Atmospheric Environment 286 (2022) 119219

Contents lists available at ScienceDirect

Atmospheric Environment

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

Mapping clear-sky surface solar ultraviolet radiation in China at 1 km spatial resolution using Machine Learning technique and Google Earth Engine

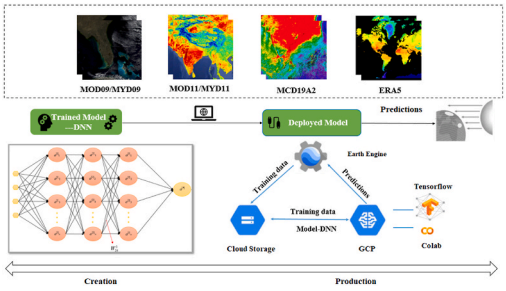
Jinyang Wu [a], Wenmin Qin [a,*], Lunche Wang [a,**], Bo Hu [b], Yan Song [a], Ming Zhang [a]

[a] Key Laboratory of Regional Ecology and Environmental Change, School of Geography and Information Engineering, China University of Geosciences, Wuhan, China
[b] Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

HIGHLIGHTS

- A Machine Learning model to estimate clear-sky UV radiation is proposed.
- This model has been evaluated in various climate zones across China.
- The model was deployed to the Google Cloud Platform for online estimation.
- The spatial and temporal variations of UV radiation in China are investigated.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Clear-sky Ultraviolet radiation
China
1 km spatial Resolution
Machine learning technique
Google earth engine

ABSTRACT

Ultraviolet (UV) radiation is an important fundamental data for solar energy utilization, climate change, human health, photochemical reaction studies, etc. However, it is still a problem to get UV radiation estimations with high spatial resolution rapidly. This study attempted to develop a Machine Learning (ML) model to estimate clear-sky UV radiation with high accuracy and high spatial resolution (1 km) in China. Based on Moderate-resolution Imaging Spectro-radiometer (MODIS) data and ERA5 reanalysis data obtained from Google Earth Engine (GEE), we established input dataset composed of different variables and developed 29 ML models to estimate clear-sky UV radiation using 37 Chinese Ecosystem Research Network (CERN) stations measurements for model training and validation. The results showed that compared with other ML models the Deep Neural Networks (DNN) model had a high and stable performance with a determination coefficient (R^2) of 0.904, a Root Mean Square Error (RMSE) of 3.100 Wm^{-2}, a Mean Absolute Error (MAE) of 2.274 Wm^{-2} for 10-fold cross-validation. To realize fast estimation of online clear-sky UV radiation, the DNN model was deployed to Google Cloud Platform. The online estimation results showed that northern China had more UV radiation than southern China, and eastern China had less radiation than western China. This study would provide a useful reference for the study of solar energy resources, human health, and ecological system studies.

* Corresponding author. Hunan Key Laboratory of remote sensing of ecological environment in Dongting Lake Area, School of Geography and Information Engineering, China University of Geosciences, Lumo road 388, Hongshan District, Wuhan, 430074, China.
** Corresponding author. Hunan Key Laboratory of Remote Sensing of Ecological Environment in Dongting Lake Area, School of Geography and Information Engineering, China University of Geosciences, Lumo road 388, Hongshan District, Wuhan, 430074, China.
E-mail addresses: qinwenmin@whu.edu.cn (W. Qin), wang@cug.edu.cn (L. Wang).

https://doi.org/10.1016/j.atmosenv.2022.119219
Received 12 October 2021; Received in revised form 3 June 2022; Accepted 4 June 2022
Available online 7 June 2022

1. Introduction

Ultraviolet (UV) radiation refers to the Solar Downward Surface Shortwave Radiation (DSSR) with the wavelength range of 100 nm–400 nm (Gueymard, 2004; Lucas et al., 2019). Although comprising a small fraction (approximately 8%) of the total solar radiation, UV radiation has significant influences on ecosystems, environments, human health, and the Earth's atmospheric processes (Ferrero et al., 2006; Thomas et al., 2012; Liu et al., 2017). UV radiation can inhibit plant photosynthesis by destroying leaves, which subsequently affects the balance of ecosystems (Williamson et al., 2014; Wang et al., 2015). UV radiation in the troposphere may accelerate photochemical reactions in the near-surface layer and produce more secondary pollutants (Liu et al., 2017). Moreover, UV radiation can cause direct and immediate harm to people, such as skin cancer and cataracts (Santos et al., 2011; Barnes et al., 2019). Therefore, accurate measurement of UV radiation and its spatial and temporal variability is necessary.

Not until the end of the 20th century, Farman et al. (1985) found that the significant loss of Antarctic ozone was associated with the unique height distribution of UV radiation in the polar stratosphere, and opened the prelude of UV radiation observation. In 1988, National Science Foundation and American Antarctic Program established a high-latitude UV radiation monitoring and observation network in Antarctica (Booth et al., 1994). In the 1990s, China also established the Brewer UV radiation observation system at Zhongshan Scientific Research Station in Antarctica and Waliguan Global Atmosphere Station (Bo et al., 2009). Subsequently, the Chinese Ecosystem Research Network (CERN), which provides long-term meteorological and radiation observations covering about 40 stations in China, was founded in 2004 (Qin et al., 2020). Topography affects the solar radiation reaching the Earth surface through self-shadowing or blocking by neighboring topography (Hao et al., 2021). Elevation, slope, slope direction (azimuth or aspect), and shadowing can all contribute to strong local gradients in solar radiation (Dubayah and Rich, 1995). However, these monitoring stations are too sparse and spread over flat terrain to routinely observe UV radiation. The most representative satellite sensor in the early years that could monitor UV radiation was the Total Ozone Mapping Spectrometer (TOMS) (Herman et al., 1999; Guide, 2012; Chubarova et al., 2020). Recently, some satellite sensors could estimate UV radiation, such as the Global Ozone Monitoring Experiment (GOME) (Krotkov et al., 2001; Lamy et al., 2021; Parisi et al., 2021), Ozone Monitoring Instrument (OMI) (Tanskanen et al., 2006; Parisi et al., 2021; Sun et al., 2021; Taipe et al., 2021), and TROPOspheric Monitoring Instrument (TROPOMI) (Lindfors et al., 2018; Lakkala et al., 2020; Kujanpää et al., 2021). Many satellites/sensors could also provide land and atmosphere parameter information with a finer spatial and temporal resolution for UV radiation estimation, such as Advanced Very High Resolution Radiometer (AVHRR) (Yang et al., 2018; Wei et al., 2019), Moderate-resolution Imaging Spectro-radiometer (MODIS) (Liang et al., 2006; Tang et al., 2016, 2017; Brown et al., 2020), Landsat 5 and 7 (Wang et al., 2019), Multi-functional Transport Satellites (MTSAT) (Tang et al., 2016; Jiang et al., 2019, 2020), Communication, Ocean, and Meteorological Satellite (COMS) Meteorological Imager (MI) (Yeom et al., 2019), Geostationary Operational Environmental Satellite-16 (GOES-16) (Hrisko et al., 2021), Himawari-8 (H-8) (Letu et al., 2020a,b; Ma et al., 2020), Fengyun-4 (FY-4) (Min et al., 2017; Yang et al., 2017), etc.

Multiple algorithms have been developed to derive UV radiation from satellite products. Empirical models are based on the mathematical relationship between satellite observation data and ground-based observation data to simplify the complex radiation transmission process. Laguarda and Abal (2019) used air quality, clearness index, and satellite-derived (OMI/TOMS) ozone column as predictors to estimate UV radiation. The model was partially adjusted at different stations. Pei and He (2019) selected MODIS data to establish two models (with or without ozone variable) based on the relationship with the solar zenith angle, clearness index, ozone, and altitude to estimate UVB radiation. Singh et al. (2022) selected Multi-angle Imaging Spectro-Radiometer (MISR) data to discuss the aerosol optical depth and its impact on the flux of UV radiation. Physical models could effectively estimate UV radiation owing to take into account the main radiation dumping and attenuation processes on UV radiation between the atmosphere and the earth's surface. For example, Verdebout (2000) proposed a look-up table method to estimate UV radiation across Europe with a spatial resolution of 0.05° selecting GOME and Meteosat data. The estimation results showed a good agreement with Ispra's measurements. Using satellite records from Solar Backscatter Ultraviolet Radiation 2 (SBUV2), OMI-DOAS and Microwave Limb Sounder (MLS) satellites sensors, Lamy et al. (2018) developed a local parameterization for retrieving clear-sky UV radiation. Based on the Leckner model, Qin et al. (2020) developed a new effective physical broadband parameterization method (hereafter, FASTUV) to estimate UV radiation under all-weather conditions. Machine Learning (ML) models focus on the flexible analysis of training data and selection of training methods, and through multi-level learning of different input data to obtain the most optimized results (Hänsch et al., 2018; Letu et al., 2020a,b). For example, Barbero et al. (2006) used two different methodologies, the traditional statistical techniques and artificial neural network methods for estimating daily solar UV radiation. Results showed that the neural network based model provides the best overall estimates in the site. Zhao and He (2022) proposed a deep learning framework combining random forests for retrieving all-sky, kilometer-level erythemal UV-B from MODIS data. The model evaluated on 2017 SURFRAD data showed an R^2 of 0.9376, an MBE of 1.24 mW/m^2, and an RMSE of 17.45 mW/m^2. Among these algorithms, ML, a subset of artificial intelligence, has the ability to automatically learn and deal with nonlinear complicated problems, showing superior performance. From the above, ML models have become promising methods for estimating UV radiation. In addition, when the remote sensing mechanism is not clear to construct an effective physical model, only the non-linear statistical method of the ML model can be used to approximate the estimation. When there is a clear physical model, but the accuracy of it is low due to insufficient remote sensing observation information or insufficient accuracy, the use of ML models could help to achieve accuracy improvement.

