Constructing a gridded direct normal irradiance dataset in China during 1981–2014

WenminQin [a], Lunche Wang [a,*], Christian A. Gueymard [b], Muhammad Bilal [c], Aiwen Lin [d], Jing Wei [e,f], Ming Zhang [a], Xuefang Yang [a]

[a] Hubei Key Laboratory of Critical Zone Evolution, School of Geography and Information Engineering, China University of Geosciences, Wuhan, China
[b] Solar Consulting Services, P.O. Box 392, Colebrook, NH, 03576, USA
[c] School of Marine Science, Nanjing University of Information Science & Technology, Nanjing, 2100444, China
[d] School of Resource and Environmental Science, Wuhan University, Wuhan, 430079, China
[e] State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China
[f] Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

ARTICLE INFO

ABSTRACT

High-quality direct normal irradiance (DNI) observations are of vital importance for the optimal design, installation and profitability of CST plants. Numerous models have been developed for estimating DNI at regional or global scales. The goal of this study is to generate a gridded DNI dataset for all-sky conditions over mainland China during 1981–2014, based on a broadband DNI estimates for clear-sky conditions (obtained with the REST2_v9.1 model) and cloud transmittance estimates using sunshine observations. The results indicate that the REST2_v9.1 model can be used to estimate DNI with high accuracy and consistency, owing to its robust two-band parameterization of the radiation transfer processes. Comparing daily DNI modeled predictions to measurements at 6 BSRN (Baseline Surface Radiation Network) stations in East Asia results in relatively low errors statistics: RMSE, MAE, RMSER, MAER and R of 1.436 MJm^{-2}, 0.900 MJm^{-2}, 22.26%, 13.95% and 0.972. Somewhat lessened agreement is found at the 17 CMA (China meteorological administrations) stations: 4.064 MJm^{-2}, 2.864 MJm^{-2}, 34.41%, 24.26% and 0.914, respectively. A gridded DNI dataset is constructed using sunshine duration measurements at 2474 CMA meteorological stations and the MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications, version 2) reanalysis products. The spatial and temporal variations of DNI in different climate zones throughout China are also investigated. The gridded DNI datasets generated in this study would assist in numerous solar resource studies and solar energy applications.

1. Introduction

Solar energy is appealing as a clean, renewable, sustainable and environmentally friendly energy source for the continuation of life on our planet [1,2]. Many countries have devoted great attention and effort on the development of solar electrical applications since the 2000s [5–8]. In particular, various types of solar energy systems have been developed to convert solar radiation into low-grade heat, process heat or electricity, using either flat-plate or concentrating solar collectors. Whereas the former collector type can benefit from both direct and diffuse forms of solar radiation, the latter type is strictly sensitive to direct irradiance [9]. Therefore, evaluating, mapping, and monitoring the direct normal irradiance (DNI) is of vital importance for the proper

design, financing, and operation of solar power plants using concentrating technologies (CSP) [10]. China is the country with the largest thermal power generation, making it the largest emitter of greenhouse gases [11]. The huge demand for electricity and energy consumption make the Chinese government turn to vigorously develop CSP industry [12]. China is in the leading position in the construction and planned installed capacity of CSP power generation [13]. Clear understanding of DNI in China would improve the efficiency of CSP resources utilization [14].

DNI is the amount of solar radiation received per unit area by a surface perpendicular to the sun rays that come in a straight line from the direction of the sun at its current position in the sky [15]. DNI is typically measured with a pyrheliometer mounted on a sun tracker. DNI

Table 1
Basic information of the CMA and BSRN stations used in this study.

