



Prediction of solar irradiance based on Python

LITING YAN, AO YU, GE ZHANG and JINYE ZHANG*

Hubei University of Technology, School of Science, Wuhan, China, 430 068

Hubei Engineering technology Research Center of Energy Photoelectric Device and System,

Hubei University of Technology

Hubei University of Technology, Hubei Collaborative Innovation Center

for High-efficiency Utilization of Solar Energy, Wuhan, China, 430068

(Received 10 August 2022, Accepted 11 May 2023)

e mail : zhangjinye@hbut.edu.cn

सार — आधुनिक औद्योगिक समाज का तीव्र विकास सस्ते और प्रचुर मात्रा में जीवाश्म ईंधन ऊर्जा पर बहुत अधिक निर्भर रहा है। हालाँकि, सतत विकास के लिए, प्रकाश वोल्टिक (पीवी) और पवन ऊर्जा जैसे नए ऊर्जा स्रोतों को विकसित करने पर ध्यान केंद्रित किया जा रहा है। बिजली उत्पन्न करने के लिए सौर विकिरण का उपयोग करने के संदर्भ में, कुशल उपयोग के लिए सौर ऊर्जा का पहले से पूर्वानुमान करना महत्वपूर्ण है। यह शोध पत्र स्पष्ट आकाश परिस्थितियों में तीन प्रकार के विकिरण का पूर्वानुमान करने के लिए pvlib-python मॉडल का उपयोग करता है: POA_DNI, POA_GHI, और POA_DHI। इसके अलावा, हम पूर्वानुमान की सटीकता में सुधार के लिए pvlib से वायु विलय डेटा शामिल करते हैं। BSRN से तीन साइटों का चयन किया जाता है और मॉडल की पूर्वानुमान प्रभावशीलता का मूल्यांकन करने के लिए अनुमानित डेटा की तुलना प्रेक्षित डेटा से की जाती है। परिणाम दर्शाता है कि POA_GHI के लिए मॉडल सबसे अच्छा प्रदर्शन करता है और पूर्वानुमान की सटीकता वास्तविक मेघावरण वितरण का पर उल्लेखनीय प्रभाव पड़ता है।

ABSTRACT. The rapid development of modern industrial society has relied heavily on cheap and abundant fossil fuel energy. However, to achieve sustainable development, there is an increasing focus on developing new energy sources such as photovoltaics (PV) and wind energy. In the context of using solar irradiance to generate electricity, predicting the solar power in advance is crucial for efficient utilization. This paper utilizes the pvlib-python model to predict three types of irradiance in clear sky conditions: POA_DNI, POA_GHI, and POA_DHI. Furthermore, we incorporate aerosol data from pvlib to improve the prediction accuracy. Three sites from BSRN are selected and the predicted data are compared with the observed data to evaluate the model's prediction effectiveness. The result reveals that the model performs best for POA_GHI and the actual cloud cover distribution has a significant impact on the prediction accuracy.

Key words – pvlib-python, Solar irradiance, Aerosol, Photovoltaics, Point prediction

1. Introduction

1.1. Background

Today, oil accounts for 40.28% of gross energy consumption, natural gas accounts for 16.37% and coal accounts for 9.52% [IEA (International Energy Agency), Electricity generation by source, 2022]. The world's oil reserves have been greatly reduced due to overdependence on fossil fuels (Zhang *et al.*, 2017). In response, there has been increasing attention on renewable energy sources as

a means of achieving long-term sustainable development. Fig. 1 illustrates global power generation from different renewable energy sources, with solar energy having experienced particularly rapid development in recent years.

In 2009, the total global solar power generation was 19,859 GWh, accounting for only 0.95% of the world's total power generation. However, by 2019, the total global solar power generation had increased to 680,952 GWh. It accounted for 2.52% of total global electricity generation and 10.0 % of the total renewable energy generation

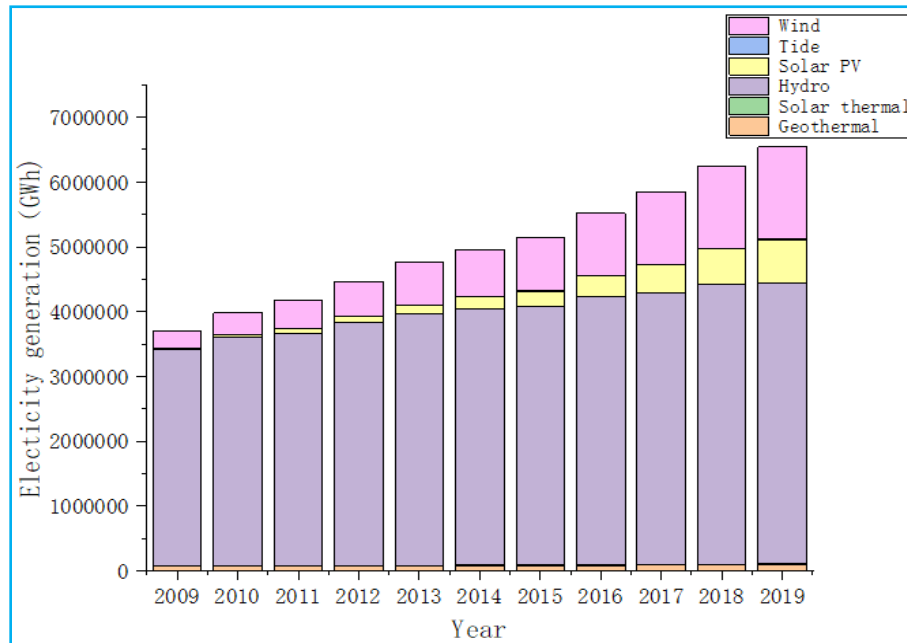


Fig. 1. Global power generation by energy sources from 2009 to 2019 (Source : International Energy Agency, 2022)

[IEA (International Energy Agency), Electricity generation by source, 2022]. Its share is still small, but the progress is fast. In China, solar power generation has also grown year by year, from 700 GWh in 2010 to 269,726 GWh in 2020. In 2020, China's solar power generation accounted for 12.98% of total renewable energy generation [IEA (International Energy Agency), Renewable electricity generation by source in China, 2022]. In the first 11 months of 2021, China's solar power generation reached 300,900 GWh, a year-on-year increase of 24.3% [NEA (National Energy Administration), New energy generation by the National Energy Administration, 2022]. With the maturation of power generation technology and the reduction of costs, the solar power generation industry is developing rapidly. Photovoltaic power generation is accounting for an increasing proportion of renewable energy generation (Inman *et al.*, 2013).

Accurate forecasting is essential for developing a reasonable power generation plan. More precise solar irradiance predictions can lead to more efficient utilization of solar energy. In this paper, we utilize pvlib-python to extract meteorological predictions from the GFS model, and then apply a PV model to process these meteorological data and get irradiance predictions.

1.2. Literature review

There are various methods for predicting PV power output. The current mainstream prediction models can be

classified into three categories : point prediction, interval prediction, and probabilistic prediction (Lai *et al.*, 2019). The following section briefly describes these three methods.

1.2.1. Point prediction model

Point prediction involves obtaining a definite value at a specific moment. Most of the current PV prediction research focuses on point prediction. Although point prediction is relatively straightforward, it is susceptible to various meteorological factors, resulting in significant uncertainty. When weather conditions change greatly, the fluctuation of meteorological data and PV output power will also be large. Moreover, the limited data contained in point prediction makes it difficult to express the randomness of the prediction results (Ding *et al.*, 2021). Point prediction can be subdivided into physical methods, statistical methods, metaheuristic learning methods, and combinatorial methods.