Google Earth Engine (GEE) is a platform that provides cloud-based high-resolution, large-scale spatial processing for inversion and analysis of geographic environmental components (Yang et al., 2019). Therefore, this study attempted to use the ML technique and GEE platform to achieve large-scale UV radiation estimation. Firstly, we used GEE to obtain the MODIS data and ERA5 reanalysis data required for clear-sky UV radiation retrieval. Then, 29 ML models for estimating UV radiation in clear sky conditions were developed and the accuracy was evaluated using UV radiation measurements at 37 CERN stations. Finally, fast inversion of online clear-sky UV radiation was realized on GEE. This method could be applied to estimate clear-sky UV radiation in regional and global scales with a high spatial resolution (1 km), fast speed, and acceptable accuracy, which would be helpful for the research of solar energy resources, human health, and ecosystems.

2. Material and method

2.1. Study area and data

2.1.1. Observation data

According to the zonal distributions of climate elements such as temperature, precipitation, and altitude, there are five climate types in China: TCZ (Temperate Continental Zone), TMZ (Temperate Monsoon Zone), MPZ (Mountain Plateau Zone), SMZ (Subtropical Monsoon Zone) and TPMZ (Tropical Monsoon Zone) (Fan et al., 2019). The hourly UV radiation (280 nm–400 nm) measurements at 37 CERN stations in China from 2004 to 2015 were used to construct a training dataset and verify clear-sky UV radiation estimation results (http://www.cern.ac.cn/). Fig. 1 shows the spatial distribution of five climatic zones and the 37

selected CERN stations. Table 1 shows the geographic information of these CERN stations. It can be seen that these stations cover most areas of China and are distributed in different climatic zones.

2.1.2. Input data

According to the theory of solar radiation transmission in the atmosphere, we take into account factors such as Rayleigh scattering, aerosol, water vapor, ozone, cloud, and other factors on UV radiation. The inputs used in this study to estimate clear-sky UV radiation include aerosol optical depth, water vapor, surface reflectance, land surface temperature, and scattering angle. These required inputs are available in the atmospheric and land products of the MODIS. The MODIS data used in this study include reflectance products (MOD09 and MYD09), land temperature products (MOD11 and MYD11), and aerosol products (MCD19A2). The MOD and MYD here represent the data collected by the Terra platform and Aqua platform, respectively (García-Mora et al., 2012). The MCD is a combination of products collected from the Terra and Aqua platform (King et al., 2003). Table 2 shows the summarized information of input variables. The surface pressure and surface solar radiation downwards from ERA5 re-analysis data were added as input variables to clear-sky UV radiation estimation models. To keep the same spatial resolution (1 km) with the MODIS products, the ERA5 data at 10 km pixel resolution was resampled to a spatial resolution of 1 km. All the above processes were realized in GEE, and specific details can be queried on the official website (https://developers.google.com/earth-engine/guides/scale; https://developers.google.com/earth-engine/guides/projections; https://developers.google.com/earth-engine/guides/resample#resampling). Both MODIS and ERA5 data are acquired from GEE (https://developers.google.com/earth-engine/datasets/catalog/). The process of data acquisition from GEE is shown in the following Fig. 2. Detail information about MODIS and ERA5 data could be found on their official websites (https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/,https://cds.climate.copernicus.eu), respectively. Besides, it is necessary to consider the influence of geographical factors such as latitude, longitude, and elevation on clear-sky UV radiation estimation.

2.2. Method

After collecting the input data, the data are randomly divided into 7:3: 70% for model training and the remaining 30% for model verification. When constructing the models, the Grid Search is used to determine the optimal parameters of each model (a parameter adjustment method, in all candidate parameter selection, try every possibility through cyclic traversal, and the best parameter is the final result).

Table 1
Summarized information of the 37 CERN stations.

Station code	Station name	Latitude	Longitude	Climate zone
SJ	Sanjiang	47.58°N	133.51°E	TMZ
DH	Donghu	30.62°N	114.35°E	SMZ
LZ	Linze	39.33°N	100.12°E	TCZ
HT	Huitong	26.85°N	109.60°E	SMZ
NMGCY	Neimenggucaoyuan	43.63°N	116.70°E	TCZ
BJSL	Beijingsenlin	39.97°N	115.43°E	TMZ
QYZ	Qianyanzhou	26.75°N	115.07°E	SMZ
ALS	Ailaoshan	24.53°N	101.02°E	TPMZ
DYW	Dayawan	22.53°N	114.52°E	TPMZ
TH	Taihu	31.42°N	120.22°E	SMZ
NM	Naiman	42.93°N	120.70°E	TCZ
AS	Ansai	36.86°N	109.32°E	TMZ
FQ	Fengqiu	35.00°N	114.40°E	TMZ
CS	Changshu	31.53°N	120.68°E	SMZ
LS	Lasa	29.67°N	91.33°E	MPZ
LC	Luancheng	37.88°N	114.68°E	TMZ
TY	Taoyuan	28.92°N	111.43°E	SMZ
SY	Shengyang	41.52°N	123.40°E	TMZ
SPT	Shapotou	37.47°N	105.00°E	TCZ
DTH	Dongtinghu	29.50°N	112.80°E	SMZ
HL	Hailun	47.45°N	116.92°E	TCZ
HB	Haibei	37.53°N	101.25°E	MPZ
HJ	Huanjiang	24.82°N	108.33°E	SMZ
YT	Yanting	31.27°N	105.45°E	SMZ
SNJ	Shennongjia	31.32°N	110.48°E	SMZ
YC	Yucheng	36.85°N	116.57°E	TMZ
MX	Maoxian	31.70°N	103.90°E	SMZ
XSBN	Xishuangbanna	21.92°N	101.27°E	TPMZ
GSSF	GonggashanF	29.58°N	101.99°E	SMZ
GSSS	GonggashanS	29.65°N	102.11°E	SMZ
EEDS	Eerduosi	39.48°N	110.18°E	TCZ
CW	Changwu	35.20°N	107.67°E	TMZ
CBS	Changbaishan	42.40°N	128.10°E	TMZ
FK	Fukang	47.29°N	87.93°E	TCZ
AKS	Akesu	40.62°N	89.82°E	TCZ
HS	Heshan	22.68°N	112.90°E	TPMZ
YT	Yingtan	28.20°N	116.92°E	SMZ

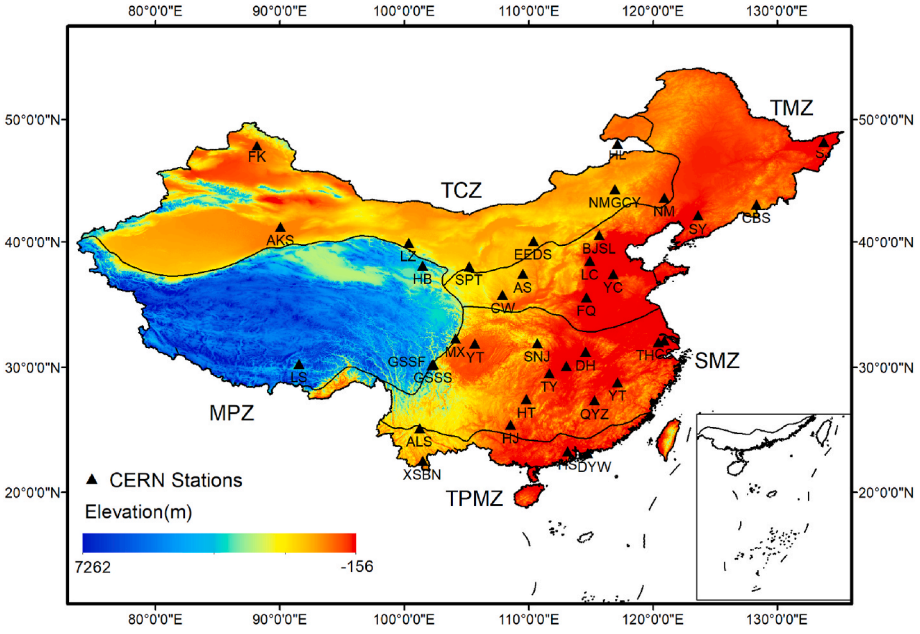


Fig. 1. Distribution of five climatic zones in China and 37 selected CERN stations.

Table 2
Summarized information about the input variables used in this study.

Dataset name	Parameters used in this study	Spatial resolution	Temporal resolution
MOD09/ MYD09	Solar Zenith Angle, Solar Azimuth Angle, Surface Reflectance at 0.47, 0.55, 0.65, 0.86, 1.24 and 2.21 μm	1 km	Instantaneous
MOD11/ MYD11	Land Surface Temperature	1 km	Instantaneous
MCD19A2	Aerosol Optical Depth (AOD) at 0.55 μm, Scattering Angle, Water Vapor	1 km	Instantaneous
ERA5	Surface Pressure, Surface Solar Radiation Downwards	10 km	Hourly

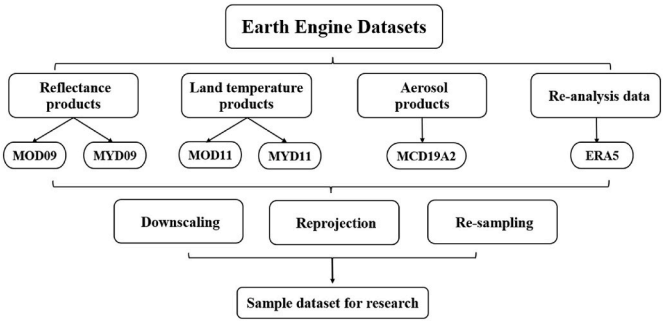


Fig. 2. The process of data acquisition from GEE.

Table 3 outlines the ML algorithm used and the ML models constructed in this study.

2.2.1. Linear regression (LR) method

In statistics, LR is a type of regression analysis that uses the least square function to establish the relationship between one or more independent and dependent variables. Its expression form is:

$$y = w'x + e \quad (1)$$

e is the error that obeys a normal distribution with a mean value of 0 (Cohen et al., 2013). LR method includes LR, ILR, MLR, SLR, and RLR.

2.2.2. Tree Method

Based on Classification and Regression Tree (CART), the algorithm of accumulating the results of all trees as the final result (i.e., using addition model) and reducing the residual error in the training process to classify or regress data is Gradient Boosting Decision Tree (GBDT), which can be described as:

$$Fm(x) = \sum_{m=1}^{M} T(x; \theta m) \quad (2)$$

The model trains M rounds in total, and each round produces a weak classifier $T(x, \theta m)$. ML created and trained four regression trees: FT, MT, CT, and OT, which are determined by the size of the smallest leaf (Clark and Pregibon, 2017).