Stations	Code	Name	Latitude	Longitude	Elevation (m)
BSRN	FUA	Fukuoka, Japan	33.58° N	130.38° E	3.0
	TAT	Tateno, Japan	36.06° N	140.13° E	25.0
	SAP	Sapporo, Japan	43.06° N	141.33° E	17.2
	LLN	Lulin, China	23.47° N	120.87° E	125.0
	ISH	Ishigakijima, Japan	24.34° N	124.16° E	5.7
	XIA	Xianghe, China	39.75° N	116.96° E	32.0
CMA	MH	Mohe	53.47° N	122.37°E	296
	HEB	Herbin	45.75° N	126.77°E	142
	WUR	Wurmuqi	43.78° N	87.62° E	918
	KAS	Kashi	39.47° N	75.98° E	1289
	EJN	Ejinaqi	41.95° N	101.07°E	941
	GEM	Germu	36.42° N	94.90°E	2808
	LAZ	Lanzhou	36.05° N	103.88°E	1517
	SHY	Shengyang	41.73° N	123.45°E	43
	BEJ	Beijing	39.93° N	116.28°E	54
	LAS	Lasha	29.67° N	91.13°E	3649
	CHD	Chengdu	30.70° N	103.83°E	539
	KUM	Kunming	25.02° N	102.68°E	1891
	ZHZ	Zhengzhou	34.72° N	113.65°E	110
	WUH	Wuhan	30.62° N	114.13°E	23
	SHH	Shanghai	31.40° N	121.48°E	4
	GUZ	Guangzhou	23.13° N	113.32°E	7
	SAY	Sanya	18.23° N	109.52°E	6

can also be modeled in different ways, e.g., through proper decomposition from global solar radiation measurements [16,17], or through direct modeling from atmospheric data [18–20]. Measured DNI information has high value because it has the lowest uncertainty of all solar radiation components, if it follows the best practices [21]. To be useful on a large scale, such measurements should be obtained from a dense network of observing stations. Despite worldwide continuous efforts to establish solar radiation measurement stations in recent years, the needed networks are still too sparse for high-end DNI applications, owing to the high cost and maintenance effort of observation platforms [22]. Additionally, equipmentmal function, operation-related problems, and data quality-control requirements are other limitations for DNI observations. Thus, it is of great significance to estimate DNI using modeling.

From a recent review on existing DNI models and their performance [23], it is clear that two types of radiation models can be used in the current practice of solar resource assessments: empirical models and physically-based models. Numerous empirical models have been developed for the estimation of DNI [24–27]. Empirical models assume that solar radiation is directly linked to meteorological variables such as sunshine duration, air temperature, relative humidity or site elevation—or even just solar position [26]. For example, Louche et al. developed an empirical formula in forms of fifth order polynomial to retrieve DNI values at Ajaccio [28]. Tiris et al. established the empirical relationship between the ratio of DNI to global solar radiation (RDNI) and the ratio of the daily sunshine hour to the daily maximum sunshine hour (n/N) in Gebze [29]. Benson et al. proposed an empirical sunshine-based model to reveal the linear regression relations for monthly RDNI and n/N with RMS of 8.8%~24.4% [30]. Li et al. developed a humidity based model for retrieving direct solar radiation in Yunnan Province, using seasonal averaged daily absolute humidity in the dry seasons and wet seasons [31]. However, empirical models are subjected to poor universality and lower accuracy [23,32], because all their numerical coefficients are calculated based on some simple relationship between DNI and a few model inputs. In contrast, physically-based models can relate DNI to various atmospheric processes, which are parameterized from strict radiative transfer theory. Many physically-based DNI estimation models have been developed to estimate DNI at regional or global scale [18,19,33–39]. Physically-based DNI models can be roughly divided into spectral models and broadband

models. Leckner [36] proposed a spectral model for the estimation of DNI that considered the main spectral transmittances of solar radiation in the atmosphere, including Rayleigh scattering, ozone absorption, absorption by uniformly gases, aerosol extinction and water vapor absorption. Gueymard [40] developed the more elaborate 'SMARTS' spectral model to predict direct normal irradiance, diffuse irradiance and global irradiance incident on the Earth's surface, covering the whole shortwave solar spectrum (280–4000 nm) with high spectral resolution. The SMARTS model is still being updated and validated [32], but its relative complexity may not be ideal for intensive simulations at continental scale. Although, the accuracy of spectral models may be somewhat higher than that of broadband models because of their superior spectral resolution, the latter are much faster to operate. For instance, the REST2 model [18] uses the parameterizations of SMARTS results, and a much coarser spectral resolution (two wide bands vs. 2002 wavelengths in SMARTS). It divides the shortwave solar spectrum into two parts including Band 1 (below 700 nm) and Band 2 (above 700 nm), each one evaluating the irradiance components from a set of transmittances. Remarkably, the performance of REST2 is comparable to that of SMARTS [32,41], despite the simplifications embedded in the former.