Meta-heuristic learning methods simulate biological activities and use algorithms to establish the relationship between input and output (Lai *et al.*, 2019). (Mellit *et al.*, 2013) and (Almonacid *et al.*, 2014) used an artificial neural network model to predict PV power generation. Such methods are the mainstream approaches for PV power generation prediction, but the prediction accuracy is still limited by the accuracy of meteorological data. (Yadav and Chandel, 2017) and (Mughal *et al.*, 2021) used Linear Regression Models to optimize power

TABLE 1

Application and research of pvlib-python in recent years

Article	Contribution
Evaluation of statistical learning configurations for gridded solar irradiance forecasting. (Gagne <i>et al.</i> , 2017)	Through evaluating different statistical learning models used in the current python database for grid irradiance prediction, they find that the choices of statistical learning model, interpolation scheme, and loss function have the biggest impacts on performance. Errors tend to be lower with sunnier weather
Use of measured aerosol optical depth and precipitable water to model clear sky irradiance. (Mikofski <i>et al.</i> , 2017)	The effects of atmospheric composition on clear sky irradiance of year to year variations in three clear sky models were studied using measurements of aerosol optical depth (AOD) and precipitable water (Pwat) at seven locations in the United States
A transferable turbidity estimation method for estimating clear-sky solar irradiance. (Chen <i>et al.</i> , 2023)	A transferable turbidity estimation method is developed involving stations with sufficient information and applied at locations with limited data availability. And the results are compared with the pvlib model
A photovoltaic power output dataset: Multi-source photovoltaic power output dataset with python toolkit. (Yao <i>et al.</i> , 2021)	They release a PV power output dataset (PVOD), which contains metadata, numerical weather prediction data, and local measurements data from 10 PV systems located in China. In PVOD, a Python toolkit with basic functions for data access and preprocessing is provided
Introduction, evaluation and application of an energy balance model for photovoltaic modules. (Heusinger <i>et al.</i> , 2020)	This study presents a new energy balance model that accurately simulates the complete diurnal dynamics of PV thermal behavior with routinely available meteorological input. The model is available as a stand-alone program written in Python
Global horizontal irradiance forecast for Finland based on geostationary weather satellite data. (Kallio-Myers and Riihelä, 2020)	They present the development and validation of a satellite-based GHI forecast for southern Finland. The forecast is formed by combining information from the clear sky (CS) model pvlib Solis with consecutive geostationary weather satellite imagery
Solar radiation and photovoltaic systems: Modeling and simulation. (Adel Mellit and Soteris Kalogirou, 2022)	The paper provides a short introduction to solar radiation and a brief introduction to photovoltaics, including solar cell conversion, photovoltaic technologies, latest solar cells efficiency, photovoltaic configuration, and type of photovoltaic systems
Cloud cover bias correction in numerical weather models for solar energy monitoring and forecasting systems with kernel ridge regression. (Deo <i>et al.</i> , 2023)	They propose a new kernel ridge regression (KRR) model to reduce bias in Total Cloud Cover (TCDC) simulations for medium-term prediction at the inter-daily. The proposed KRR model is evaluated against multivariate recursive nesting bias correction (MRNBC), a conventional approach and eight machine learning (ML) methods
An archived dataset from the ECMWF Ensemble Prediction System for probabilistic solar power forecasting. (Wang <i>et al.</i> , 2022)	To facilitate the uptake of ensemble NWP forecasts in solar power forecasting research, this paper offers an archived dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System, over a four-year period (2017–2020) and over an extensive geographical region (<i>e.g.</i> , most of Europe and North America), under the Creative Commons Attribution 4.0 International (CC BY 4.0) license

prediction. (Theocharides *et al.*, 2020) used linear regression analysis and K-means clustering evaluation to improve the accuracy of the predicted data. (Fermín Rodríguez *et al.*, 2021) proposed a novel solar irradiance forecaster which combines the advantages of machine learning and optimization of spatial-temporal parameters. (Wang *et al.*, 2017) developed a new hybrid PV power prediction method based on wavelet transform (WT) and deep convolutional neural network (DCNN).

A Markov chain model has been developed to predict the output of PV plants using historical power data (Ding and Xu, 2011). To enhance the accuracy and autocorrelation characteristics of PV output series, Ding Ming refined data fragment classification and built an improved Markov chain model (Jiang *et al.*, 2019). Jiang Feng applied the gray model for the overall trend prediction of PV power output and established a gray-weighted Markov Chain estimation model (Jiang *et al.*, 2019).

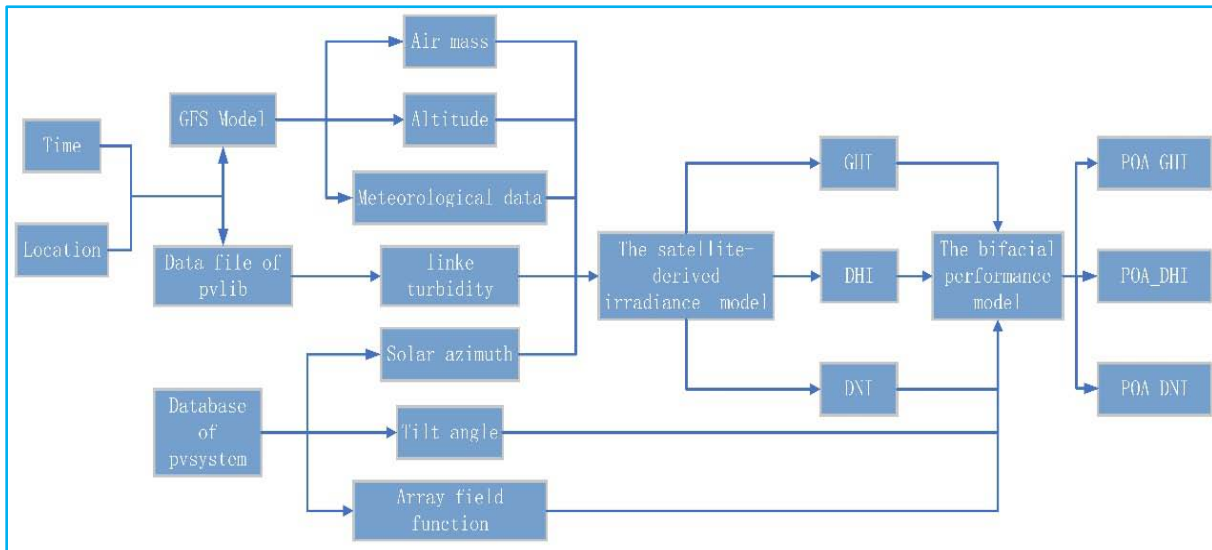


Fig. 2. Function flowchart for a procedure

The physical method uses meteorological data and physical correlations to predict solar irradiance (Lai *et al.*, 2019). The 5-parameter model is a simulation physical model used to fit the field test I-V curves of the complete PV module, resulting in more consistent prediction data with the measured values (Ma *et al.*, 2014). A behavioral electrical power model is used to compute the electrical power (Razagui *et al.*, 2020).

To consider the time lag power effect with high flexibility, S. Ding developed a new gray system model that detected the trends of nonlinearity, periodicity, and volatility in each time series through inputting time-varying parameters. A new discrete grey model was created with time-varying characteristics, which can quickly find the optimal solution to improve the prediction accuracy (Ding *et al.*, 2021).

1.2.2. Interval prediction

Interval prediction aims to provide upper and lower limits of the predicted power output at a certain confidence level, offering more accurate auxiliary information to the decision-maker (Lai *et al.*, 2019).

(Zhang, 2018) proposed an interval prediction method based on kernel density distribution estimation, which considers the nonlinear characteristics of PV power. (Sperati *et al.*, 2016) analyzed the sharpness by plotting the relative frequencies of the probability intervals. Additionally, (Ni *et al.*, 2017) combined lower-upper interval estimation with machine-learning to construct prediction intervals.

1.2.3. Probabilistic prediction

Probabilistic prediction can obtain the estimation of the expected value of power for a future period, as well as information about the probability distribution of the forecast results or errors. (Lai *et al.*, 2019).

A copula-based Bayesian strategy was proposed by Hossein Panamtash (Panamtash *et al.*, 2020). It captured the joint distribution between solar power and ambient temperature. (Schinke-Nendza *et al.*, 2021) proposed an integrated approach that employed D-vine copula to generate probabilistic PV power forecasts and considered the spatial interrelationship of prediction errors. (Bessa *et al.*, 2015) proposed a vector autoregressive framework for probabilistic forecasts of residential PV and secondary substation levels.

In recent years, the use of pvlib-python has become increasingly popular and many scholars have used it to conduct valuable research with promising results. We conducted a literature review of studies that employed pvlib-python in the past five years and summarized their findings as shown in Table 1.