2.2.3. Deep Neural Networks (DNN)

DNN is a kind of artificial network composed of a large number of neurons with strong universality. The extensive connection between neurons can simulate the structure and function of the nervous system. Here we built a neural network with an input layer, three hidden layers, with 128, 128, 128 nodes, an output layer, and one dropout layer. W_{jk}^{l} is the weight from the kth neuron in the (l-1)th layer to the *jth* neuron in the *lth* layer. The specific hyper-parameter settings are summarized in Table 4.

To make DNN more robust, the following technologies are used:

(1) The input variables were preprocessed by the standardization method z-score so that the variables followed the normal distribution with a mean of 0 and a variance of 1.
(2) Select Scaled Exponential Linear Units (SELU) as the activation function. It is more flexible than other activation functions and allows DNN to be highly robust (Ma et al., 2020).
(3) Add the dropout layer, it can effectively prevent the occurrence of over-fitting, and achieve the regularization effect to a certain extent.

2.2.4. Support vector machine (SVM)

The SVM algorithm, developed earlier by Vapnik, is a supervised ML

Table 4
The hyper-parameter settings of Deep Neural Networks used in this study.

Term numbers	Term names	Term properties
1	Hidden layer (Third)	128 Neuron
2	Dropout layer	30% (or 0.3)
3	Optimization function	Adam
4	Activation function	SELU
5	Learning rate	2e-4
6	Batch size	32
7	Epoch	2500

Table 3
A list of the ML algorithm and ML models used in this study.

Method	Model	Abbreviation	Method	Model	Abbreviation
Linear Regression Method	Linear Regression	LR	**Tree Method**	Classification and Regression Tree	CART
	Multiple Linear Regression	MLR		Gradient Boosting Decision Tree	GBDT
	Interaction Linear Regression	ILR		Fine Tree	FT
	Robust Linear Regression	RLR		Medium Tree	MT
	Stepwise Linear Regression	SLR		Coarse Tree	CT
				Optimized Trees	OT
Support Vector Machine	Support Vector Regression	SVR	**Ensemble Algorithm**	eXtreme Gradient Boosting	XGB
	Coarse Gaussian Support Vector Machine	CGSVM		Random Forest	RF
	Quadratic Support Vector Machine	QSVM		Bagging	Bagging
	Fine Gaussian Support Vector Machine	FGSVM		Extremely Randomized Trees	Extra Tree
	Cubic Support Vector Machine	CSVM		Adaptive Boosting	Adaboosting
	Linear Support Vector Machine	LSVM		Boosted Trees	Boosted Trees
	Medium Gaussian Support Vector Machine	MGSVM		Bagged Trees	Bagged Trees
Gaussian Process Regression	Squared Exponential Gaussian Process Regression	SEGPR		Optimized Ensemble	OE
	Mastern 5/2 Gaussian Process Regression	MGPR	**Deep Neural Networks**	Deep Neural Networks	DNN

model for data analysis and image recognition (Vapnik, 2013). The SVM estimates regression based on a series of kernel functions, which can transform the low-dimensional input data into high-dimensional feature space, to maximize the distance between points and surface (Fan et al., 2019). SVM algorithm includes SVR, CGSVM, QSVM, FGSVM, CSVM, LSVM and MGSVM.

2.2.5. Gaussian Process regression (GPR)

GPR is a nonparametric model, it gives more uncertainty to the model through Gaussian Process (GP), which is also its advantage. Although GP is complex, it is simple to calculate when used in the regression. The nonparametric model allows it to model arbitrary functions, and the kernel function also allows this model to use the internal structure of data.

For the selection of kernel function, one is Radial Basis Function (RBF), and the general form is:

$$k(Xi, Xj) = \exp\left(-\frac{1}{2\theta^2}\|Xi - Xj\|^2\right) \tag{3}$$

The other is Mastern kernel, the general form is:

$$k(Xi, Xj) = \frac{1}{2\varsigma - {}^1\Gamma(\varsigma)}\left(2\sqrt{\varsigma}\|Xi - Xj\|\right)^{\varsigma} H\varsigma\left(2\sqrt{\varsigma}\|Xi - Xj\|\right) \tag{4}$$

ζ Called smoothing coefficient, r and H are Gamma function and Bassel function. GPR is divided into SEGPR and MGPR.

2.2.6. Ensemble algorithm

Ensemble Algorithm combines several weak learners to get a better and more comprehensive model, to achieve the effect of reducing variance (Bagging), Boosting or improving forecasting. The bagging algorithm is a sampling method with a return. Boosting algorithm is a lifting algorithm, which pays more attention to the error samples in the previous model. Ensemble Algorithm includes XGB, RF, Bagging, Extra Tree, Adaboosting, Boosted Trees, Bagged Trees, and OE.

2.3. Google Earth Engine

GEE is a cloud platform developed by Google for geographic data monitoring and analysis. GEE provides users with free access to (1) petabytes of publicly available remote sensing images and other ready-to-use products; (2) high-speed computing facilities and ML algorithms provided by Google; and (3) Application Programming Interfaces (APIs) libraries for JavaScript and Python (Agapiou, 2017; Tamiminia et al., 2020). GCP was launched in November 2011, which provides calculating, storing, network, database, security services, and so on (Bisong, 2019).

2.4. Modeling performance criteria

The 10-fold cross-validation technique is commonly used to objectively evaluate the generalization ability of the model. The principle of cross-validation is depicted in Fig. 3. It means that the whole dataset is randomly divided into K equal subsets, one of which is rotationally used for testing, while the other K-1 subsets are used for training.

The determination Coefficient (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are commonly used to evaluate the performance of the model. Equations for these indicators are as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{N}(Qi - \overline{Q})(Ti - \overline{T})\right]^2}{\sum_{i=1}^{N}(Qi - \overline{E})^2 \sum_{i=1}^{N}(Ti - \overline{T})^2} \tag{5}$$

$$MAE = \frac{\sum_{i=1}^{N}|Qi - Ti|}{N} \tag{6}$$

Fig. 3. Schematic diagram of K-fold cross-validation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Qi - Ti)^2}{N}} \tag{7}$$

Q_i means the estimated value, T_i means the measured value, $\overline{Q}$ and $\overline{T}$ are the average of the estimated and the measured, and N represents the total number of samples. The ideal values of MAE and RMSE are zero, and R^2 is 1. Therefore, as the values of MAE and RMSE approach zero and R^2 approaches 1, the performance of the model becomes higher.

3. Result and analysis

The experimental flow of clear-sky UV radiation estimation is shown in Fig. 4, including data preprocessing, model building, model comparison, and accuracy evaluation.

3.1. Comparison of overall estimation performance

One of the important goals of training ML models is to evaluate the model's generalization performance on a new dataset. The methods used for model accuracy verification can be roughly divided into the hold-out method and K-fold cross-validation.

In this study, the total capacity of the dataset is 22,872. In the hold-out method, 70% of the data is used for training (16,010), and the remaining 30% is used for verification (6862). The comparison between the estimated and the measured clear-sky UV radiation is shown in Fig. 5. In general, the overall accuracy of the ML models is ideal, that is, the R^2 values of all the models are in the range of 0.702–0.912. Among the six types of ML algorithms, DNN performs best, followed by ensemble algorithm and GPR. The accuracy of LR and Tree methods on the test set is low. Among the 29 ML models, XGB has the highest accuracy (R^2 = 0.912, RMSE = 2.978 Wm-2, MAE = 2.167 Wm-2), followed by DNN (R^2 = 0.903, RMSE = 3.125 Wm-2, MAE = 2.292 Wm-2). As shown in Fig. 5, the scatter graph points of FGSVM are too scattered, and the model has the worst accuracy in the test set, which is related to the Gaussian kernel function in SVM. When the Kernel = *'rbf'* is used, which indicates that the deformation of the sample data is very complicated.

Fig. 6 shows the results of 10-fold cross-validation for each selected model. In general, ML models perform well in cross-validation, with R^2 values ranging from 0.700 to 0.918, RMSE values ranging from 2.890 to 5.510 Wm-2, and MAE values ranging from 2.087 to 4.066 Wm-2. Of six ML algorithms, the generalization ability of DNN, GPR, and ensemble algorithm is relatively stable with mean R^2 values of 0.904, 0.905, and 0.868, mean RMSE values of 3.100 Wm-2, 3.110 Wm-2, and 3.600 Wm-2, respectively, mean MAE values of 2.274 Wm-2, 2.266 Wm-2 and 2.655 Wm-2. In terms of algorithms performance, DNN has the deep and complex neural network structure that could operate on larger datasets and attain optimal performance. In addition, this study applied standardization method, activation function, and dropout layer to make the DNN model more robust. Ensemble algorithm is a combination of numerous weak learners to produce a better and more complete strong learner, which helps to reduce variation, bias, and enhance prediction. Both GPR and SVM use covariance functions as their cores to solve