As with all radiation models, the accuracy is ultimately dependent on that in their key inputs, as demonstrated elsewhere [41]. Since, moreover, irradiance predictions are needed on a large scale, an ideal situation would be that all the inputs required by a radiation models are accurately available at high spatiotemporal resolution. In practice, however, the required inputs are observed at only sparse meteorological or sun photometric stations. Popular satellite-derived data, such as atmosphere and land products from MODIS, Meteos at or MTSAT, provide an efficient way to retrieve some or all of the required inputs to DNI models at regional or global scales. Numerous models [42–46] have been developed to estimate DNI using satellite signals. Despite the effectiveness of satellite-based DNI models, the accuracy of satellite retrievals is subject to a lot of uncertainties like calibration, cloud screening, aerosol model or surface-albedo artifacts. Moreover, satellite records are most often not perfectly continuous over time or space, particularly over decadal periods. In contrast, reanalysis databases, such as NASA's Modern Era Retrospective-Analysis for Research and Applications, version 2 (MERRA-2) from NASA or ECMWF's CAMS cover the whole globe at hourly resolution over many decades.

In China, the solar radiation observation network has been carried out by the China Meteorological Administration (CMA) since the 1960s, providing daily global, direct and diffuse solar radiation measurements. These CMA radiation stations cover most areas of China and are scattered over a number of climate zones [47]. However, these CMA stations with high-quality DNI measurements were still too sparse for solar radiation applications. For example, there are only 17 first-class radiation stations currently providing daily DNI data in China. Constructing a gridded and gap less DNI dataset for China spanning 34 years (1981–2014) is one main justification to this study, aimed at helping the solar industry's development in China.

In what follows, DNI measurements at 17 first-class CMA meteorological stations over mainland China are used to evaluate the performance of the latest version (9.1) of REST2 during 1993–2014. In a subsequent step, a DNI database is constructed using MERRA-2 products and sunshine duration measurements at 2474 CMA stations, combined with the clear-sky REST2 predictions. Finally, the spatiotemporal variations of DNI over different climate zones and terrains in mainland China are investigated. Overall, this study should prove helpful in solar resource and energy applications that need long-term gridded DNI data with moderate spatio-temporal resolution and acceptable accuracy.