In this paper, we employ the model from pvlib-python to achieve point predictions of direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), and global horizontal irradiance (GHI) from the plane-of-array (POA) irradiance. The predicted values are POA_DNI, POA_GHI and POA_DHI, which are calculated by a bifacial performance model based on common data types of DNI, GHI and DHI (Fig. 2).

2. Methodology

2.1. Pvlib-python

Pvlib is a comprehensive set of performance tools designed to simulate PV energy systems. Created by Sandia National Laboratories, it provides users with a set of well-documented functions for PV scenario simulation. The tool is available in two versions: pvlib-python and pvlib for MATLAB, both of which are open-source software projects (Github, 2022).

Specifically, pvlib-python is a powerful computational library written in Python. It includes a wide range of functions, including time zone and area data processing, irradiance calculation and analysis, atmosphere functions, translation functions, and PV system functions (Ineichen and Perez, 2002). In addition, it provides users with a collection of classes that simplify data management and facilitate the implementation of PV system models using well-tested procedural code. These features make it easy to calculate performance metrics and conduct data analysis. In this study, we use pvlib-python to implement solar irradiance predictions.

2.2. Acquisition and processing of input data

2.2.1. Weather forecast data of the GFS model

The geographic location and time range can be set in the program. We chose 7-day prediction cycle and use Unidata's Siphon library to access the Unidata THREDDS catalog. The catalog includes Global Forecast System (GFS), North American Model (NAM), Rapid Refresh (RAP), High-Resolution Rapid Refresh (HRRR) and National Numerical Forecast Database (NDFD). We used the GFS model from THREDDS to get the forecast meteorological data (Ineichen and Perez, 2002).

2.2.2. Data file of pvlib

Pvlib provides a datafile named "linketurbidities H5". The file contains a data matrix that provides link turbidity values worldwide. When the month and the position are given, the required link turbidity can be extracted to simulate the photovoltaic model.

2.2.3. Database reference of PV system

The PVsystem can represent a photovoltaic system with a single module array or with multiple arrays. For a PV system with a single array, the corresponding function is usually used to directly specify the attribute parameters. For a PV system with multiple arrays, parameters are specified for each array by passing different information

(these property parameters usually include surface tilt angle, the azimuth angle of the module surface, and the parameter set of the battery temperature model, etc.). Many functions can be wrapped in a function group to simplify the application of functions. The operation of the PV system requires various array and module attribute values. The imported data usually include the following four categories: CEC module database, Sandia Module database, CEC Inverter database, and Anton Driesse Inverter database. The data is imported for a full implementation of the PVsystem model. The parameters of modules and inverters are generally provided by the SAM website (Github, 2022). But if the required module can't be found, the parameters need to be imported manually.

2.3. Transformation of data

Pvlib provides PV scenario model for prediction. Perez's operational model is used to calculate and process the extracted data, combining solar geometric information, time zone, region and other data to obtain irradiance data such as DHI, DNI and GHI. However, there are substantial discrepancies in format between the data obtained from the GFS model and the input data of Perez's operational model (*e.g.*, the column names are non-standard, the temperature is in Kelvin, etc.). So, it is necessary to convert them to the corresponding format. The 'Process_data' function is used to convert the retrieved standardized meteorological data into the format that is required in the pvlib model.

2.4. Input-to-Prediction conversation

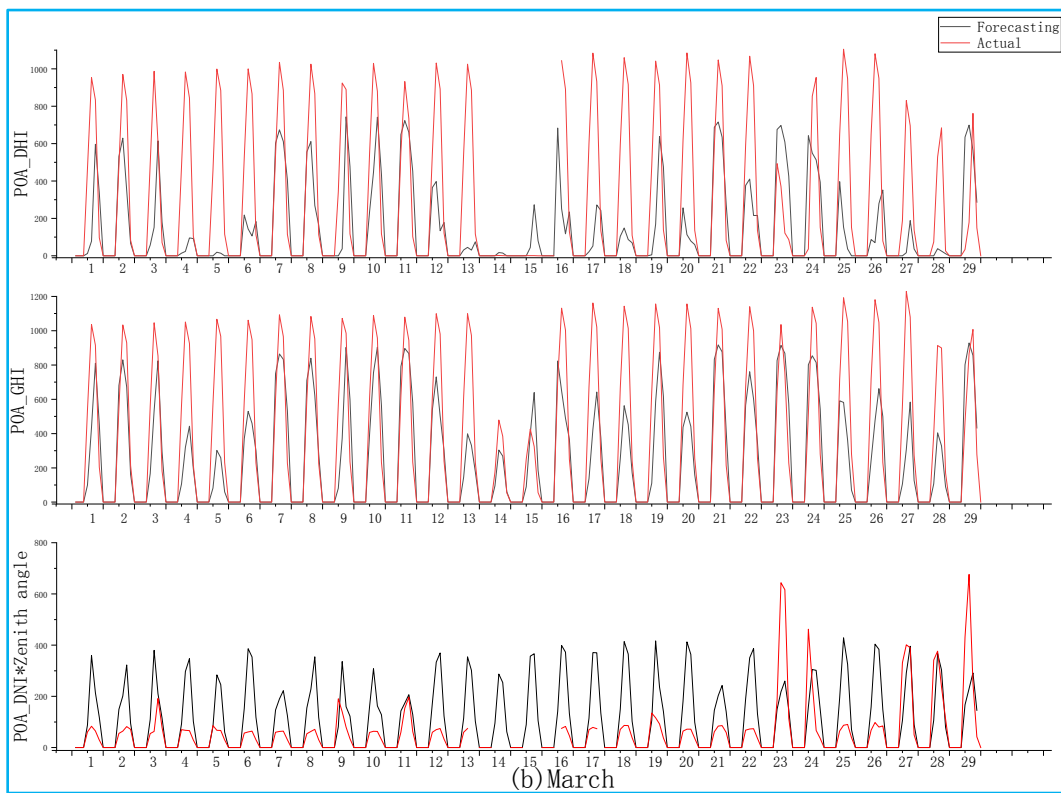
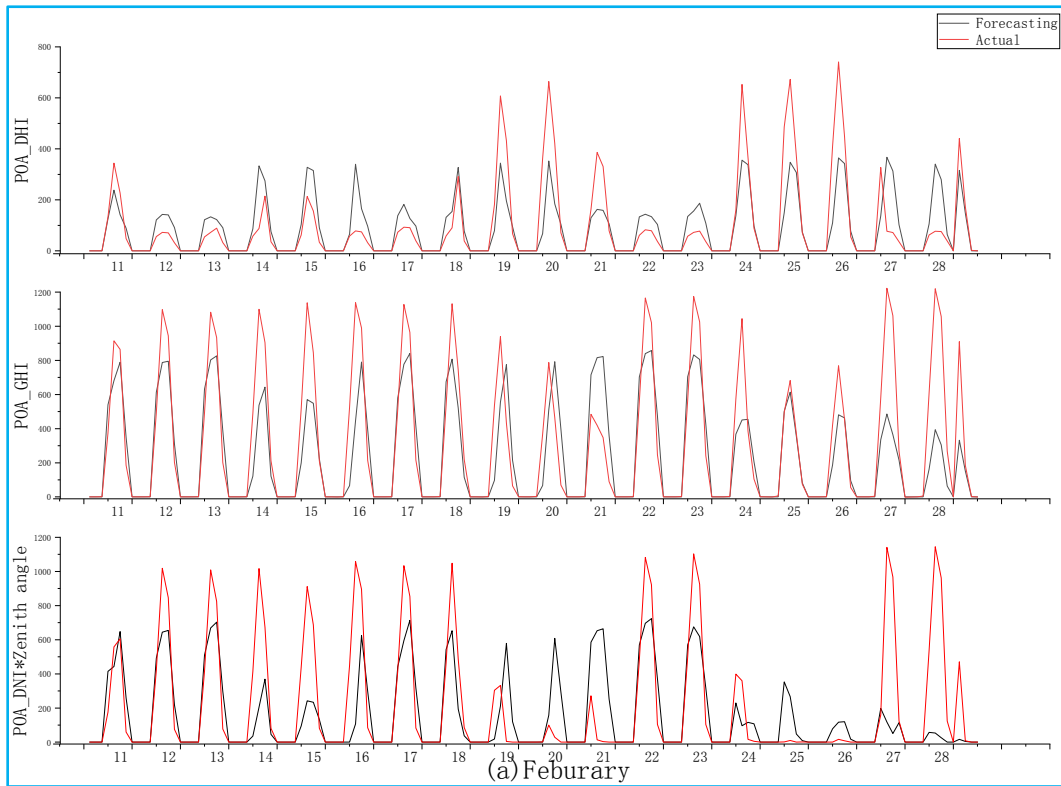
2.4.1. The satellite-derived irradiances model