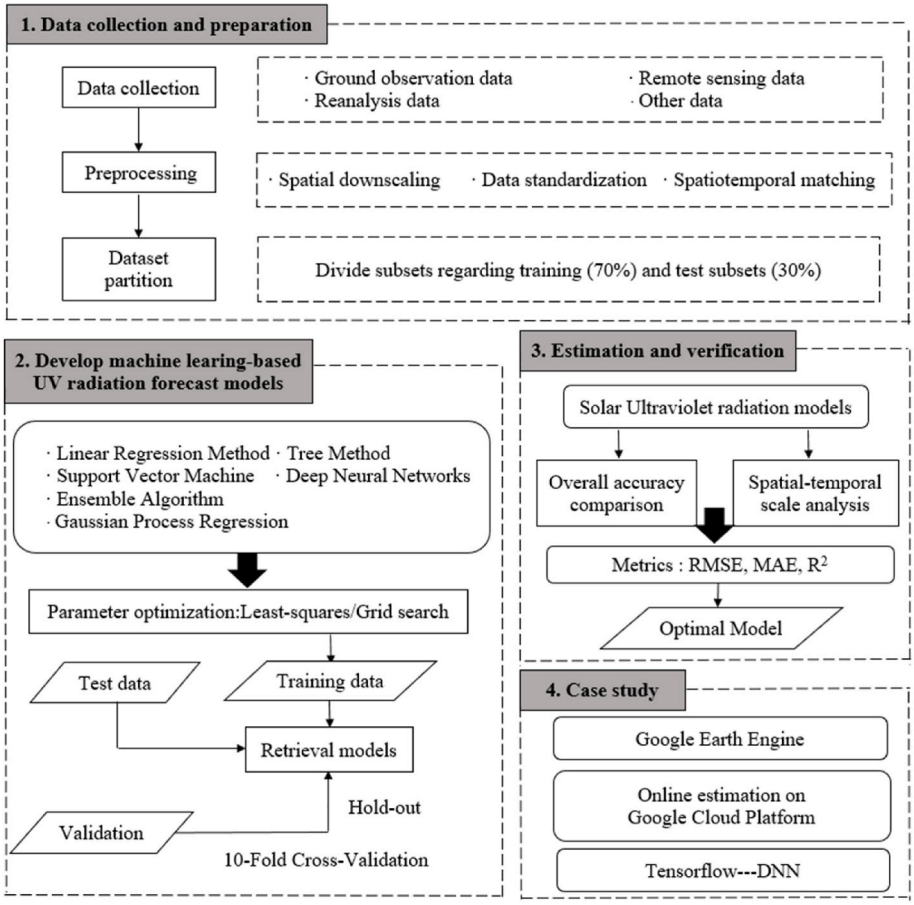


Fig. 4. The flowchart of solar clear-sky UV radiation estimation in this study.

complex regression problems with high dimensionality, small samples, and nonlinearity. However, the performance of SVM (mean $R^2 = 0.839$, mean RMSE = 3.950 Wm-2, mean MAE = 2.850 Wm-2) is lower than GPR because when SVM attempts to map sample data into a high-dimensional space using kernel functions with strong localization, which is prone to deformation of the sample data after infinite multi-dimensional transformations. The lower overall performance of LR method is due to the absence of a nonlinear relationship between the variables and the dependent variable. Among the 29 ML models, the XGB model has the highest the estimation accuracy with R^2 value of 0.918, RMSE value of 2.890 Wm-2, and MAE value of 2.087 Wm-2. The accuracy of MGPR is slightly lower than that of XGB, with R^2 value of 0.910, RMSE value of 3.060 Wm-2, and MAE value of 2.228 Wm-2. The DNN model ranked third with R^2 value of 0.904, RMSE value of 3.100 Wm-2, and MAE value of 2.274 Wm-2. FGSVM has the lowest accuracy with R^2 value of 0.700, RMSE value of 5.510 Wm-2, and MAE value of 4.066 Wm-2. The site-based cross-validation results is similar to the cross-validation results, as shown in the Appendix. From the above, we used two methods to verify the overall accuracy of the 29 ML models. To facilitate the subsequent analysis, the results of the two validation methods were integrated. The models with the best performance among all kinds of ML algorithms were selected, including ILR (LR Method)、GBDT (Tree Method)、MGSVM (SVM)、MGPR (GPR)、XGB (Ensemble Algorithm), and DNN.

3.2. Estimation accuracy analysis of the models in different seasons

Fig. 7 illustrates the validation results of the seasonal clear-sky UV radiation estimations of the six models (ILR, GBDT, MGSVM, MGPR, XGB, and DNN) with the best overall accuracy among the six types of ML algorithms obtained in 3.1. In general, the prediction results of the six models all have the same characteristics, that is, poor in spring and summer, while better in autumn and winter. According to the RMSE, R^2, and MAE values in Fig. 7, the performance and accuracy of the six models are ranked from best to worst, followed by DNN, MGSVM, MGPR, XGB, GBDT, and ILR. The prediction results of the DNN model in four seasons are the best among these six models with relatively lower RMSE, higher R, and lower MAE value. In general, the DNN model shows the most stable performance, with an R^2 value greater than 0.750 in all four seasons. As we all know, UV radiation in summer is the highest in all four seasons, but it can be seen from Fig. 7 that the prediction accuracy of the six models in summer is slightly lower, relatively. Meanwhile, the prediction results of the DNN model are also relatively good in summer. In summer, the subtropical monsoon climate causes the rain heat over the same period. The special topography of the basin leads to the accumulation of water vapor, which affects the transmission process of solar radiation. At the same time, it also provides the environmental conditions for cloud formation Thus, the prediction accuracy of the six models is affected relatively.

3.3. Validation of clear-sky UV radiation estimation in different climate zones and stations in China

Fig. 8 is the boxplot of the differences between the estimated and measured clear-sky UV radiation in different climate zones using six ML models including DNN, GBDT, ILR, MGPR, XGB, and MGSVM. For each box, the lower and upper solid lines in the box represent the 25th and 75th percentiles of the clear-sky UV radiation bias; the lower and upper whiskers are the minimum and maximum errors. The middle red line in each box represents the median value of the clear-sky UV radiation

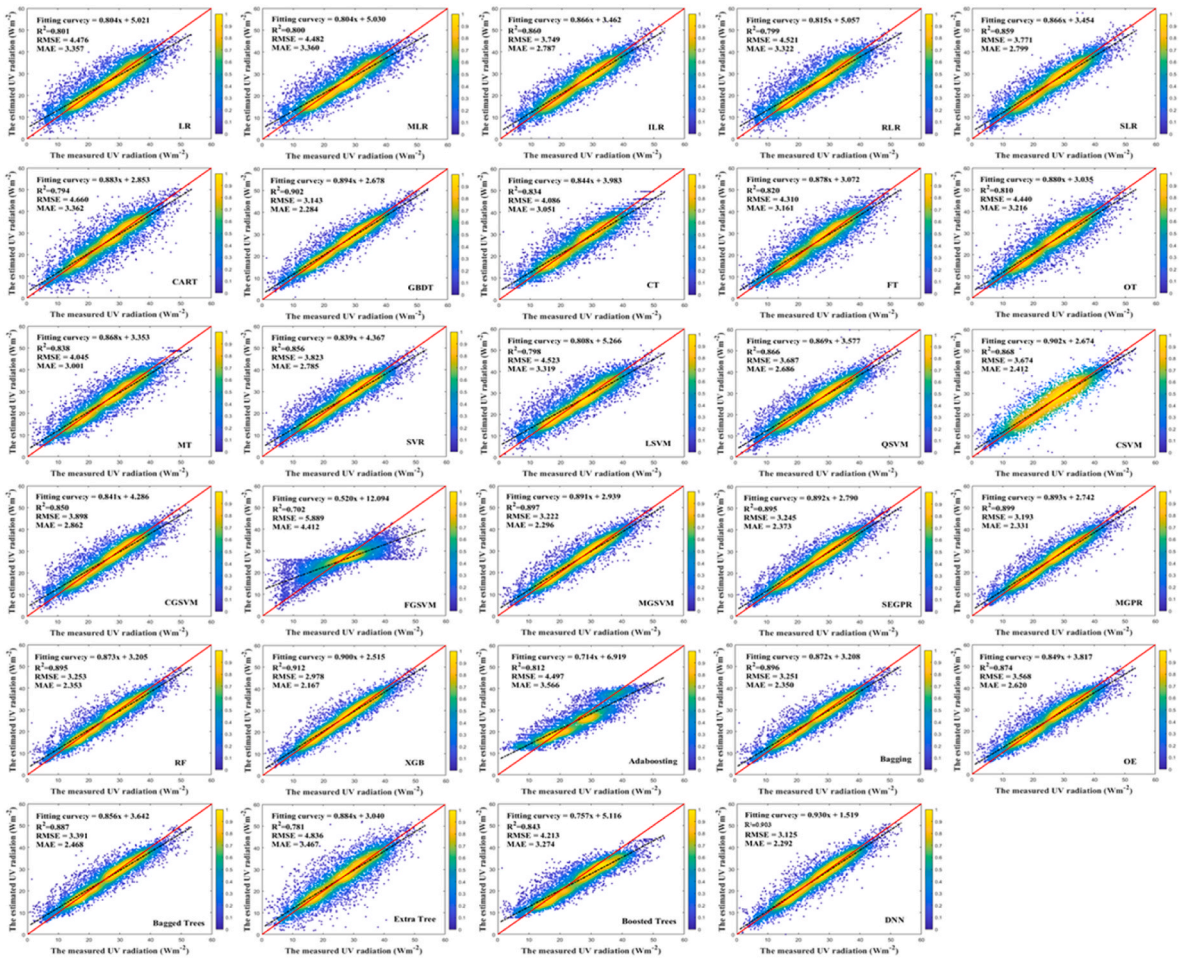


Fig. 5. Validation of the density scatterplots of clear-sky UV radiation estimation results by the hold-out method.

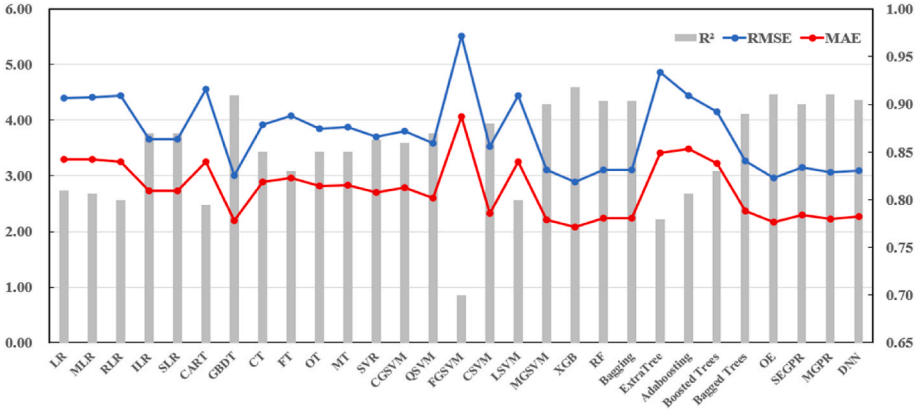


Fig. 6. The changes in the estimated indicators of each model during the 10-fold cross-validation.

biases, and the middle green triangle in each box shows the average value of the clear-sky UV radiation biases (the average values are represented on the graph). Overall, by comparing the estimation accuracy of clear-sky UV radiation of six ML models, we found that no matter in which climate zones, the estimated clear-sky UV radiation of the DNN model can maintain a good agreement with the observed clear-sky UV radiation values. Comparing the accuracy of the DNN model in five climate zones, it is found that the accuracy of the DNN model in SMZ is lower, which is related to the reason that SMZ is often overcast and rainy. The ILR model shows a large max-min error in the five climate zones, and its maximum error occurred in TPMZ.