2. Materials and methods

2.1. Sites and data processing

Daily direct normal irradiance (*DNI*) measurements during
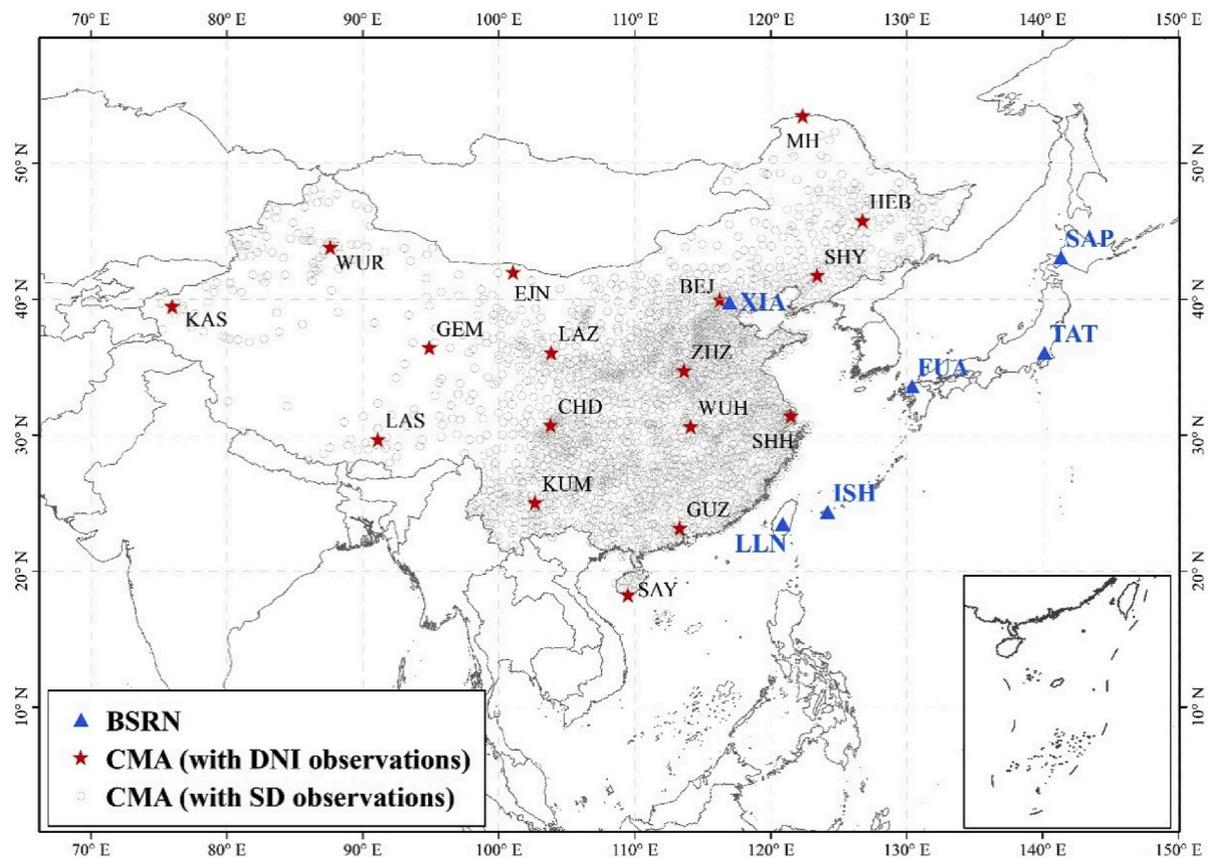
Fig. 1. Spatial distributions of the BSRN and CMA stations that are used in this study.
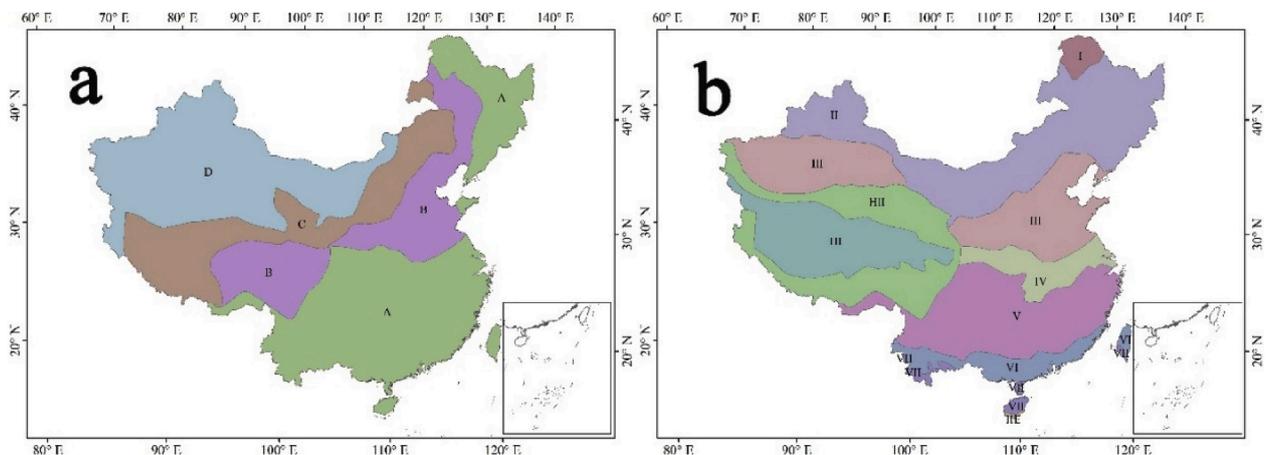
Fig. 2. The climatic zones over mainland China.

1993–2014 at 17 CMA stations over mainland China and 6 Baseline Surface Radiation Network (BSRN) stations in East Asia were used for model validations of REST2_V9.1 model. Before 1993, the observation radiometers that were used at these CMA stations were similar to those used in the Soviet Union. However, the observation precision of these radiometers was impacted by imprecise instrument calibration, sensor aging and data quality control method [48]. To address these issues, the CMA updated the observation instruments to a modern design of thermopile pyrheliometer (DFY-4) with higher accuracy and robustness [49]. The Baseline Surface Radiation Network (BSRN; https://bsrn.awi. de/) provides 1-min DNI data at many stations around the world. These data have been checked for ensuring the data quality using various data quality control methods. Table 1 illustrated the basic information of

these CMA and BSRN stations. Fig. 1 showed the spatial distributions of the CMA and BSRN meteorological stations that were used in this study. These stations covered most areas of China (18.22°Ñ53.47°N, 75.75°E~126.77°E) and East Asia (23.47°Ñ43.06°N, 120.87°W~141.33°W) with distinct climatic and terrain features. Meanwhile, 2474 CMA stations with sunshine duration measurements were used to calculated the cloud transmittance on *DNI* (see Fig. 2).