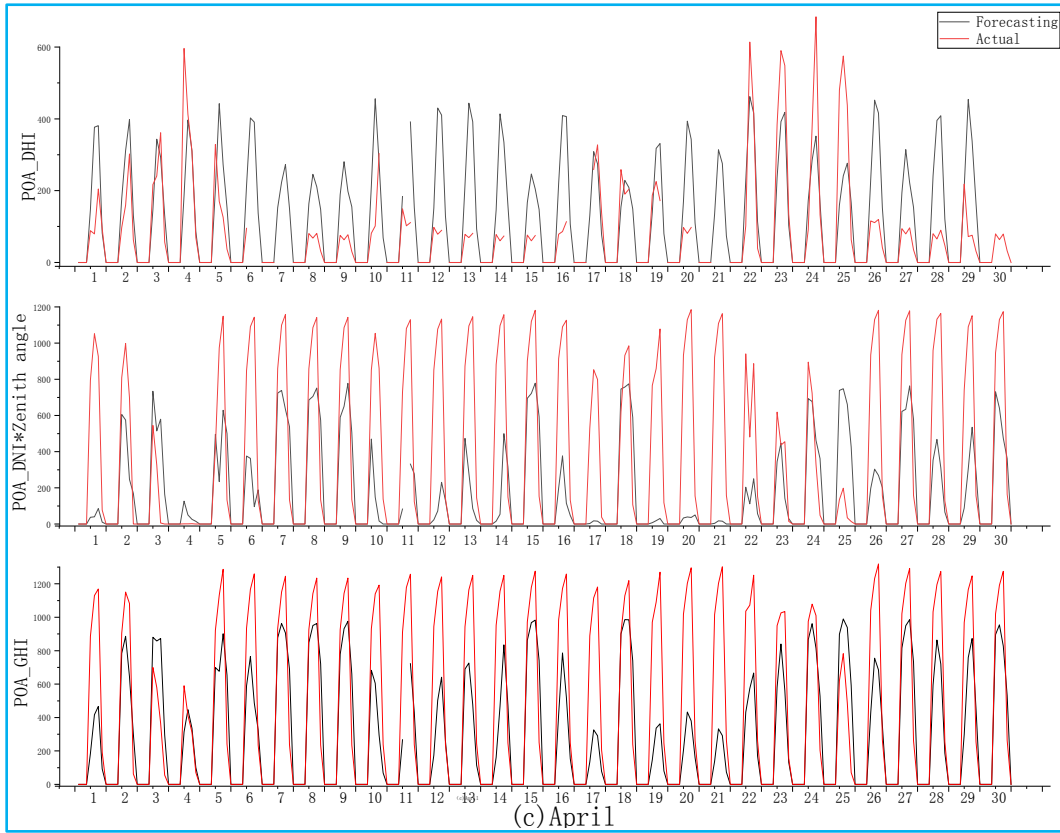
The suspended solid and liquid particles in the atmosphere can affect solar irradiance on a cloudless day. Ineichen and R. Perez discovered a link turbidity factor T_L by the inversion of the beam clear sky irradiance model (Ineichen and Perez, 2002). The Ineichen model is suitable for low solar altitudes and T_L values below 2, making it independent of solar geometry. We obtain the turbidity value from the data file "linketurbidities. H5" provided by pvlib. With the extracted T_L , air mass (AM), cloud index (CI), altitude (ALT), array module data and other meteorological data, the irradiance GHI, DHI and DNI are calculated by Perez's operational model (Perez *et al.*, 2002) :

$$GHI = K_{tm} \cdot G_{hc} (0.0001 K_{tm} \cdot G_{hc} + 0.9) \quad (1)$$

$$DNI = B_c \cdot DNI_1 \cdot (DNI_2)^{-1} \quad (2)$$

$$DNI = GHI - DNI \cdot \cos(Z) \quad (3)$$





Figs. 3(a-c). Forecasting and actual value on IZA in (a) February, (b) March and (c) April

Where G_{hc} is the clear sky global irradiance, B_c is the beam clear sky irradiance, Z is the solar zenith angle, DNI_1 is the intermediate direct irradiance, DNI_2 is the intermediate clear sky direct irradiance and

$$G_{kc} = cg_1 \cdot I_0 \cdot \cos(Z) \cdot e^{-cg_2 \cdot AM} [fh_1 + fh_2(T_L - 1)] \cdot e^{0.01AM^{1.8}} \quad (4)$$

where :