Fig. 9 indicates the spatial distribution of these statistical indicators representing the model accuracy for each model. It can be seen that the larger RMSE and MAE value appears in SMZ and TCZ, which is similar to the results of model accuracy verification in different climate zones. The typical station with poor estimation accuracy in the TCZ area is the AKS station. There is a difference in accuracy which is caused by terrain change. The cloud distribution in the SMZ area is relatively complex and thick, which has a certain influence on clear-sky UV radiation estimation. For most stations, the estimated results of clear-sky UV radiation

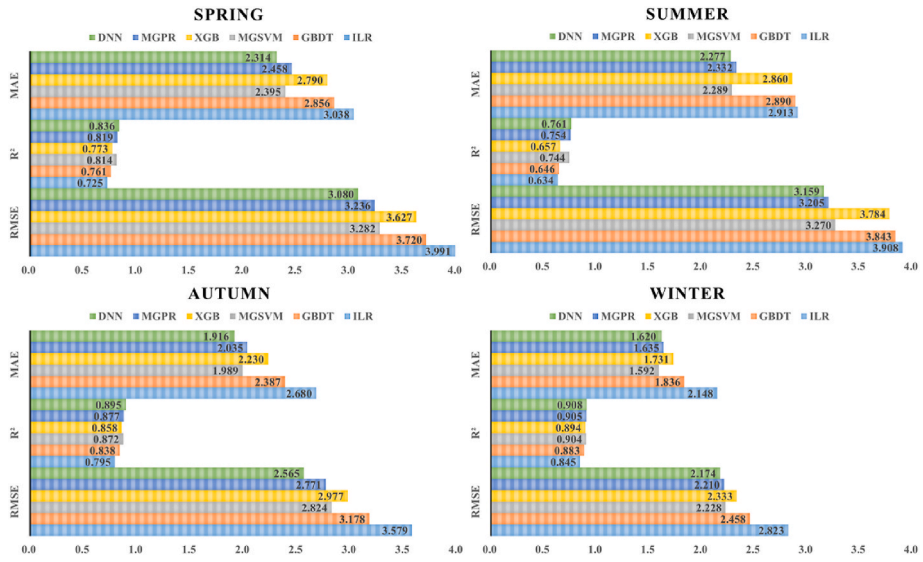


Fig. 7. Clear-sky UV radiation prediction results of different season models.

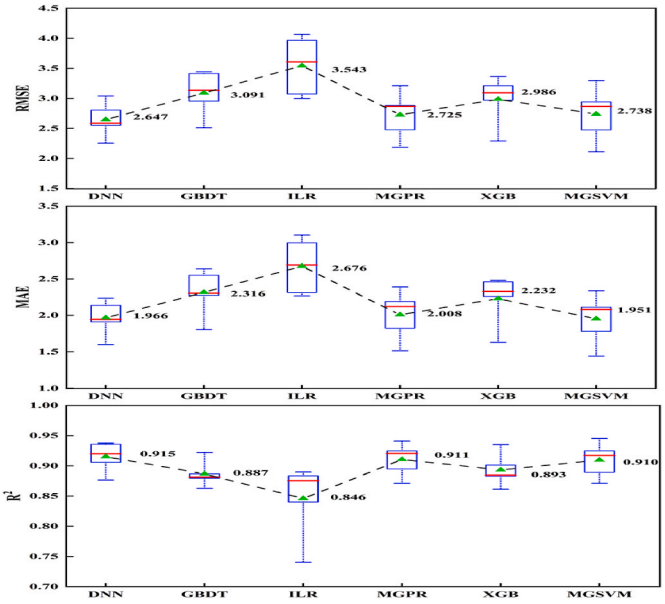


Fig. 8. Box plots analysis of the estimated and observed clear-sky UV radiation of six ML models in different climate zones.

from the DNN model are in good agreement with the measured clear-sky UV radiation. In terms of the RMSE value, in DNN, GBDT, ILR, MGPR, MGSVM and XGB models, the proportion of RMSE greater than 3.5 Wm-2 is 11.11%, 38.89%, 36.11%, 13.89%, 13.89% and 36.11%, respectively. From the perspective of R^2 value, in DNN, GBDT, ILR, MGPR, MGSVM and XGB models, the proportion of R^2 greater than 0.900 is 63.89%, 19.44%, 25%, 55.56%, 58.33% and 25%, respectively. For MAE, in DNN, GBDT, ILR, MGPR, MGSVM and XGB models, the proportion of MAE greater than 3.0 Wm-2 is 5.56%, 19.44%, 25%, 8.33%, 8.33% and 25%, respectively.

3.4. Verification of model prediction results under different solar angles

In this study, we consider the influence of solar angles such as Solar Zenith Angle (SZA), Solar Azimuth Angle (SAA), and Scattering Angle (SA) on clear-sky UV radiation estimation. Fig. 10 illustrates the prediction accuracy of each model in different SZA ranges. The range of SZA in MODIS data is 0°~180°. We selected the appropriate interval range (7°) and divided the original sample data into eight intervals with different sample numbers. In terms of the RMSE value in Fig. 10, the maximum RMSE value of the DNN model in eight SZA intervals is 3.1927 Wm-2. The maximum RMSE values of the other five models are 3.821 Wm-2 (GBDT), 3.9122 Wm-2 (XGB), 3.9855 Wm-2 (ILR), 3.2399 Wm-2 (MGPR), 3.2943 Wm-2 (MGSVM). For R value, the R^2 value of the DNN model in eight SZA intervals is 0.738–0.813. The scope of R^2 value for the other five models is 0.637–0.746 (GBDT), 0.606–0.772 (XGB), 0.495–0.679 (ILR), 0.745–0.800 (MGPR), 0.733–0.794 (MGSVM). The maximum MAE of the DNN model is 2.2619 Wm-2 in eight SZA intervals. The MAE values for the other five models are 2.9104 Wm-2 (GBDT), 2.9183 Wm-2 (XGB), 3.0034 Wm-2 (ILR), 2.3873 Wm-2 (MGPR), 2.2599 Wm-2 (MGSVM). In conclusion, the DNN model has the smallest error in different SZA intervals, which shows good prediction results. The estimation results of the DNN model are relatively stable. At the same time, there are seasonal changes in SZA in China. The estimated results of the DNN model verified by SZA are consistent with the results obtained from the evaluation on the seasonal scale in 3.2 above, that is, the DNN models have shown both good prediction results.

The range of SAA in MODIS data is −180°~180°. The original sample data is divided into five intervals: (I) −180°~-150°; (II) −150°~-115°; (III) −115°~130°; (Ⅳ) 130°~165°; (Ⅴ) 165°~180°. Table 5 shows the prediction results of each model in different SAA intervals. According to Table 5, the prediction results of the DNN model are the best among the six models in the five SAA intervals, that is, the R^2 values of the five intervals were all greater than 0.880, and the average R^2 value was 0.912. In GBDT, ILR, MGPR, MGSVM and XGB models, the proportion of R^2 greater than 0.880 is 40%, 40%, 60%, 60% and 40%, respectively. Due to the small number of samples in the first interval, the accuracy of the DNN model is affected in this case. However, MGPR and MGSVM are mostly suitable for a small number of samples, so the accuracy of the DNN model is slightly lower than that of MGPR and MGSVM in the first interval. The prediction results of the ILR model are still the worst among the six models, at the same time the MGSVM model shows similar potential to the DNN model.

Fig. 11 outlines the line chart of prediction accuracy of each model in different SA ranges. According to the maximum (178.69°) and the minimum (79.75°) of SA value in the sample data, we selected the appropriate range of SA as 10°. The fluctuation of the line chart is obvious. The prediction results of the DNN model in the five SA intervals are still the best among the six ML models, with relatively lower RMSE value, higher R^2 value, and lower MAE value. The fluctuation trend of

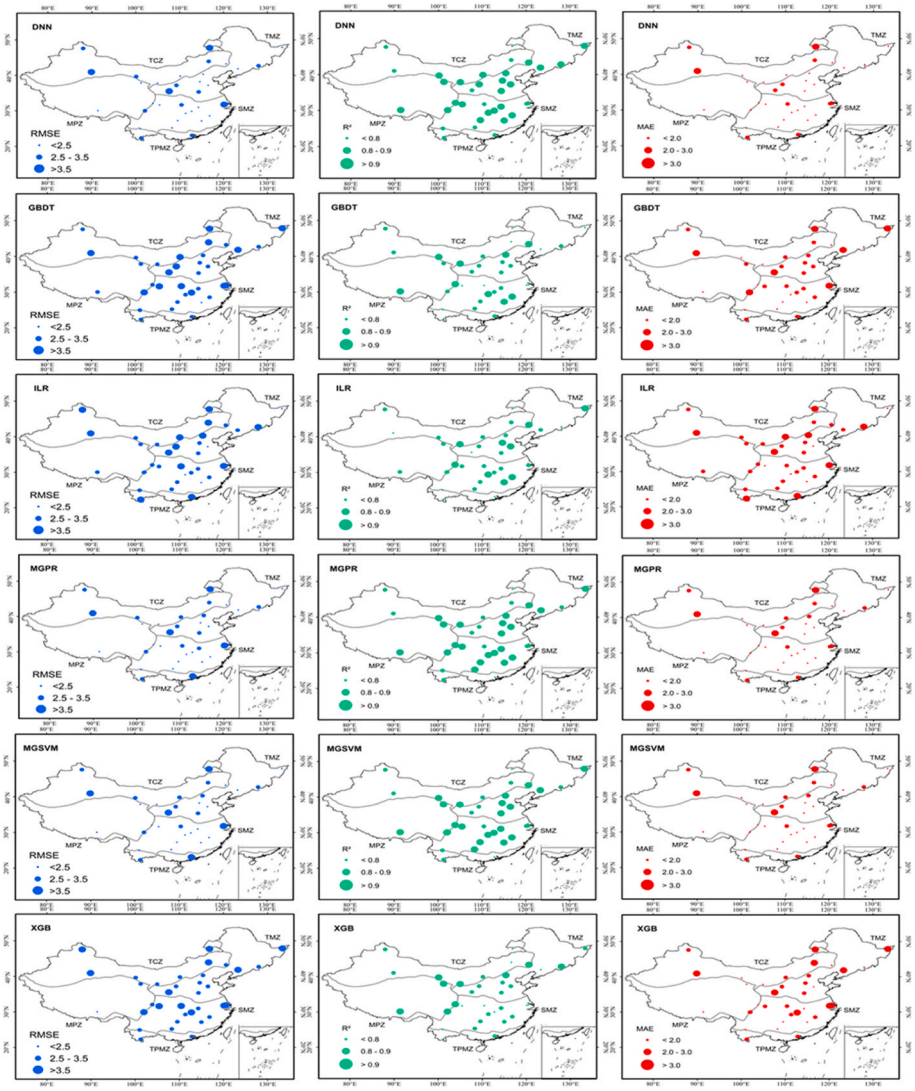


Fig. 9. Spatial distribution of six ML models estimation indicator values at different stations in China.

