived irradiance, NWP irradiance, and NWP cloud amount can result in improved forecasts when multiple inputs are integrated (see Section 3.2). However, forecasts based on either a single type of input or their combinations cannot ensure superior performance across all forecast horizons and stations. To further enhance forecasting performance, ensemble

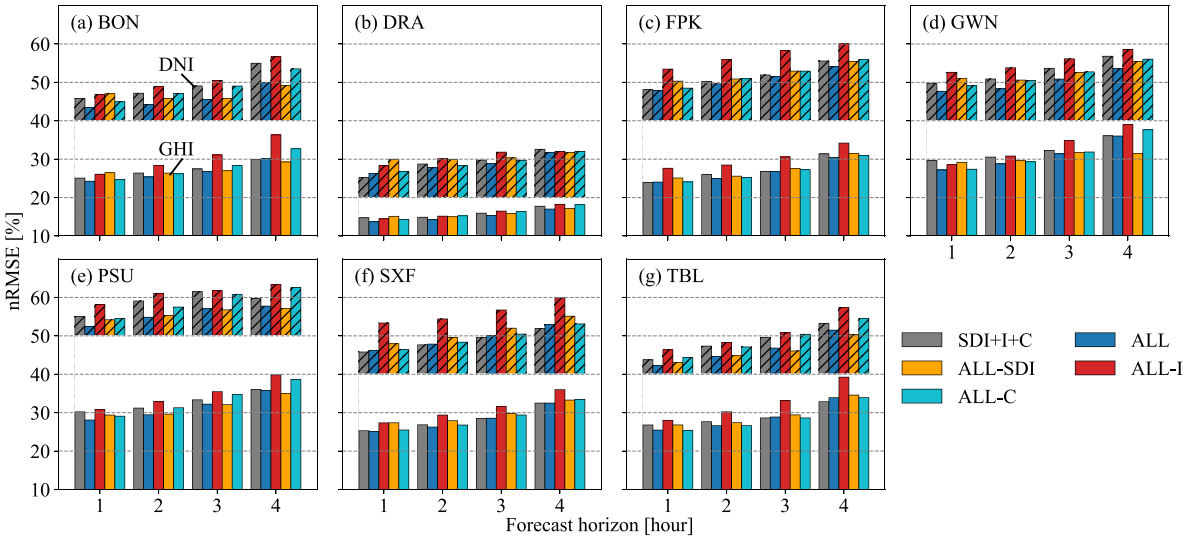


Fig. 5. The nRMSE [%] for GHI and DNI forecasts using different types of spatio-temporal inputs and their ensemble forecasts up to 4 h ahead at seven SURFRAD stations. "SDI+I+C" refers to forecasts using satellite-derived irradiance as well as NWP irradiance and cloud cover, "ALL" is the averaged forecast of the seven ensemble members, "ALL-SDI", "ALL-I", "ALL-C" represent the averaged ensemble forecast excluding members using "I", "SDI", and "C", respectively.

Table 3
The proposed strategies for weighted ensemble forecasting when the importance of used inputs is considered. The weights are equal for used members and are 0 for unused members.

Ensemble method	Forecast horizon	Used members (weights)	Unused members (weights)
ENS1	1–4-h ahead	I, SDI+I, I+C, SDI+I+C (0.25)	SDI, C, SDI+C (0)
ENS2	1–2-h ahead	SDI, I, SDI+I, SDI+I+C (0.25)	C, SDI+C, I+C (0)
	3–4-h ahead	I, C, I+C, SDI+I+C (0.25)	SDI, SDI+I, SDI+C (0)

forecasting is employed. Therefore, this section further analyzes the sensitivity of different spatio-temporal inputs in the ensemble forecasting scheme. Instead of simply averaging the forecasts generated using different inputs, two additional strategies for forecast combination are proposed based on the results of the sensitivity analysis.

The results for the sensitivity of different spatio-temporal inputs on GHI and DNI forecasts are presented in Fig. 5. It is shown that the averaged ensemble forecasts for both GHI and DNI generally outperform those using the SDI+I+C combination as inputs. As shown in Fig. 5, NWP irradiance generally exerts a relatively greater impact on ensemble forecasts for both GHI and DNI, with higher nRMSE observed when it is excluded from the ensemble calculations across different stations and forecast horizons. At shorter forecast horizons (i.e., 1–2-h ahead), satellite-derived irradiance typically has a greater influence than NWP cloud amount and, in some cases, even surpasses the impact of NWP irradiance. However, as forecast horizons extend, NWP cloud cover becomes increasingly influential, particularly for DNI forecasts.