$$fh_1 = e^{-ALT/8000} \quad (5)$$

$$fh_2 = e^{-ALT/1250} \quad (6)$$

$$cg_1 = 0.0000509ALT + 0.868 \quad (7)$$

$$cg_2 = 0.0000392ALT + 0.0387 \quad (8)$$

$$Ktm = 2.36CI^5 - 6.2CI^4 + 6.22CI^3 + 2.63CI^2 - 0.58CI + 1 \quad (9)$$

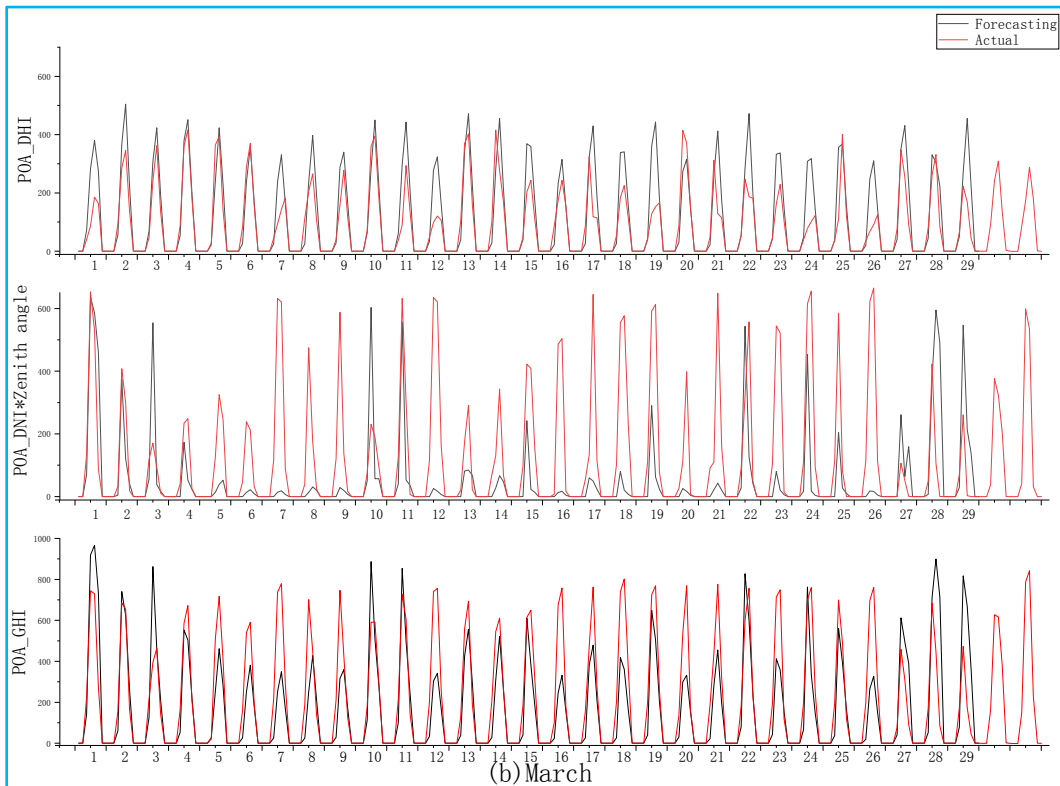
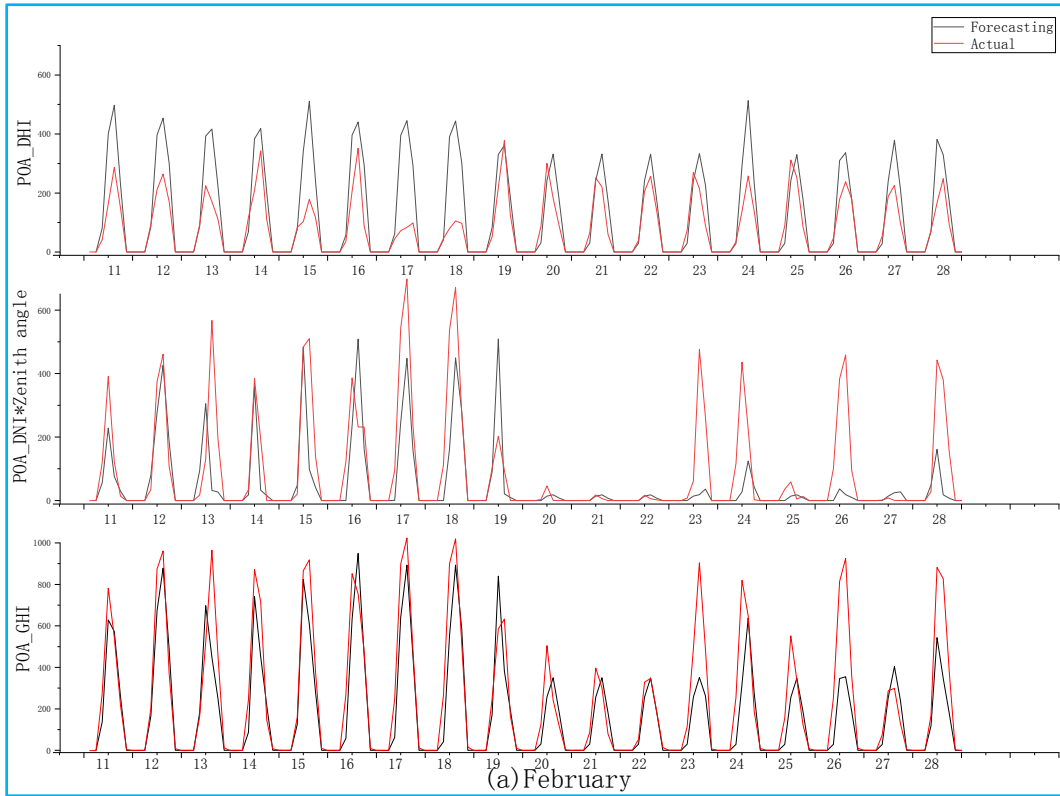
$$B_c = \min \left[0.831I_0 \cdot e^{0.09AM(T_L - 1)} \cdot (0.81 + 0.196 / fh_1), (G_{hc} - D_c) \cdot \cos^{-1}(Z) \right] \quad (10)$$

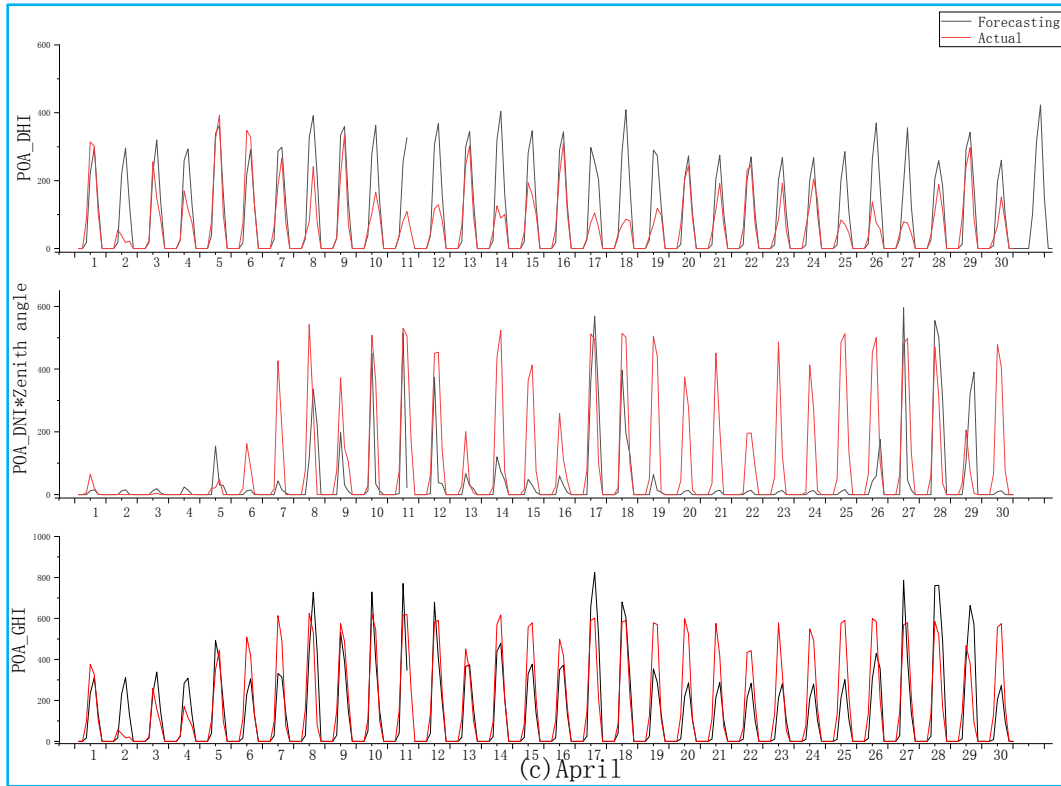
$$D_c = G_{hc} \cdot (0.1 - 0.2e^{2T_L}) \cdot (0.1 + 0.882 / fh_1)^{-1} \quad (11)$$

where I_0 is the extraterrestrial normal incident irradiance, AM is the elevation-corrected air mass, T_L is the Linke turbidity coefficient, ALT is the altitude in meters, CI is the cloud index, and D_c is the minimum clear sky diffuse irradiance.

2.4.2. The bifacial performance model

The bifacial performance model is used to collect the irradiance from both the front and back surfaces of the array. It can be used to estimate the double-sided gain of photovoltaic arrays. It assumes that the horizontal line of the photovoltaic system is uniform and infinite. The model takes into account the shadow of the photovoltaic bracket on the ground, the angle of view of the sky field on the





Figs. 4(a-c). Forecasting and actual value at RUN in (a) February, (b) March and (c) April

ground and the field of view between the module surface and the sky or ground caused by adjacent lines. At the same time, the model considers important data named ground albedo. The ground albedo is defined as the percentage of the sunlight reflected by the surface to the incident sunlight and determines the effective reflected radiation of photovoltaic energy conversion. Through the shadow function, the field of view function, ground albedo and other data of a photovoltaic array, the model can better simulate the photovoltaic gain of the double-sided array. We enter GHI, DHI and DNI into the bifacial performance model to estimate various array irradiance (the front irradiance and back irradiance of the array system): POA_GHI, POA_DHI, and POA_DNI (Mikofski *et al.*, 2019).

3. Results evaluation and discussion

3.1. Metrics to evaluate model prediction accuracy

The relative error is the ratio of the measurement's absolute error to the real value.

$$\delta = \Delta/L \tag{12}$$

where L is the actual value.

$$\Delta = |Predicted\ value - Actual\ value| \tag{13}$$

In general, the relative error is a better indicator of the degree of confidence in the measurement than the absolute error.