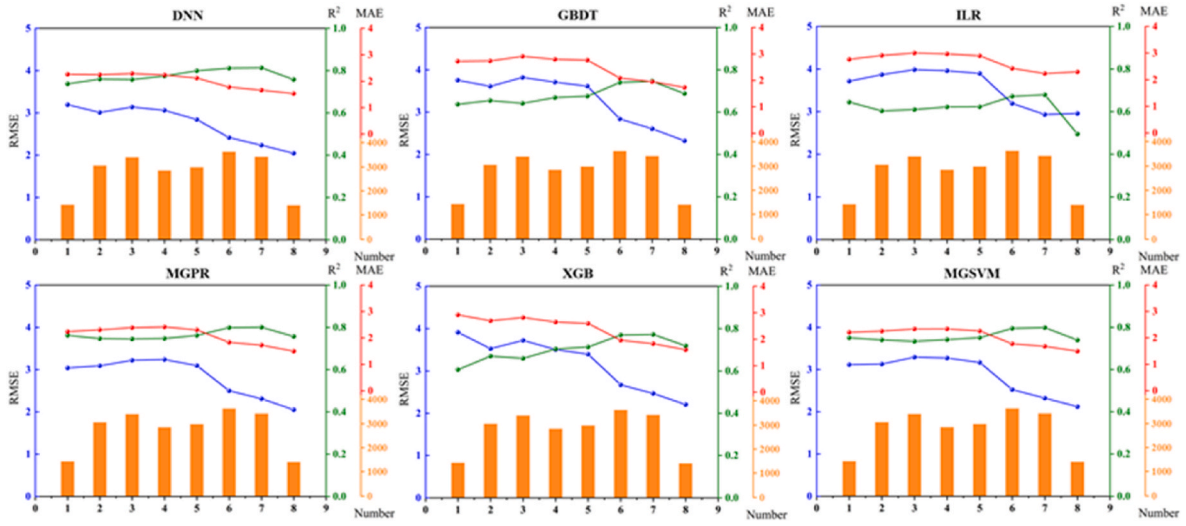


Fig. 10. Prediction accuracy of each model in different SZA ranges. (1) 0°~21°; (2) 21°~28°; (3) 28°~35°; (4) 35°~42°; (5) 42°~49°; (6) 49°~56°; (7) 56°~63°; (8) 63°~77°.

Table 5
The prediction results of each model in different SAA intervals.

Statistics indicator	Model	Range of Solar Azimuth Angle					Mean
		(I)	(II)	(III)	(IV)	(V)	
RMSE	DNN	2.679	2.962	2.992	2.920	2.913	2.893
	GBDT	3.476	3.709	3.942	3.625	3.397	3.630
	XGB	3.647	3.646	3.737	3.669	3.549	3.650
	ILR	3.307	3.645	3.532	3.869	3.999	3.670
	MGPR	2.657	2.915	3.190	3.164	2.962	2.978
	MGSVM	2.673	2.969	3.188	3.214	3.030	3.015
R^2	DNN	0.927	0.925	0.889	0.893	0.924	0.912
	GBDT	0.878	0.883	0.807	0.835	0.897	0.860
	XGB	0.865	0.887	0.827	0.831	0.888	0.859
	ILR	0.889	0.887	0.845	0.812	0.857	0.858
	MGPR	0.929	0.928	0.874	0.874	0.922	0.905
	MGSVM	0.928	0.925	0.874	0.870	0.918	0.903
MAE	DNN	1.955	2.155	2.250	2.177	2.200	2.147
	GBDT	2.586	2.804	2.923	2.825	2.583	2.744
	XGB	2.689	2.740	2.814	2.860	2.682	2.757
	ILR	2.490	2.750	2.615	2.963	2.896	2.743
	MGPR	1.946	2.164	2.341	2.373	2.199	2.205
	MGSVM	1.902	2.125	2.259	2.302	2.192	2.156

Ps: The range of SAA is −180°∼180°, select the appropriate interval range and divide the original sample data into five intervals. (I) −180°∼-150°; (II) −150°∼-115°; (III) −115°∼130°; (IV) 130°∼165°; (V) 165°∼180°.

each indicator for the DNN model in the line chart is relatively flat, which indicates the DNN model is relatively stable. The RMSE and MAE values of the ILR model are significantly higher than other models, which indicates that the ILR model has the worst estimation performance at this stage. The GBDT model also has poor prediction performance, that is, within the same input and interval, the predicted results of ILR and GBDT models are greatly different from the measured values. Different from the analysis in SZA, the prediction results of MGSVM are also considerable, ranking the second among the six selected ML models.

The above analysis provides a comprehensive review and validation of six different ML algorithms for estimating clear-sky UV radiation values. By comparing the overall accuracy of 29 ML models, the estimation accuracy of clear-sky UV radiation under different scenarios (including season, climate zones, space location, and solar angle) was comprehensively considered. It is strongly recommended that the DNN model be used to estimate clear-sky UV radiation based on astronomical, meteorological, geographical, and other factors.

3.5. Case study

Here, we deploy the DNN model, which is the optimal model that is obtained from offline training to GCP based on Tensorflow. Then we can implement online prediction through GEE calls, as shown in Fig. 12. Starting with authentication and authorization, that is, set up the software library as needed and log in to the Google account. Then we need to prepare data and define common variables, before this, we need to build engineering projects and storage buckets for storing training and test data on GCP in advance. The following is defining a model with the same structure on Google Colab as the offline DNN model and process the model for uploading to GCP. Finally, the online estimation of clear-sky UV radiation based on the GEE and DNN model can be realized.

After the deployment of the DNN model, we can obtain the clear-sky UV radiation of a certain day across China in less than 1 min, only in 10 min can obtain the full-year estimated data. Fig. 13 shows the spatial characteristics of clear-sky UV radiation in January 2018 throughout China. The spatial distribution of UV radiation was complex and extremely uneven across the country. Generally, northern China had more radiation than southern China, and eastern China had less radiation than western China. The highest UV radiation on the Qinghai-Tibet Plateau is mainly affected by altitude. The plateau areas with higher altitudes have lower air quality and relatively less scattering and absorption of UV radiation, so the intensity of UV radiation is greater than that in plain areas. The lower UV radiation in the south is mainly affected by the weather. When the southern rainy season comes, there is more cloudy and foggy weather, which greatly weakens the UV radiation.

4. Conclusion

In this study, the accuracy and applicability of 29 ML models based on six types of ML algorithms in clear-sky UV radiation estimation in China were investigated. A large-scale (national scale) ML model with a high spatial resolution (1 km) was constructed to estimate clear-sky UV radiation using MODIS and ERA5 data obtained from GEE and UV radiation measurement. The observation data from 37 CERN stations were used to verify the models.

Compared with other ML models, the DNN model was the model with the highest accuracy. The validation results showed that the R^2,

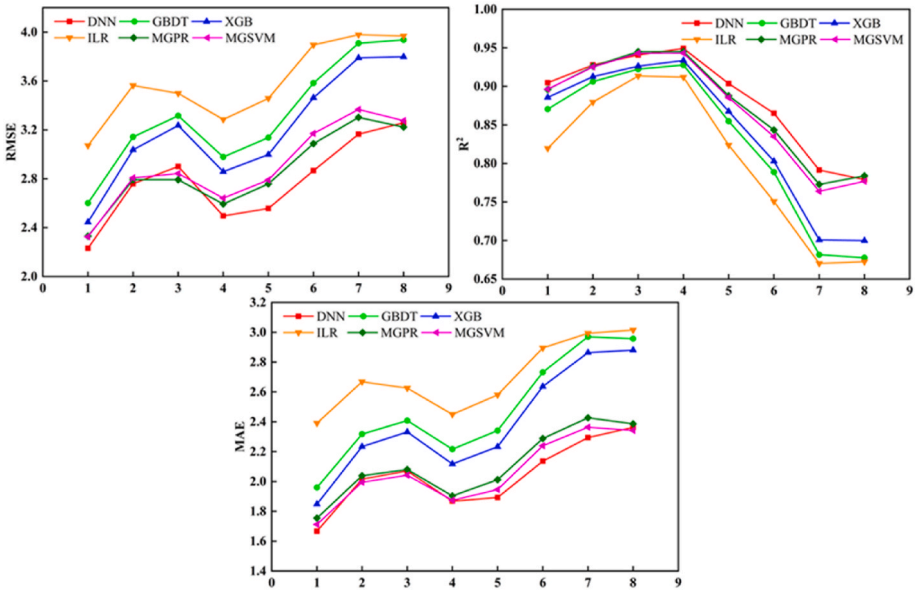


Fig. 11. Line chart of prediction accuracy of each model in different SA ranges. (1) 79.75°∼99°; (2) 99°∼109°; (3) 109°∼119°; (4) 119°∼129°; (5) 129°∼139°; (6) 139°∼149°; (7)149°∼159°; (8) 159°∼179°.