Based on the sensitivity analysis, two weighted ensemble forecasting strategies are proposed to further improve forecasting performance. The basic concept of "weighted ensemble" is to produce averaged forecasts using selected ensemble members, whose weights are therefore higher than those of the excluded members (with a weight of 0). This illustrates the importance of the members used in ensemble forecasting. As shown in Table 3, one strategy (denoted as ENS1) is to generate ensemble forecasts based on NWP irradiance, either entirely or partially, since this input has a relatively higher impact than the others. In this case, the averages are derived from forecasts based on I, SDI+I, I+C, and SDI+I+C. Given that satellite-derived irradiance has a higher impact for shorter forecast horizons, and NWP irradiance has an overall greater influence on the ensemble forecast, the other strategy is to use members based on SDI for shorter forecast horizons (1–2-h ahead), and members based on NWP irradiance and cloud cover for longer forecast horizons (3–4-h ahead). As the combination of SDI+I+C shows relatively better performance, the forecast based on SDI+I+C is used for all forecast horizons. Specifically, the 1–2-h-ahead forecasts are averages of SDI, I, SDI+I, and SDI+I+C, while the 3–4-h forecasts are based on I, C, I+C, and SDI+I+C (denoted as ENS2 in Table 3).

As shown in Table 4, averaged ensemble forecasts (denoted as "ALL") typically have higher forecast skills than the best member (which is SDI+I+C as presented in Section 3.2), across different locations and forecast horizons. When comparing the averaged and proposed ensemble methods, ENS1 and ENS2 could further improve performance in most cases, especially for longer forecast horizons (i.e., 3–4-h ahead), as shown in Table 4. Although there are some exceptions where ALL performs slightly better, ENS1 and ENS2 still demonstrate higher forecast skills than SDI+I+C (e.g., at DRA and GWN). When comparing ENS1 and ENS2, neither exhibits dominant performance over the other; however, ENS1 performs slightly better in more cases.

To further evaluate the effectiveness of weighted ensemble forecasting methods, comparisons of 1–4-h-ahead GHI and DNI forecasts are performed at three representative SURFRAD stations, namely DRA, PSU, and SXF, for clear, mostly cloudy and medium cloudy sky conditions, respectively [13]. The marginal distributions of the measurements and 1–4-h-ahead forecasts of GHI and DNI are shown in Fig. 6. The numbers annotated in subplots represent the Wasserstein distances [56], which quantify the distance between two probability distributions. Lower Wasserstein distances indicate greater similarities between the two distributions [39]. As presented in Fig. 6, 1–4-h-ahead GHI forecasts of ENS1 and ENS2 generally show higher similarities to GHI measurements than those based on ALL. ENS1 performs slightly better than ENS2, as evidenced by lower values of Wasserstein distance in most conditions. Meanwhile, the Wasserstein distances between DNI forecasts and measurements are comparatively larger than those of GHI. Nevertheless, DNI forecasts based on ENS1 and ENS2 are likely to have closer distributions with DNI measurements than those of ALL. Similarly, DNI forecasts of ENS1 distribute closer to the measurements

Table 4
The forecast skill score [%] over smart persistence for both GHI and DNI forecasts up to 4 h ahead using different ensemble methods at all the SURFRAD stations.

Station	Method	GHI				DNI			
		1 h	2 h	3 h	4 h	1 h	2 h	3 h	4 h
BON	SDI+I+C[a]	6.9	15.8	27.3	33.1	4.8	15.6	21.6	22.3
	ALL[b]	**9.9**	**19.1**	29.0	32.5	**9.6**	**20.8**	27.3	29.4
	ENS1[c]	6.5	18.6	**30.5**	**36.2**	7.7	20.3	27.5	30.2
	ENS2[d]	9.7	18.4	29.4	35.3	9.4	19.1	**28.0**	**30.6**
DRA	SDI+I+C	4.7	18.1	21.0	20.8	**11.6**	16.6	22.7	24.1
	ALL	11.3	**21.2**	**24.0**	24.1	7.7	**19.4**	**25.1**	**28.4**
	ENS1	9.1	19.4	23.8	24.4	7.4	18.2	23.8	26.0
	ENS2	**11.9**	19.2	23.9	**24.9**	9.6	19.3	23.5	27.1
FPK	SDI+I+C	15.2	19.2	27.8	24.9	16.8	23.3	26.4	24.3
	ALL	14.9	22.4	27.7	27.0	17.1	24.3	27.0	26.3
	ENS1	**17.2**	**23.6**	**28.5**	**27.2**	17.4	**25.4**	**28.5**	**26.7**
	ENS2	16.5	23.0	27.8	26.4	**17.6**	24.3	26.4	25.5
GWN	SDI+I+C	0.1	14.1	23.1	26.8	−1.8	11.7	20.7	26.6
	ALL	**8.5**	**19.1**	25.2	26.9	**2.5**	**16.1**	**24.8**	**30.6**
	ENS1	1.9	14.9	23.9	27.1	1.0	15.6	24.1	29.5
	ENS2	8.3	18.5	**25.4**	**28.4**	1.8	14.9	23.5	29.4
PSU	SDI+I+C	11.3	23.3	29.4	36.4	13.0	19.9	24.5	30.9
	ALL	**17.5**	27.6	31.8	36.8	**17.1**	25.6	30.0	33.2
	ENS1	17.4	**28.4**	32.3	37.9	16.3	**25.7**	29.2	33.9
	ENS2	16.1	25.3	**32.6**	**38.6**	15.1	23.2	**30.5**	**34.6**
SXF	SDI+I+C	8.0	16.3	22.1	24.6	7.4	15.0	21.3	23.6
	ALL	8.7	18.1	21.9	24.6	6.7	14.6	20.4	22.1
	ENS1	7.6	18.6	**22.7**	**25.4**	**11.4**	**18.3**	**23.5**	**25.2**
	ENS2	**9.4**	**19.3**	21.8	25.0	8.1	15.3	20.1	21.9
TBL	SDI+I+C	8.7	20.8	28.6	30.8	12.3	19.8	25.9	28.8
	ALL	13.2	23.7	28.0	28.5	15.1	24.4	30.1	31.1
	ENS1	**14.5**	**25.9**	**30.9**	**31.8**	**15.5**	23.8	30.2	32.8
	ENS2	13.7	24.3	28.3	29.6	12.9	**24.7**	**31.1**	**33.0**