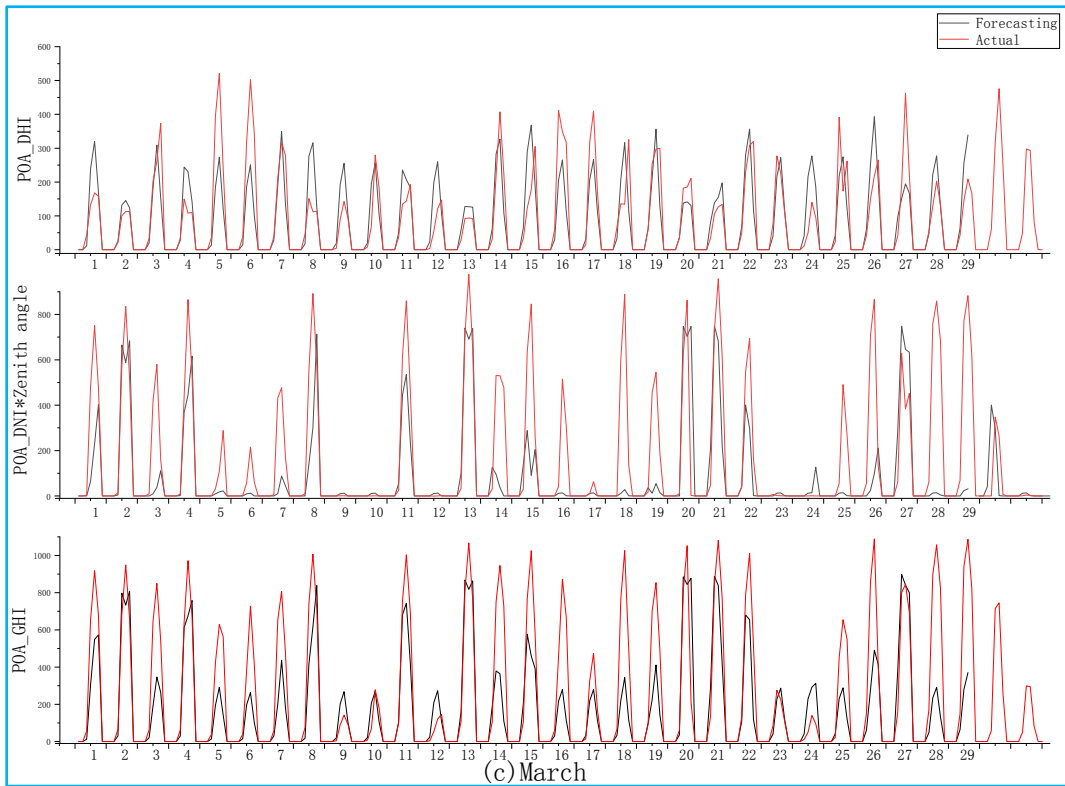
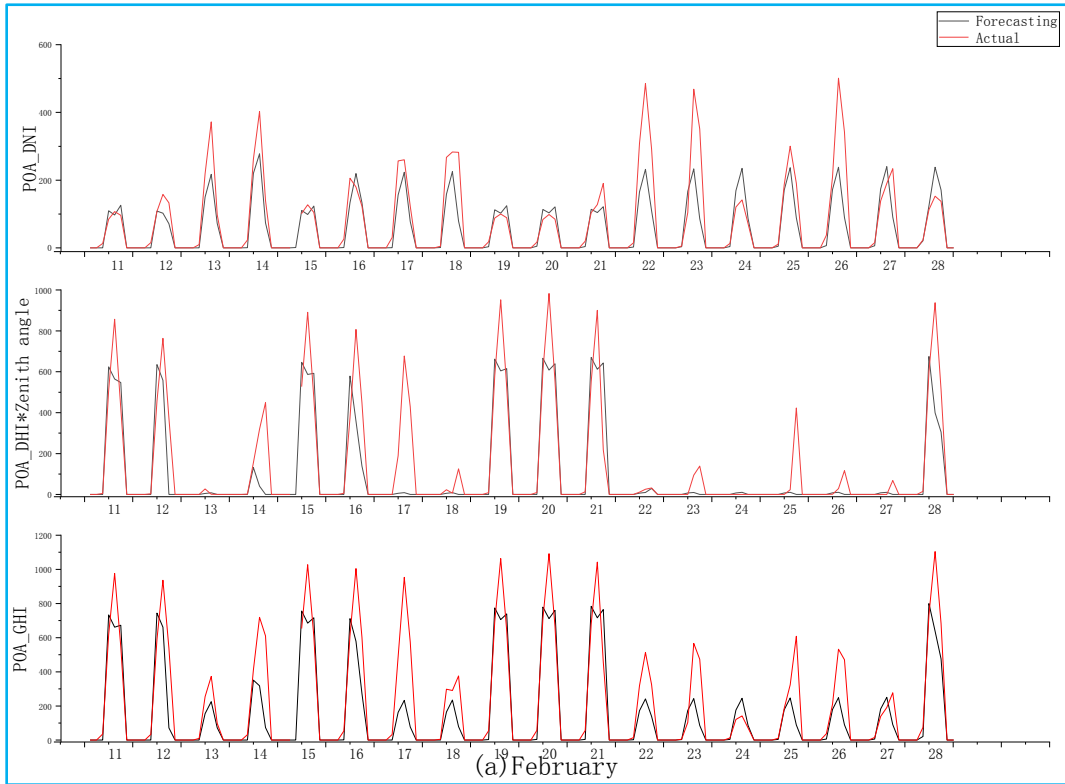
3.2. Introduction of sites

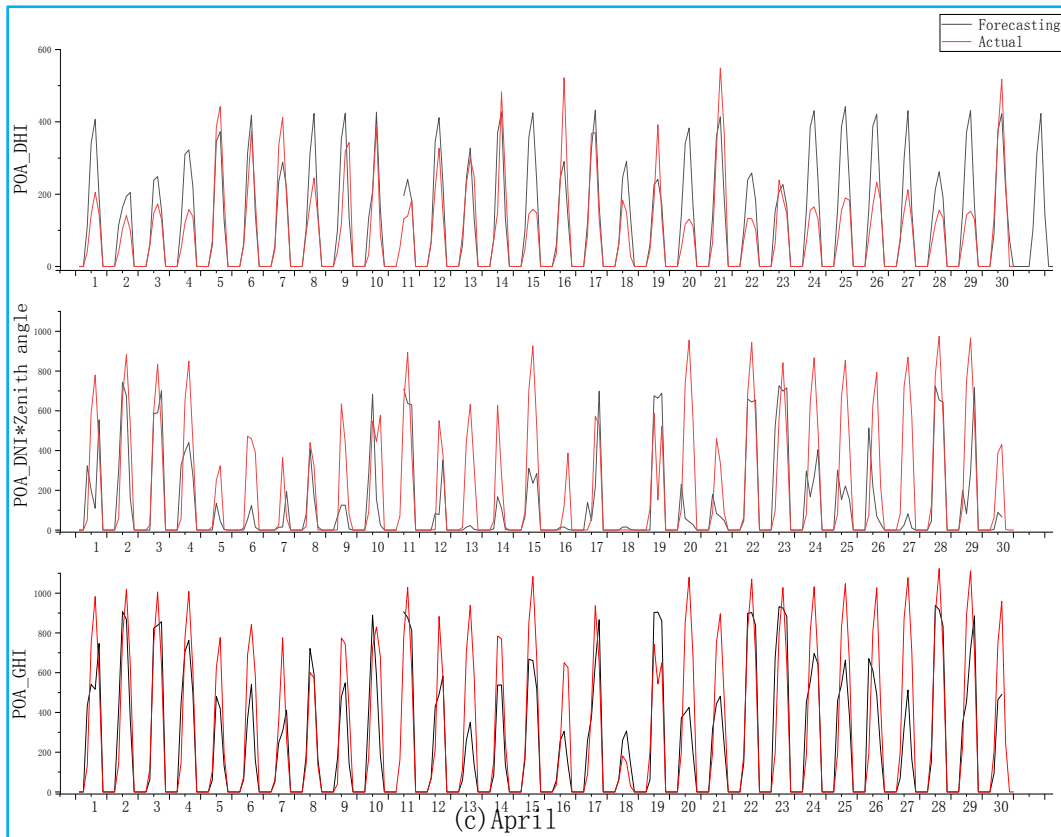
3.2.1. IZANA (IZA)

IZA is site 61 of BSRN at an altitude of 2372.9 meters in a hilltop terrain. The IZA site is located in Tenerife, Spain, at 28°18'N, 16°29'W. Tenerife is the largest island of the Canary Islands in the North Atlantic Ocean, with a population density of 305.5 people per square kilometer and a warm tropical climate type. The location has a dry climate, characterized by abundant sunshine and clean air.