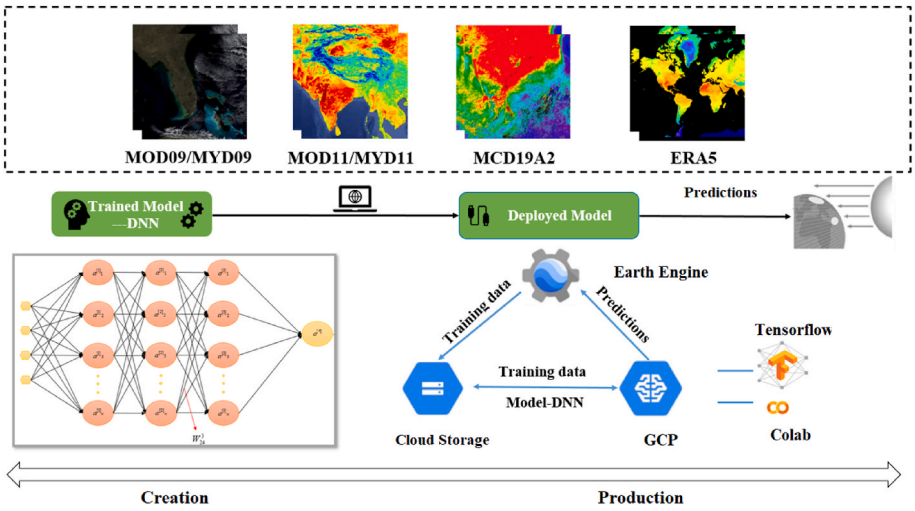


Fig. 12. Flowchart of deploying DNN model into GEE to achieve online prediction.

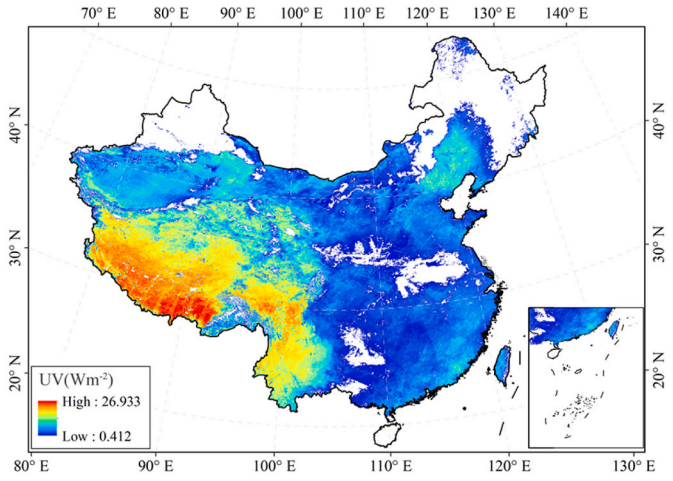


Fig. 13. The spatial distribution of clear-sky UV radiation in January 2018 over China.

RMSE, and MAE values of the DNN model in the hold-out method were 0.903, 3.125 Wm-2, and 2.292 Wm-2, respectively. And the R^2, RMSE, and MAE values of the DNN model in the cross-validation method were 0.904, 3.100 Wm-2, and 2.274 Wm-2, respectively. At the same time, the DNN model had stronger robustness and higher universality than that of other ML models under different scenarios (including season, climate zones, space location, and solar angle). The DNN model could operate on larger datasets and attain optimal performance because of its deep and complex neural network structure. In addition, this study applied standardization method, activation function, and dropout layer to make the DNN model more robust. Although the overall accuracy of XGB and GBDT models was relatively high in the sample data, their accuracy for estimating clear-sky UV radiation is unstable in different climate zones, seasons, and observation angles. Therefore, the DNN model was used to realize the fast estimation of online clear-sky UV radiation using the GEE platform. Overall, the online estimation results showed that northern China had more radiation than southern China, and eastern China had less radiation than western China. The highest UV radiation on the Qinghai-Tibet Plateau is mainly affected by altitude. The lower UV radiation in the south is mainly affected by the weather.

Compared with previous studies, this study provides an effective case of online clear-sky UV radiation estimation with high spatial resolution (1 km) and high accuracy using ML techniques and GEE platform across China. And we can obtain the clear-sky UV radiation of a certain day in less than 1 min, only in 10 min can obtain the full-year estimated data. The method can also be applied to other similar areas around the world for further study. In addition, this study didn't compare the consistencies with other existing UV radiation products and lacked validation in other regions of the world. The uncertainty caused by angular and topographic effects were not adequately considered. These works will be further carried out in the future.

CRediT authorship contribution statement

Jinyang Wu: Conceptualization, Formal analysis, Writing – original draft, Conception and design of study, analysis and/or interpretation of data, Drafting the manuscript, Approval of the version of the manuscript to be published. **Wenmin Qin:** Conceptualization, Writing – original draft, Conception and design of study, Drafting the manuscript, Approval of the version of the manuscript to be published. **Lunche Wang:** Funding acquisition, Writing – review & editing, acquisition of data, revising the manuscript, critically for important intellectual content, Approval of the version of the manuscript to be published. **Bo Hu:** Funding acquisition, Writing – review & editing, acquisition of data, revising the manuscript, critically for important intellectual content, Approval of the version of the manuscript to be published. **Yan Song:** Formal analysis, Writing – review & editing, Formal analysis, and/or interpretation of data, revising the manuscript critically for important intellectual content, Approval of the version of the manuscript to be published. **Ming Zhang:** Writing – review & editing, Funding acquisition, acquisition of data, revising the manuscript critically for important intellectual content, Approval of the version of the manuscript to be published.

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

This work was financially supported by the National Natural Science Foundation of China (No. 42001016), the Special Fund for Basic Scientific Research of Central Colleges, China University of Geosciences, Wuhan (No. 111–162301182738). We would like to thank the NASA Earth data website (https://search.earthdata.nasa.gov/search) for providing MODIS aerosol and auxiliary products and the NASA Goddard

Space Flight Center (https://aeronet.gsfc.nasa) for providing AERONET ground measurements.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2022.119219.

References

Agapiou, A., 2017. Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine© applications. INT J DIGIT EARTH 10, 85–102.

Barbero, F.J., Lopez, G., Batlles, F.J., 2006. Determination of Daily Solar Ultraviolet Radiation Using Statistical Models and Artificial Neural Networks. Annales Geophysicae. Copernicus GmbH, pp. 2105–2114.

Barnes, P.W., Williamson, C.E., Lucas, R.M., Robinson, S.A., Madronich, S., Paul, N.D., Bornman, J.F., Bais, A.F., Sulzberger, B., Wilson, S.R., 2019. Ozone depletion, ultraviolet radiation, climate change and prospects for a sustainable future. Nat. Sustain. 2, 569–579.

Bisong, E., 2019. An overview of Google cloud platform services. Building Machine Learning and Deep Learning Models on Google Cloud Platform 7–10.

Bo, H., Yue-Si, W., Guang-Ren, L., 2009. Properties of solar radiation over Chinese arid and semi-arid areas. Atmospheric and Oceanic Science Letters 2, 183–187.

Booth, C.R., Lucas, T.B., Morrow, J.H., Weiler, C.S., Penhale, P.A., 1994. The United States National Science Foundation's polar network for monitoring ultraviolet radiation. Ultraviolet radiation in Antarctica: Measurements and biological effects 62, 17–37.

Brown, M.G., Skakun, S., He, T., Liang, S., 2020. Intercomparison of machine-learning methods for estimating surface shortwave and photosynthetically active radiation. REMOTE SENS-BASEL 12, 372.

Chubarova, N.E., Pastukhova, A.S., Zhdanova, E.Y., Volpert, E.V., Smyshlyaev, S.P., Galin, V.Y., 2020. Effects of ozone and clouds on temporal variability of surface UV radiation and UV resources over Northern Eurasia derived from measurements and modeling. ATMOSPHERE-BASEL 11, 59.

Clark, L.A., Pregibon, D., 2017. Tree-based Models. Statistical Models in S. Routledge, pp. 377–419.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. Applied Multiple Regression/correlation Analysis for the Behavioral Sciences. Routledge.

Dubayah, R., Rich, P.M., 1995. Topographic solar radiation models for GIS. Int. J. Geogr. Inf. Syst. 9, 405–419.

Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X., Zou, H., 2019. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: a review and case study in China. Renew. Sustain. Energy Rev. 100, 186–212.

Farman, J.C., Gardiner, B.G., Shanklin, J.D., 1985. Large losses of total ozone in Antarctica reveal seasonal ClO x/NO x interaction. Nature 315, 207–210.

Ferrero, E., Eöry, M., Ferreyra, G., Schloss, I., Zagarese, H., Vernet, M., Momo, F., 2006. Vertical mixing and ecological effects of ultraviolet radiation in planktonic communities. Photochem. Photobiol. 82, 898–902.

García-Mora, T.J., Mas, J., Hinkley, E.A., 2012. Land cover mapping applications with MODIS: a literature review. INT J DIGIT EARTH 5, 63–87.

Gueymard, C.A., 2004. The sun's total and spectral irradiance for solar energy applications and solar radiation models. Sol. Energy 76, 423–453.

Guide, O.D.U.S., 2012. Ozone Monitoring Instrument (OMI) Data User's Guide. Guide, Tech.

Hänsch, R., Schulz, K., Sörgel, U., 2018. Machine Learning Methods for Remote Sensing Applications: an Overview. Berlin: Earth Resources and Environmental Remote Sensing/GIS Applications IX. Inter Soc Opt Photo 10790, 1079002.

Hao, D., Bisht, G., Gu, Y., Lee, W., Liou, K., Leung, L.R., 2021. A parameterization of sub-grid topographical effects on solar radiation in the E3SM Land Model (version 1.0): implementation and evaluation over the Tibetan Plateau. Geosci. Model Dev. (GMD) 14, 6273–6289.

Herman, J.R., Krotkov, N., Celarier, E., Larko, D., Labow, G., 1999. Distribution of UV radiation at the Earth's surface from TOMS-measured UV-backscattered radiances. J. Geophys. Res. Atmos. 104, 12059–12076.