[a] "SDI+I+C" refers to forecasts based on the combination of satellite-derived irradiance, NWP irradiance and cloud cover.
[b] "ALL" is the averaged forecast based on all the seven ensemble members.
[c] "ENS1" is weighted ensemble method as presented in Table 3.
[d] "ENS2" is weighted ensemble method as presented in Table 3.

than those from ENS2, while ENS2 is sometimes inferior to ALL in terms of DNI forecasting, as the Wasserstein distance could become larger. The joint distributions are presented in Appendix C.

4. Discussion

When a single type of spatio-temporal input is considered (Section 3.1), the largest errors in raw NWP irradiance forecasts of GHI and DNI indicate that calibration is necessary when such predictions are directly used in resource-to-power models or grid-integration applications [39]. In contrast, data-driven forecasts generated using NWP irradiance exhibit higher accuracy, while those based on satellite-derived irradiance or NWP cloud cover are comparatively less skillful, often resulting in negative skill scores for 1-h-ahead forecasts. This could be attributed to the following reasons:

- Satellite-derived irradiance products based solely on past atmospheric information might not effectively extrapolate to predict future conditions.
- NWP cloud forecasts are typically associated with greater uncertainties compared to NWP irradiance predictions.
- The influence on ground-level irradiance is governed by nonlinear radiative transfer processes in the atmosphere, which may not be adequately captured by deep learning models.
- The relatively limited spatial coverage of satellite-based inputs could also impact forecasting performance, particularly for longer forecast horizons.

Nevertheless, satellite-derived irradiance has positive impacts on forecasting at shorter horizons compared to NWP cloud cover, as evidenced by the higher 1-h-ahead forecast skills (see Table 1). The importance of satellite-derived irradiance for shorter horizons has also been demonstrated in [11]. This suggests the need for improvements in the accuracy of satellite-to-irradiance for solar forecasting [13]. As the forecast horizon extends, NWP cloud amount becomes more beneficial because NWP cloud cover is more informative than satellite-derived irradiance regarding future atmospheric conditions [11].

When comparing forecasts based on combined spatio-temporal inputs (Section 3.2), the integrated use of SDI+I+C generally produces more skillful GHI and DNI forecasts. The benefits of input combination have also been demonstrated in other studies; for example, combined sky and satellite images in [28]. This is because integrating all spatio-temporal inputs provides more information on atmospheric status [11]. However, more careful consideration should be given to site-specific variations among stations and forecast horizons:

- Although the combination of inputs generally leads to improved forecasts, the high uncertainty of NWP cloud cover remains an issue. When integrated with satellite-derived irradiance for intra-day solar forecasting, NWP irradiance is found to be a better input than NWP cloud amount. Given its site-specific impacts on different climate zones, the use of NWP cloud cover in solar forecasting should be further investigated.
- Satellite-derived irradiance demonstrates its importance in shorter horizon solar forecasting, even when combined with other spatio-temporal inputs [11]. This again demonstrates the necessity of improvements in satellite-to-irradiance.
- The understanding of the impacts of different inputs on forecasting performance could be enhanced with explainable machine learning [57], which could provide more information on input selection and support the design of solar forecasting models.