3.2.2. Réunion Island, University (RUN)

The Réunion Island University site (RUN) is BSRN site 82 at 116.0 meters above sea level in a flat area of the island. The RUN site is located at 20°54'S, 55°28'E on the French island of Réunion, which has a population density of 341.9 people per square kilometer. It sits above a





Figs. 5(a-c). Forecasting and actual value at LRC in (a) February, (b) March and (c) April

crustal hotspot and has a tropical rainforest climate at the shore and an alpine climate on the island. The population density of RUN is 329.2 people per square kilometer. There is little human involvement and the seasonal variation of solar irradiance is not very large.

3.2.3. Langley Research Center (LRC)

The Langley Research Center (LRC) is BSRN site 49. It is located at a flat location at 3.0 meters above sea level. The LRC is located at 37°6'N, 76°23'W in Hampton, Virginia, USA, which has a population density of 389.5 people per square kilometer. The climate is subtropical and humid, with four distinct seasons. The population density is 389.3 people per square kilometer, there are also few human activities.

3.3. Data analysis

The time resolution of the data is 3 hours. At IZA, we analyzed 77 days' data from February to April. 15 days' average relative errors of POA_DHI were below 50%. For example, the average relative error of POA_DHI on 19 February was 37.76% and that on 23 March was 43.13%.

5 days' average relative errors of POA_DNI were below 50%. They were 35.50% on 14 February, 27.41% on 18 February, 30.95% on 2 March, 39.67% on 8 March and 44.07% on 2 April. And 36 days' average relative error of POA_GHI was below 50%. Such as the average relative error of POA_GHI on 25 February was 28.85% and that on 4 April was 26.36%. So, GHI has the best accuracy at IZA.

The average relative errors of POA_GHI and POA_DNI were both below 50% on 14 February, 18 February, 2 March and 8 March. On these days, the parameters of mid-high and mid-low clouds were under 5%. Especially, there were no mid-low clouds on 18 February. On 25 February, 26 February, 14 March, 15 March, 3 April, 4 April and 25 April, the average errors of the POA_DNI were over 1000%. And the parameter of mid-low clouds was over 10% [Figs. 3(a-c)].

At RUN, we analyzed 77 days of data from February to April. 24 days' average relative errors of POA_DHI were below 50%. Such as the average relative error of POA_DHI on 19 February was 41.25% and that on 20 March was 29.01%. Only one day's average relative error

of POA_DNI was below 50%. It is 46.06% on 18 April. And 40 days' average relative errors of POA_GHI were below 50%. Such as the average relative error of POA_GHI on 15 February was 14.49% and that on 18 April was 28.18%. So, GHI has the best accuracy at RUN.

The average relative error of POA_GHI and POA_DHI were both below 50% on 2 March, 4 March, 6 March, 10 March, 1 April, 6 April and 9 April. On these days, the parameters of mid-high clouds were under 5%. On 14, 27, 29 February, 3-5, 7 and 8 April, the average errors of the POA_DNI were over 1000%. On these days, the parameter of mid-low clouds was over 10% [Figs. 4(a-c)].

At LRC, we also analyzed 77 days of data from February to April. 38 days' average relative errors of POA_DHI were below 50%. Such as the average relative error of POA_DHI on 21 February was 35.89% and that on 22 March was 35.28%. 10 days' average relative errors of POA_DNI were below 50%. They were 46.56% on 11 February, 27.09% on 15 February, 41.59% on 19 February, 41.04% on 20 February, 36.81% on 2 March, 40.35% on 4 March and 46.17% on 2 March, 43.59% on 3 April, 20.19% on 22 April, 29.07% on 28 April. And 32 days' average relative errors of POA_GHI were below 50%. Such as the average relative error of POA_GHI on 22 March was 35.76% and that on 28 April was 18.07%. So, DNI has the best accuracy at LRC.

The average relative errors of POA_GHI and POA_DNI were both below 50% on 11, 15, 20 February and 28 April. On these days, the parameters of mid-high and mid-low clouds were under 5%. On 23, 27 February, 20 March and 9, 12, 17 April, the average errors of the POA_DNI were over 1000%. On these days, the parameter of mid-low clouds was over 10% [Figs. 5(a-c)].

4. Conclusion

In this study, we used the Ineichen model implemented in pvlib-python to estimate POA_DNI, POA_GHI, and POA_DHI under clear sky conditions at three different locations within the BSRN network. Our results show that the model provides the most accurate prediction for POA_GHI and that the accuracy of radiation prediction is greatly influenced by cloud cover, with an average relative error that is smaller when cloud cover is below 5%. The population size of the studied regions is small and has a negligible impact on the accuracy of our predictions. Among the three sites, the LRC site provide the largest amount of data with an average relative error of less than 50%.

Solar irradiance data is widely utilized for independent and financing solar resource assessment,

operational monitoring and solar forecasting. Constructing a photovoltaic power plant is a substantial investment, requiring thorough analysis and calculations during the early stages to ensure a return on investment. Accurate PV forecasting is crucial in assessing the optimal location for PV installations and reducing monetary losses due to improper placement. With pvlib, users can choose the appropriate model and algorithm to fit their needs and combine various calculation models to obtain the final results.

Moving forward, we plan to improve our predictions by incorporating atmospheric factors and applying the PV prediction method of pvlib-python to real-world solar energy resource assessments. In a basic scenario, we use pvlib to calculate the economic benefits of building a PV power plant at a known latitude and longitude. We predict irradiance and other data based on the location's meteorological forecast data, then combine them with PV module parameters and configuration information to estimate power generation.

Acknowledgments

Thanks to the BSRN website for the measurement of DNI, DHI and GHI at the RUN, IZA and LRC sites. This project is supported by the Undergraduate Innovation and Entrepreneurship Training Program of Hubei University of Technology under Grant No. X202110500134.

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

- Adel Mellit and Soteris Kalogirou, 2022, "Solar radiation and photovoltaic systems: Modeling and simulation", Academic Press. <https://doi.org/10.1016/B978-0-12-820641-6.00001-6>.
- Almonacid, F., Pérez-Higueras, P. J., Fernández, E. F. and Hontoria, L., 2014, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator", *Energy Convers. Manag.*, **85**, 389-398. <https://doi.org/10.1016/j.enconman.2014.05.090>.
- Bessa, R. J., Trindade, A., Silva, C. S. P. and Miranda, V., 2015, "Probabilistic solar power forecasting in smart grids using distributed information", *Int. J. Electr. Power Energy Syst.*, **72**, 16-23. <https://doi.org/10.1016/j.ijepes.2015.02.006>.
- Chen, S., Liang, Z., Dong, P., Guo, S. and Li, M., 2023, "A transferable turbidity estimation method for estimating clear-sky solar irradiance", *Renew. Energy*, **206**, 635-644. <https://doi.org/10.1016/j.renene.2023.02.096>.
- Deo, R. C., Ahmed, A. A. M., Casillas-Pérez, D., Pourmousavi, S. A., Segal, G., Yu, Y. and Salcedo-Sanz, S., 2023, "Cloud cover bias correction in numerical weather models for solar energy monitoring and forecasting systems with kernel ridge