Hrisko, J., Ramamurthy, P., Gonzalez, J.E., 2021. Estimating heat storage in urban areas using multispectral satellite data and machine learning. Remote Sens. Environ. 252, 112125.

Jiang, H., Lu, N., Huang, G., Yao, L., Qin, J., Liu, H., 2020. Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data. APPL ENERG 270, 115178.

Jiang, H., Lu, N., Qin, J., Tang, W., Yao, L., 2019. A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. Renew. Sustain. Energy Rev. 114, 109327.

King, M.D., Menzel, W.P., Kaufman, Y.J., Tanré, D., Gao, B., Platnick, S., Ackerman, S.A., Remer, L.A., Pincus, R., Hubanks, P.A., 2003. Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS. IEEE T GEOSCI REMOTE 41, 442–458.

Krotkov, N.A., Herman, J.R., Bhartia, P.K., Fioletov, V., Ahmad, Z., 2001. Satellite estimation of spectral surface UV irradiance: 2. Effects of homogeneous clouds and snow. J. Geophys. Res. Atmos. 106, 11743–11759.

Kujanpää, J., Lakkala, K., Lindfors, A., Kalakoski, N., Sundström, A., Ialongo, I., Arola, A., Hassinen, S., Tamminen, J., 2021. TROPOMI UV Radiation Product and Recent Applications. EGU General Assembly Conference Abstracts, pp. U21–U15853.

Laguarda, A., Abal, G., 2019. Assessment of empirical models to estimate UV-A, UV-B and UV-E solar irradiance from GHI. In: Proceedings of the ISES Solar World Congress 2019, Santiago de Chile, Chile. ISES, 4-7 de Noviembre, 2019.

Lakkala, K., Kujanpää, J., Brogniez, C., Henriot, N., Arola, A., Aun, M., Auriol, F., Bais, A. F., Bernhard, G., De Bock, V., 2020. Validation of the TROPOspheric Monitoring Instrument (TROPOMI) surface UV radiation product. Atmos. Meas. Tech. 13, 6999–7024.

Lamy, K., Portafaix, T., Brogniez, C., Godin-Beekmann, S., Bencherif, H., Morel, B., Pazmino, A., Metzger, J.M., Auriol, F., Deroo, C., 2018. Ultraviolet radiation modelling from ground-based and satellite measurements on Reunion Island, southern tropics. Atmos. Chem. Phys. 18, 227–246.

Lamy, K., Portafaix, T., Brogniez, C., Lakkala, K., Pitkänen, M.R., Arola, A., Forestier, J., Amelie, V., Toihir, M.A., Rakotoniaina, S., 2021. UV-Indien network: ground-based measurements dedicated to the monitoring of UV radiation over the western Indian Ocean. Earth Syst. Sci. Data 13, 4275–4301.

Letu, H., Shi, J., Li, M., Wang, T., Shang, H., Lei, Y., Ji, D., Wen, J., Yang, K., Chen, L., 2020a. A review of the estimation of downward surface shortwave radiation based on satellite data: methods, progress and problems. Sci. China Earth Sci. 63, 774–789.

Letu, H., Yang, K., Nakajima, T.Y., Ishimoto, H., Nagao, T.M., Riedi, J., Baran, A.J., Ma, R., Wang, T., Shang, H., 2020b. High-resolution retrieval of cloud microphysical properties and surface solar radiation using Himawari-8/AHI next-generation geostationary satellite. Remote Sens. Environ. 239, 111583.

Liang, S., Zheng, T., Liu, R., Fang, H., Tsay, S.C., Running, S., 2006. Estimation of incident photosynthetically active radiation from Moderate Resolution Imaging Spectrometer data. J. Geophys. Res. Atmos. 111.

Lindfors, A.V., Kujanpää, J., Kalakoski, N., Heikkilä, A., Lakkala, K., Mielonen, T., Sneep, M., Krotkov, N.A., Arola, A., Tamminen, J., 2018. The TROPOMI surface UV algorithm. Atmos. Meas. Tech. 11, 997–1008.

Liu, H., Hu, B., Zhang, L., Zhao, X.J., Shang, K.Z., Wang, Y.S., Wang, J., 2017. Ultraviolet radiation over China: spatial distribution and trends. Renew. Sustain. Energy Rev. 76, 1371–1383.

Lucas, R.M., Yazar, S., Young, A.R., Norval, M., De Gruijl, F.R., Takizawa, Y., Rhodes, L. E., Sinclair, C.A., Neale, R.E., 2019. Human health in relation to exposure to solar ultraviolet radiation under changing stratospheric ozone and climate. PHOTOCH PHOTOBIO SCI 18, 641–680.

Ma, R., Letu, H., Yang, K., Wang, T., Shi, C., Xu, J., Shi, J., Shi, C., Chen, L., 2020. Estimation of surface shortwave radiation from Himawari-8 satellite data based on a combination of radiative transfer and deep neural network. IEEE T GEOSCI REMOTE 58, 5304–5316.

Min, M., Wu, C., Li, C., Liu, H., Xu, N., Wu, X., Chen, L., Wang, F., Sun, F., Qin, D., 2017. Developing the science product algorithm testbed for Chinese next-generation geostationary meteorological satellites: Fengyun-4 series. J METEOROL RES-PRC 31, 708–719.

Parisi, A.V., Igoe, D., Downs, N.J., Turner, J., Amar, A., Jebar, M.A., 2021. Satellite monitoring of environmental solar ultraviolet A (UVA) exposure and irradiance: a review of OMI and GOME-2. REMOTE SENS-BASEL 13, 752.

Pei, C., He, T., 2019. UV radiation estimation in the United States using MODIS data. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1880–1883.

Qin, W., Wang, L., Wei, J., Hu, B., Liang, X., 2020. A novel efficient broadband model to derive daily surface solar Ultraviolet radiation (0.280–0.400 μm). Sci. Total Environ. 735, 139513.

Santos, J.B., Villán, D.M., Castrillo, A.D.M., 2011. Analysis and cloudiness influence on UV total irradiation. Int. J. Climatol. 31, 451–460.

Singh, S., Mishra, A.K., Jose, S., Lodhi, N.K., 2022. Atmospheric pollution and solar ultraviolet radiation in Asia. Asian Atmospheric Pollution. Elsevier 129–146.

Sun, J., Veefkind, P., Van Velthoven, P., Levelt, P., 2021. Aerosol absorption over land derived from the Ultra-Violet aerosol index by deep learning. IEEE J-STARS.

Taipe, C.W., Mendoza, E.G., Flores, H.H., 2021. Validation of ultraviolet index data from the Ozone Monitoring Instrument (OMI) based on measurements from meteorological stations in the city of Puno. In: Journal of Physics: Conference Series. IOP Publishing, 12005.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: a meta-analysis and systematic review. ISPRS J PHOTOGRAMM 164, 152–170.

Tang, W., Qin, J., Yang, K., Liu, S., Lu, N., Niu, X., 2016. Retrieving high-resolution surface solar radiation with cloud parameters derived by combining MODIS and MTSAT data. Atmos. Chem. Phys. 16, 2543–2557.

Tang, W., Qin, J., Yang, K., Niu, X., Min, M., Liang, S., 2017. An efficient algorithm for calculating photosynthetically active radiation with MODIS products. Remote Sens. Environ. 194, 146–154.

Tanskanen, A., Krotkov, N.A., Herman, J.R., Arola, A., 2006. Surface ultraviolet irradiance from OMI. IEEE T GEOSCI REMOTE 44, 1267–1271.

Thomas, P., Swaminathan, A., Lucas, R.M., 2012. Climate change and health with an emphasis on interactions with ultraviolet radiation: a review. Global Change Biol. 18, 2392–2405.

Vapnik, V., 2013. The Nature of Statistical Learning Theory. Springer science & business media.

Verdebout, J., 2000. A method to generate surface UV radiation maps over Europe using GOME, Meteosat, and ancillary geophysical data. J. Geophys. Res. Atmos. 105, 5049–5058.

Wang, L., Gong, W., Feng, L., Hu, B., 2015. UV variability in an arid region of Northwest China from measurements and reconstructions. Int. J. Climatol. 35, 1938–1947.

Wang, Y., Jiang, B., Liang, S., Wang, D., He, T., Wang, Q., Zhao, X., Xu, J., 2019. Surface Shortwave net radiation estimation from Landsat TM/ETM+ data using four machine learning algorithms. REMOTE SENS-BASEL 11, 2847.

Wei, Y., Zhang, X., Hou, N., Zhang, W., Jia, K., Yao, Y., 2019. Estimation of surface downward shortwave radiation over China from AVHRR data based on four machine learning methods. Sol. Energy 177, 32–46.

Williamson, C.E., Zepp, R.G., Lucas, R.M., Madronich, S., Austin, A.T., Ballaré, C.L., Norval, M., Sulzberger, B., Bais, A.F., McKenzie, R.L., 2014. Solar ultraviolet radiation in a changing climate. Nat. Clim. Change 4, 434–441.

Yang, J., Zhang, Z., Wei, C., Lu, F., Guo, Q., 2017. Introducing the New Generation of Chinese Geostationary Weather Satellites, Fengyun-4, vol. 98. B AM METEOROL SOC, pp. 1637–1658.

Yang, L., Zhang, X., Liang, S., Yao, Y., Jia, K., Jia, A., 2018. Estimating surface downward shortwave radiation over China based on the gradient boosting decision tree method. REMOTE SENS-BASEL 10, 185.

Yang, Z., Li, W., Chen, Q., Wu, S., Liu, S., Gong, J., 2019. A scalable cyberinfrastructure and cloud computing platform for forest aboveground biomass estimation based on the Google Earth Engine. INT J DIGIT EARTH 12, 995–1012.

Yeom, J., Park, S., Chae, T., Kim, J., Lee, C.S., 2019. Spatial assessment of solar radiation by machine learning and deep neural network models using data provided by the COMS MI geostationary satellite: a case study in South Korea. SENSORS-BASEL 19, 2082.

Zhao, R., He, T., 2022. Estimation of 1-km resolution all-sky instantaneous erythemal UV-B with MODIS data based on a deep learning method. REMOTE SENS-BASEL 14, 384.