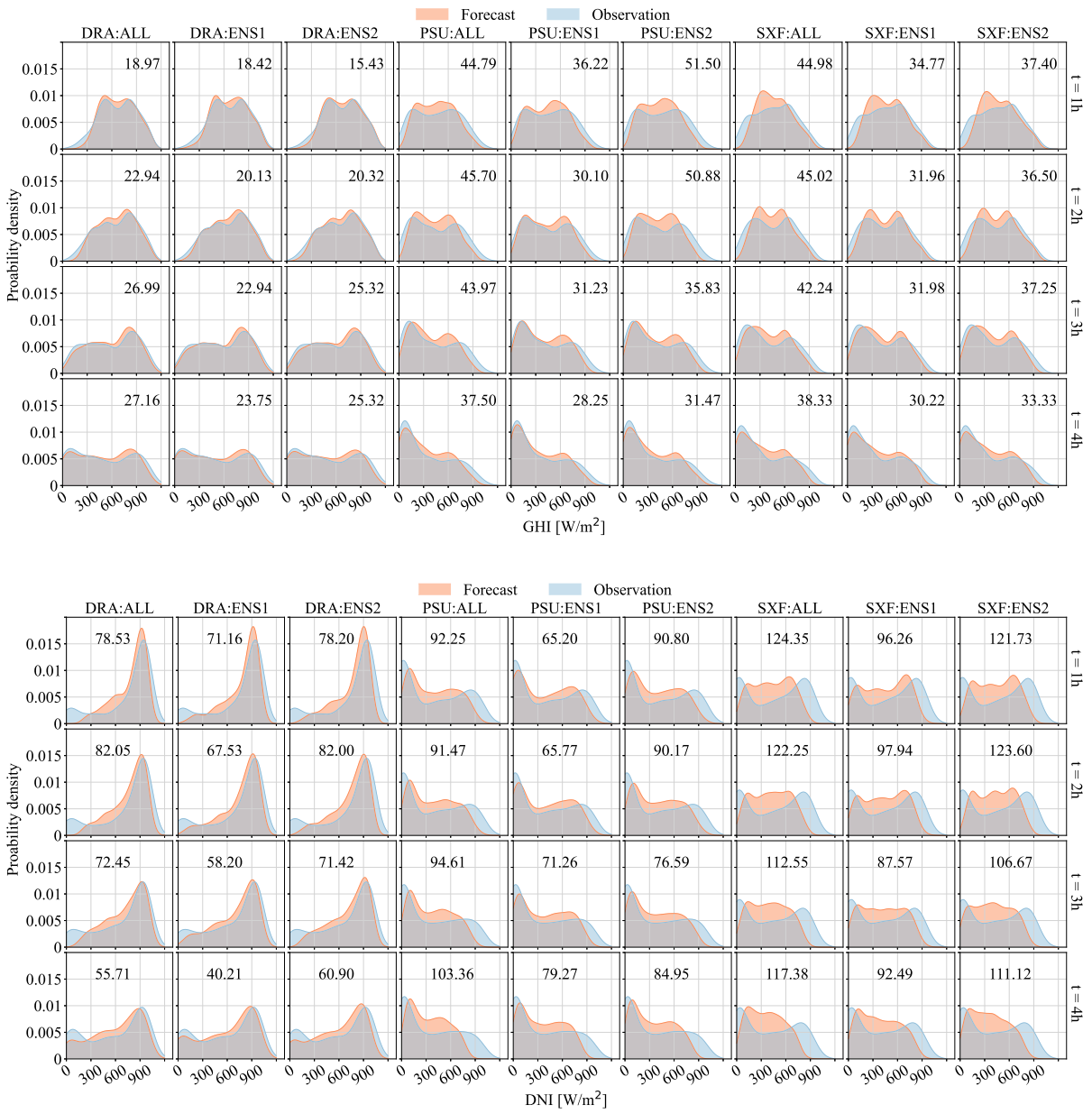


Fig. 6. Marginal distributions of 1–4-h-ahead GHI and DNI forecasts and measurements, from three ensemble forecasting methods (i.e., ALL, ENS1, and ENS2), at DRA, PSU, and SXF. The annotated number in each subplot is the Wasserstein distance between two distributions, a smaller value indicates a higher similarity.

Forecasts of GHI and DNI based on SDI+I+C do not always demonstrate dominant performance; therefore, sensitivity analysis and ensemble forecasting are further considered (Section 3.3). Ensemble methods can further reduce forecast errors due to the advantage of forecast combinations [7]. However, most ensemble methods are based solely on the arithmetic mean of all members without considering their impact on forecasting [7]. Based on the sensitivity analysis, two strategies, ENS1 and ENS2, for weighted ensembles that consider the impact of different inputs are proposed:

- Both ENS1 and ENS2 lead to improvements in forecasting performance, with ENS1 performing slightly better than ENS2 in most cases. This difference may stem from the features of ensemble methods, such as target variables, climate zones, and forecast horizons.
- Although ensemble methods can enhance forecasting performance by leveraging the forecasts of ensemble members, it remains essential to ensure that the ensemble members provide high-quality forecasts.
- In the context of data-driven intra-day solar forecasting, a weighted average of ensemble members, considering the importance of spatio-temporal inputs and their influences on different forecast horizons, demonstrates greater potential compared to the simple averaging.

Since high-quality solar power predictions are usually generated using accurate irradiance forecasts [1], this study on intra-day GHI and DNI forecasting provides novel insights for practical solar energy applications. In a real-world case, satellite-derived irradiance can be obtained using near real-time satellite measurements (a 30-min latency is considered when designing the model, as detailed in Section 2.4), and hourly updated HRRR forecasts are already operationally accessible. This paves the way for operational forecasting, where solar power generation can be obtained almost simultaneously using the solar power curve and irradiance forecasts [23]. These improved forecasts are essential for integrating such non-dispatchable solar energy into the power grid by supporting operations, scheduling, and unit commitment

with reduced costs [3]. It is reported that a 3% accuracy improvement in solar power forecasts can reduce scheduling reserve costs by 5% and generation curtailment costs by 38% [9]. This is particularly important for regions with high solar energy potential and supportive policies, such as Australia [58], Spain [59], and China [60]. Moreover, geostationary satellites and NWP models provide near-global coverage, even though the data are from different sources. This means that the proposed data-driven approach has superior adaptability regarding the choice of spatio-temporal inputs and ensemble forecasting compared with other methods for operational solar forecasting.