- regression”, *Renew. Energy*, **203**, 113-130. <https://doi.org/10.1016/j.renene.2022.12.048>.
- Ding, M. and Xu, N., 2011, “Short-term prediction method of photovoltaic power generation system output based on Markov chain”, *Power Syst. Technol.*, **35**, 152-157. <https://doi.org/10.1016/j.pst.2011.04.031>.
- Ding, S., Li, R. and Tao, Z., 2021, “A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting”, *Energy Convers. Manag.*, **227**, 113644. <https://doi.org/10.1016/j.enconman.2020.113644>.
- Fermín Rodríguez, Martín, F., Fontán, L. and Galarza, A., 2021, “Ensemble of machine learning and spatiotemporal parameters to forecast very short-term solar irradiation to compute photovoltaic generators’ output power”, *Energy*, **229**, 120647. <https://doi.org/10.1016/j.energy.2021.120647>.
- Gagne, D. J., McGovern, A., Haupt, S. E. and Williams, J. K., 2017, “Evaluation of statistical learning configurations for gridded solar irradiance forecasting”, *Sol. Energy*, **150**, 383-393. <https://doi.org/10.1016/j.solener.2017.04.031>.
- GitHub, Github libraries, Accessed 23 January, 2022 [Online] Available: <https://github.com/NREL/SAM/tree/develop/deploy/libraries>.
- Heusinger, J., Broadbent, A. M., Sailor, D. J. and Georgescu, M., 2020, “Introduction, evaluation and application of an energy balance model for photovoltaic modules”, *Sol. Energy*, **195**, 382-395. <https://doi.org/10.1016/j.solener.2019.11.041>.
- IEA (International Energy Agency), Share of oil consumption, Accessed 20 January, 2022 [Online] Available: <https://www.iea.org/data-and-statistics/data-browser?country=WORLD&fuel=Energy%20consumption&indicator=TFChySource>.
- IEA (International Energy Agency), Electricity generation by source, Accessed 20 January, 2022 [Online] Available: <https://www.iea.org/data-and-statistics/data-browser?country=WORLD&fuel=Electricity%20and%20heat&indicator=ElecGenByFuel>.
- IEA (International Energy Agency), Renewable electricity generation by source in China, Accessed 20 January, 2022 [Online] Available: <https://www.iea.org/data-and-statistics/data-browser?country=CHINA&fuel=Electricity%20and%20heat&indicator=RenewGenBySource>.
- Ineichen, P. and Perez, R., 2002, “A new airmass independent formulation for the Linke turbidity coefficient”, *Sol. Energy*, **73**, 151-157. [https://doi.org/10.1016/S0038-092X\(02\)00045-2](https://doi.org/10.1016/S0038-092X(02)00045-2).
- Inman, R. H., Pedro, H. T. C. and Coimbra and C. F. M., 2013, “Solar forecasting methods for renewable energy integration”, *Prog. Energy Combust. Sci.*, **39**, 535-576. <https://doi.org/10.1016/j.pecs.2013.06.002>.
- Jiang, F., Wang, Z. and Zhang, P., 2019, “Grey-Weighted Markov Chain-Based Forecasting of Photovoltaic Power Generation”, *Power Syst. Prot. Control*, **47**, 55-60. <https://doi.org/10.19783/j.cnki.pspc.180634>.
- Kallio-Myers, V. and Riihelä, A., 2020, “Global horizontal irradiance forecast for Finland based on geostationary weather satellite data”, *Sol. Energy*, **198**, 68-80. <https://doi.org/10.1016/j.solener.2020.01.008>.
- Lai, C., Li, J., Chen, B., Huang, Y. and Wei, S., 2019, “Review of photovoltaic power output prediction technology research”, *Trans. China Electrotech. Soc.*, **34**, 1201-1217. <https://doi.org/10.19595/j.cnki.1000-6753.tces.180326>.
- Ma, T., Yang, H. and Lu, L., 2014, “Solar photovoltaic system modeling and performance prediction”, *Renew. Sustain. Energy Rev.*, **36**, 304-315. <https://doi.org/10.1016/j.rser.2014.04.057>.
- Mellit, A., Sağlam, S. and Kalogirou, S. A., 2013, “Artificial neural network-based model for estimating the produced power of a photovoltaic module”, *Renew. Energy*, **60**, 71-78. <https://doi.org/10.1016/j.renene.2013.04.011>.
- Mikofski, M. A., Darawali, R., Hamer, M., Neubert, A. and Newmiller, J., 2019, “Bifacial Performance Modeling in Large Arrays, in : 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)”, Presented at the 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), IEEE, Chicago, IL, USA, 1282-1287. <https://doi.org/10.1109/PVSC40753.2019.8980572>.
- Mikofski, M. M., Hansen, C. W., Holmgren, W. F. and Kimball, G. M., 2017, “Use of measured aerosol optical depth and precipitable water to model clear sky irradiance, in : 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)”, Presented at the 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC), 110-116. <https://doi.org/10.1109/PVSC.2017.8366314>.
- Mughal, S. N., Sood, Y. R. and Jarial, R. K., 2021, “Design and optimization of photovoltaic system with a week ahead power forecast using autoregressive artificial neural networks”, *Mater. Today Proc.*, **S2214785321067304**. <https://doi.org/10.1016/j.matpr.2021.10.223>.
- NEA (National Energy Administration), New energy generation by the National Energy Administration, Accessed 20 January, 2022 [Online] Available: http://www.nea.gov.cn/2021-12/31/c_1310404016.htm.
- Ni, Q., Zhuang, S., Sheng, H., Kang, G. and Xiao, J., 2017, “An ensemble prediction intervals approach for short-term PV power forecasting”, *Sol. Energy*, **155**, 1072-1083. <https://doi.org/10.1016/j.solener.2017.07.052>.
- Panamtash, H., Zhou, Q., Hong, T., Qu, Z. and Davis, K. O., 2020, “A copula-based Bayesian method for probabilistic solar power forecasting”, *Sol. Energy*, **196**, 336-345. <https://doi.org/10.1016/j.solener.2019.11.079>.
- Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R. and Vignola, F., 2002, “A new operational model for satellite-derived irradiances : description and validation”, *Sol. Energy*, **73**, 307-317. [https://doi.org/10.1016/S0038-092X\(02\)00122-6](https://doi.org/10.1016/S0038-092X(02)00122-6).
- Razagui, A., Abdeladim, K., Semaoui, S., Hadj Arab, A. and Boulahchiche, S., 2020, “Modeling the forecasted power of a photovoltaic generator using numerical weather prediction and radiative transfer models coupled with a behavioral electrical model”, *Energy Rep.*, **6**, 57-62. <https://doi.org/10.1016/j.egy.2019.08.018>.
- Schinke-Nendza, A., von Loeper, F., Osinski, P., Schaumann, P., Schmidt, V. and Weber, C., 2021, “Probabilistic forecasting of photovoltaic power supply - A hybrid approach using D-vine copulas to model spatial dependencies”, *Appl. Energy*, **304**, 117599. <https://doi.org/10.1016/j.apenergy.2021.117599>.
- Sperati, S., Alessandrini, S. and Delle Monache, L., 2016, “An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting”, *Sol. Energy*, **133**, 437-450. <https://doi.org/10.1016/j.solener.2016.04.016>.
- Theocharides, S., Makrides, G., Livera, A., Theristis, M., Kaimakis, P. and Georghiou, G. E., 2020, “Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing”, *Appl. Energy*, **268**, 115023. <https://doi.org/10.1016/j.apenergy.2020.115023>.

- Wang, H., Yi, H., Peng, J., Wang, G., Liu, Y., Jiang, H. and Liu, W., 2017, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network", *Energy Convers. Manag.*, **153**, 409-422. <https://doi.org/10.1016/j.enconman.2017.10.008>.
- Wang, W., Yang, D., Hong, T. and Kleissl, J., 2022, "An archived dataset from the ECMWF Ensemble Prediction System for probabilistic solar power forecasting", *Sol. Energy*, **248**, 64-75. <https://doi.org/10.1016/j.solener.2022.10.062>.
- Yadav, A. K. and Chandel, S. S., 2017, "Identification of relevant input variables for prediction of 1-minute time-step photovoltaic module power using Artificial Neural Network and Multiple Linear Regression Models", *Renew. Sustain. Energy Rev.*, **77**, 955-969. <https://doi.org/10.1016/j.rser.2016.12.029>.
- Yao, T., Wang, J., Wu, H., Zhang, P., Li, S., Wang, Y., Chi, X. and Shi, M., 2021, "A photovoltaic power output dataset: Multi-source photovoltaic power output dataset with Python toolkit", *Sol. Energy*, **230**, 122-130. <https://doi.org/10.1016/j.solener.2021.09.050>.
- Zhang, D., Wang, J., Lin, Y., Si, Y., Huang, C., Yang, J., Huang, B. and Li, W., 2017, "Present situation and future prospect of renewable energy in China", *Renew. Sustain. Energy Rev.*, **76**, 865-871. <https://doi.org/10.1016/j.rser.2017.03.023>.
- Zhang, Q., 2018, "Study of short-term photovoltaic power output forecasting methods (Master)", Guangdong University of Technology.