Although the data used in this work are from the United States, and the relative importance of satellite- and NWP-based inputs may vary for regions with significantly different climates, the methodology can be applied elsewhere and tailored according to local data availability. While high-quality and long-term on-site irradiance measurements remain the primary option for developing and testing data-driven solar forecasting models, satellite-derived irradiance has demonstrated its effectiveness in solar forecasting when used as the target [61]. Therefore, the need for ground irradiance measurements is no longer the primary constraint; rather, the regional NWP forecasts serve as the main limitation in the development of data-driven and ensemble solar forecasting methods. Future work should include more locations with varying climates and international stations when the data become available.

5. Conclusion

With advancements in regional weather forecast models and improved computing capabilities, the spatio-temporal resolution of NWP models has significantly increased. On this account, this study incorporates high-resolution spatio-temporal NWP forecasts with satellite-based irradiance products in data-driven intra-day solar forecasting (both GHI and DNI) for the first time. Since each spatio-temporal input has its unique characteristics and varying impacts on solar forecasting, the effects of different types of spatio-temporal inputs, namely satellite-derived irradiance and NWP forecasts of irradiance and cloud amount, are examined using data-driven methods for intra-day GHI and DNI forecasting. The analysis is conducted from three perspectives: a single type of input, a combination of different inputs, and ensemble forecasting. The findings suggest that the inclusive use of satellite- and NWP-based inputs generally improves forecasting accuracy compared to their individual use. While NWP irradiance shows a higher overall influence on the forecast, satellite-derived irradiance contributes more to shorter-term predictions (less than 2 h ahead) and NWP cloud cover is more influential for longer forecast horizons. This suggests that selecting informative inputs can enhance forecasts without the need to advance forecasting models.

Based on the sensitivity analysis and the impacts of various inputs on the forecast, two strategies for weighted ensemble forecasting, which account for the importance of ensemble members and the types of spatio-temporal inputs, are proposed to further enhance overall forecasting accuracy. These simple yet effective ensemble strategies can be important tools for operational forecasting. Given the growing importance of accurate solar irradiance forecasts, this work provides practical guidelines for designing more effective data-driven forecasting methods using spatio-temporal satellite- and NWP-based inputs to support the integration of solar energy. Even for regions with completely different climates, the methodology on quantifying the input importance and designing the ensemble method can still be applied considering that satellite and NWP data have global coverage. For future research, several aspects can be explored: (i) a more generalized model can be developed for a wider range of international and climatic regions as suitable data become available; (ii) the sensitivity analysis can be investigated using explainable machine learning to provide additional insights for designing forecasting models; and (iii) more advanced strategies for weighted ensemble forecasting can be developed to further enhance the performance.

CRediT authorship contribution statement

Shanlin Chen: Writing – original draft, Visualization, Methodology, Investigation, Data curation. **Tao Jing:** Writing – original draft, Methodology. **Dazhi Yang:** Writing – review & editing, Methodology. **Yuntian Chen:** Writing – review & editing, Methodology. **Mengying Li:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Yinghao Chu:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

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

During the preparation of this work, the authors used 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 published article.

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 of the Hong Kong University Grants Committee (UGC) through the Research Grants Council under project numbers 25213022, C6003-22Y, and 21200424. Additional support from the Guangdong Basic and Applied Basic Research Foundation under project number 2024A1515010117 is also sincerely acknowledged.

Appendix A. Satellite-to-irradiance with deep learning

The deep learning satellite-to-irradiance method is implemented through the following two steps: (i) a pre-trained deep learning model is developed for irradiance estimations at a single location centered in the domain of satellite images (11 × 11 pixels); (ii) the pre-trained deep learning model is subsequently applied to obtain the spatial irradiance estimates for the area of interest. As shown in Fig. A.1, to obtain irradiance estimations for the area of interest (11 × 11 pixels), a larger region of satellite images (21 × 21 pixels) is required since the inputs of the deep learning satellite-to-irradiance model for a single location are satellite images measuring 11 × 11 pixels. Therefore, an 11 × 11-pixel window is shifted pixel-wise over the region of satellite images (21 × 21 pixels) to generate irradiance estimations for 121 locations within the area of interest. The spatial estimates of GHI and DNI are obtained using the same methodology, while the pre-trained models are developed separately.

Appendix B. Deep learning framework and hyperparameter tuning

The structure and hyperparameters of the deep learning model are detailed in Table B.1. It should be noted that although the forecasting frameworks for GHI and DNI are similar, the models are trained, and hyperparameters are tuned separately. To enhance generalizability and prevent overfitting, regularization techniques such as L2 regularization and early stopping are applied, as outlined in Table B.1.

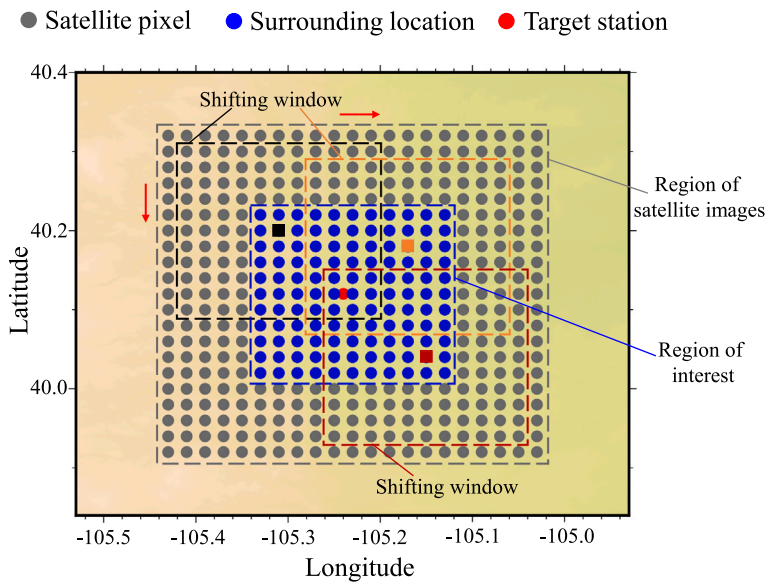


Fig. A.1. The illustration of spatial irradiance estimations using the pre-trained deep learning model with satellite images as inputs. The 11 × 11 shifting window moves pixel-wisely for 121 times over the region of satellite images (21 × 21 pixels) to obtain the spatial irradiance estimates for the area of interest (11 × 11 pixels) [13].

Table B.1
The structure of the data-driven deep learning forecasting model and hyperparameters tuning ranges. The hyperparameters are tuned with Bayesian optimization using the KerasTuner [62].

Structure	Hyperparameter[a]	Values
Inputs (single or combined) → BatchNormalization	–	–
→ Conv2D(x_1, 3 × 3, activation$_1$) + MaxPool2D(2 × 2)	x_1	Range(min = 8, max = 64, step = 8)
	activation$_1$	[relu, gelu, selu, tanh]
→ Conv2D(x_2, 3 × 3, activation$_2$) + Attention layer	x_2	Range(min = 8, max = 64, step = 4)
	activation$_2$	[relu, gelu, selu, tanh]
→ Flatten	–	–
→ Concatenate (two or more types of inputs)	–	–
→ Dense(x_3, activation$_3$, L2 regularization)	x_3	Range(min = 8, max = 64, step = 4)
	activation$_3$	[relu, gelu, selu, tanh]
→ Dense(x_4, activation$_4$)	x_4	Range(min = 4, max = 16, step = 4)
	activation$_4$	[relu, gelu, selu, linear]
→ Dense(4, relu) → Outputs	–	–
	Optimizer	Adam
	Loss function	Huber
	Learning rate	[1e−5, 1e−4, 2e−4, 5e−4, 1e−3, 2e−3]
	Early Stopping	With the patience equal to 5

[a] x_1 and x_2 refer to the number of filters in the convolution, x_3 and x_4 are the number of units, and activation$_i$ means the activation function.

Appendix C. Further evaluation of the weighted ensemble forecasting

The joint distributions of 1–4-h-ahead GHI and DNI forecasts and measurements are presented in Fig. C.1, where the number in each subplot is the coefficient of determination (R^2). The density of both GHI and DNI forecast-measurement pairs is higher along the identity line. Compared to GHI forecasts, DHI forecasts show a sparser distribution of forecast-measurement points and lower R^2 values. Despite some variations, DNI and GHI forecasts based on ENS1 and ENS2 show higher R^2 values than those of ALL in most cases. Overall, it is observed that forecasts for ENS1 and ENS2 have distributions (marginal and joint) that are generally closer to the measurements than forecasts of ALL, and ENS1 typically performs slightly better than ENS2 for both GHI and DNI forecasting. Moreover, there are still some under- and over-predictions in the forecasts from the ensemble methods.

Data availability

The data used in this work are publicly available. SURFRAD measurements are available at https://gml.noaa.gov/aftp/data/radiation/surfrad/, satellite data of GOES-16 can be accessed at https://noaa-goes16.s3.amazonaws.com/index.html, NSRDB data are accessible via https://nsrdb.nrel.gov/data-viewer, and HRRR forecasts can be found at https://noaa-hrrr-bdp-pds.s3.amazonaws.com/index.html.

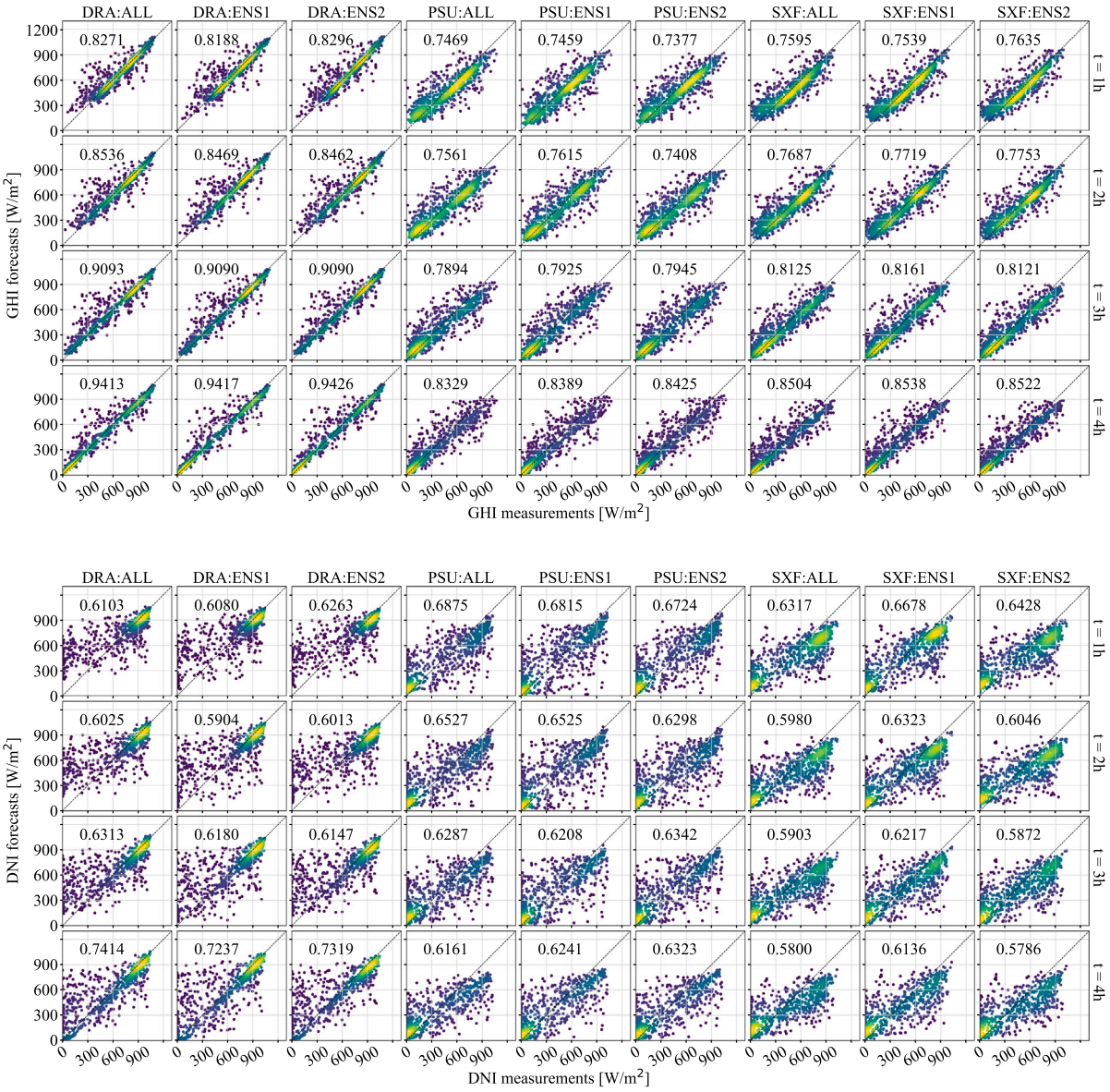


Fig. C.1. Scatter plots of 1–4-h-ahead GHI and DNI forecasts and measurements, from three ensemble forecasting methods (i.e., ALL, ENS1, and ENS2), at DRA, PSU, and SXF. The annotated number in each subplot is the coefficient of determination (R^2).

References

[1] 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.

[2] 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.

[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] Chu Y, Li M, Coimbra CFM, Feng D, Wang H. Intra-hour irradiance forecasting techniques for solar power integration: A review. iScience 2021;103136.

[6] Alkhayat G, Mehmood R. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. Energy AI 2021;4:100060.

[7] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. Renew Sustain Energy Rev 2022;154:111768.

[8] 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.

[9] Wang W, Guo Y, Yang D, Zhang Z, Kleissl J, van der Meer D, et al. Economics of physics-based solar forecasting in power system day-ahead scheduling. Renew Sustain Energy Rev 2024;199:114448.

[10] Sardarabadi A, Ardakani AH, Matrone S, Ogliari E, Shirazi E. Multi-temporal PV power prediction using long short-term memory and wavelet packet decomposition. Energy AI 2025;100540.

[11] Chen S, Jing T, Li M, Lee HH, Lam MC, Bu S. Interpretable transformer based intra-day solar forecasting with spatiotemporal satellite and numerical weather prediction inputs. Energy AI 2025;100667.

[12] Liu M, Ling Q. Spatial–temporal multimodal fusion model for intra-hour solar power forecasting under variable weather conditions. Renew Energy 2025;248:123043.

[13] 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.

[14] Song Z, Xiao F, Chen Z, Madsen H. Probabilistic ultra-short-term solar photovoltaic power forecasting using natural gradient boosting with attention-enhanced neural networks. Energy AI 2025;20:100496.

[15] 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.

[16] Aicardi D, Musé P, Alonso-Suárez R. A comparison of satellite cloud motion vectors techniques to forecast intra-day hourly solar global horizontal irradiation. Sol Energy 2022;233:46–60.

[17] 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.
[18] Cui Y, Wang P, Meirink JF, Ntantis N, Wijnands JS. Solar radiation nowcasting based on geostationary satellite images and deep learning models. Sol Energy 2024;282:112866.
[19] Stull RB. Practical meteorology: an algebra-based survey of atmospheric science. University of British Columbia; 2015.
[20] Larson VE. Forecasting solar irradiance with numerical weather prediction models. Sol Energy Forecast Resour Assess 2013;299–318.
[21] Mathiesen P, Collier C, Kleissl J. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. Sol Energy 2013;92:47–61.
[22] Yang D. On post-processing day-ahead NWP forecasts using Kalman filtering. Sol Energy 2019;182:179–81.
[23] Yang D, Liu B, Zhang H, Xia X, Shen Y, Mayer MJ. A second tutorial review of the solar power curve: Applications in energy meteorology. Adv Atmospheric Sci 2025;42(2):269–96.
[24] Voyant C, Muselli M, Paoli C, Nivet M-L. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. Energy 2012;39(1):341–55.
[25] Phan Q-T, Wu Y-K, Phan Q-D. Short-term solar power forecasting using xgboost with numerical weather prediction. In: 2021 IEEE international future energy electronics conference. IFEEC, IEEE; 2021, p. 1–6.
[26] Andrade JR, Bessa RJ. Improving renewable energy forecasting with a grid of numerical weather predictions. IEEE Trans Sustain Energy 2017;8(4):1571–80.
[27] Suksamosorn S, Hoonchareon N, Songsiri J. Influential variable selection for improving solar forecasts from numerical weather prediction. In: 2018 15th international conference on electrical engineering/electronics, computer, telecommunications and information technology. ECTI-CON, IEEE; 2018, p. 333–6.
[28] 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.
[29] Hashimoto A, Yoshimoto K. Development of a short-term solar irradiance forecasting using satellite image in combination with numerical weather prediction model. Electr Eng Japan 2023;216(3):e23432.
[30] Lorenz E, Kühnert J, Heinemann D. Short term forecasting of solar irradiance by combining satellite data and numerical weather predictions. In: Proceedings of the 27th European PV solar energy conference (EU PVSEC), frankfurt, Germany. vol. 2428, 2012, 44014405.
[31] Miller SD, Rogers MA, Haynes JM, Sengupta M, Heidinger AK. Short-term solar irradiance forecasting via satellite/model coupling. Sol Energy 2018;168:102–17.
[32] Lopes FM, Conceição R, Silva HG, Salgado R, Collares-Pereira M. Improved ECMWF forecasts of direct normal irradiance: A tool for better operational strategies in concentrating solar power plants. Renew Energy 2021;163:755–71.
[33] 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.
[34] 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.
[35] Yang D. Ensemble model output statistics as a probabilistic site-adaptation tool for satellite-derived and reanalysis solar irradiance. J Renew Sustain Energy 2020;12(1).
[36] 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.
[37] 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.
[38] 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.
[39] 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.
[40] Yang D. SolarData: An R package for easy access of publicly available solar datasets. Sol Energy 2018;171:A3–12.
[41] Chen S, Li M. Improved turbidity estimation from local meteorological data for solar resourcing and forecasting applications. Renew Energy 2022;189:259–72.
[42] 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.
[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] 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.
[45] Buster G, Bannister M, Habte A, Hettinger D, Maclaurin G, Rossol M, et al. Physics-guided machine learning for improved accuracy of the National Solar Radiation Database. Sol Energy 2022;232:483–92.
[46] Yang D. Validation of the 5-min irradiance from the National Solar Radiation Database (NSRDB). J Renew Sustain Energy 2021;13(1):016101.
[47] Chen S, Jing T, Li M, Lee HH, Bu S. Improved satellite-based estimation of direct normal solar irradiance using clearness index with deep learning methods. Sol Energy 2025;302:113991.
[48] 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.
[49] Yang D. Choice of clear-sky model in solar forecasting. J Renew Sustain Energy 2020;12(2):026101.
[50] 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.
[51] 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.
[52] Matsunobu LM, Pedro HTC, Coimbra CFM. Cloud detection using convolutional neural networks on remote sensing images. Sol Energy 2021;230:1020–32.
[53] Liao Z, Coimbra CFM. Hybrid solar irradiance nowcasting and forecasting with the scope method and convolutional neural networks. Renew Energy 2024;232:121055.
[54] 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.
[55] 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.
[56] Panaretos VM, Zemel Y. Statistical aspects of Wasserstein distances. Annu Rev Stat Appl 2019;6:405–31.
[57] Machlev R, Heistrene L, Perl M, Levy KY, Belikov J, Mannor S, et al. Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy AI 2022;9:100169.
[58] Poddar S, Kay M, Prasad A, Evans JP, Bremner S. Changes in solar resource intermittency and reliability under Australia's future warmer climate. Sol Energy 2023;266:112039.
[59] Cabello-López T, Carranza-García M, Riquelme JC, García-Gutiérrez J. Forecasting solar energy production in Spain: A comparison of univariate and multivariate models at the national level. Appl Energy 2023;350:121645.
[60] Yang D. The future of solar forecasting in China. J Renew Sustain Energy 2023;15(5).
[61] 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.
[62] O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al. KerasTuner. 2019, https://github.com/keras-team/keras-tuner.