The climate regionalization data provided by Resource and environment science data center of Chinese Academy of Sciences (http://www.resdc.cn) were used to investigate the spatial-temporal variations of *DNI* in different climate zones throughout China. (A for humid, B for semi-humid, C for semi-arid, D for arid; I for cold temperate, II for mid temperate, III for warm temperate, IV for north
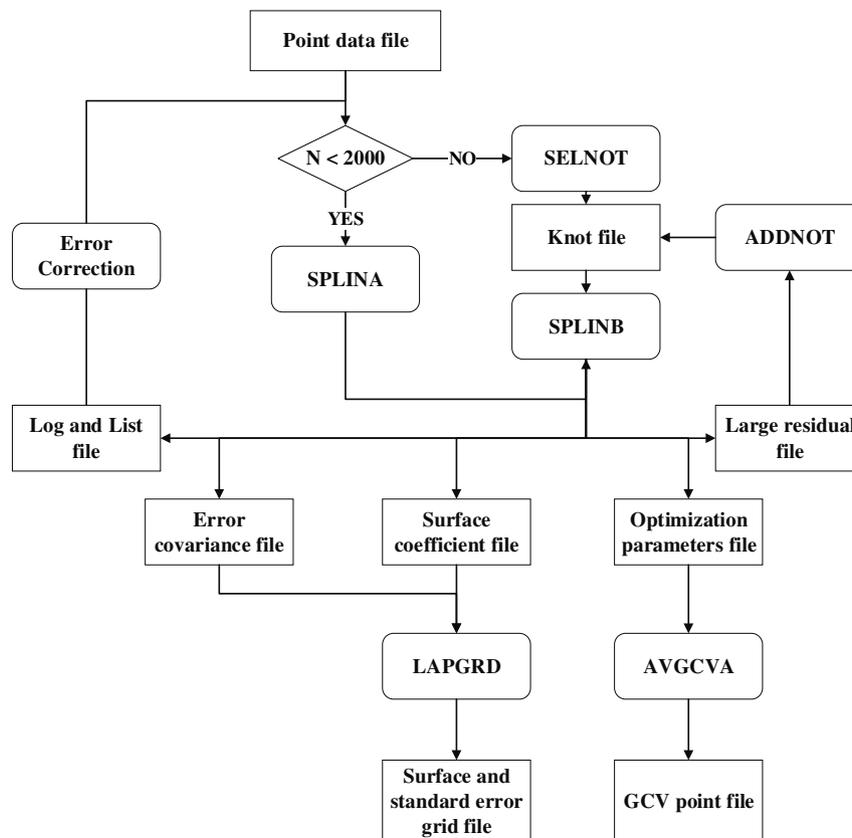
Fig. 3. The flowchart of the processes of Anusplin tool.

subtropical zone, V for the mid-subtropics, VI for the south subtropics, VII for the edge of tropical zone, HI for sub-frigid zone in plateau, HII for temperature zone in plateau, IIE for mid tropical zone with humid weather).

2.2. REST2_V9.1 model

REST2_V9.1 model is a physically based model for predicting hourly and daily cloudless-sky broadband irradiance, illuminance and *DNI*. REST version 1 was firstly developed by Gueymard [34] in November 2003, then corrected and modified by Gueymard [50] (REST to REST2_V9.1). REST2 have been validated as one of the best broadband solar radiation estimation models, which has been widely used in lot of solar radiation researches [34]. The REST2_V9.1 model has corrected the diffuse calculation under low-AOD, near-Rayleigh conditions in the model. The *DNI* in REST2_V9.1 could be obtained using following equations:

$$DNI = T_R T_g T_o T_n T_w T_a E_{0ni} \qquad (1)$$

where $T_R T_g T_o, T_n$, T_w and T_a were the transmittances for Rayleigh scattering, uniformly mixed gases absorption, ozone absorption, nitrogen dioxide absorption, water vapor absorption and aerosol extinction, respectively. E_{0ni} is the extraterrestrial solar radiation (H) at a given location. These transmittances have been obtained accurately by fitting a large number of parametric runs of the SMARTS code to computationally efficient polynomial ratios [50]. More detail description and resulting equations of REST2 model could be found in Ref. [18,50].

There are 15 input parameters in REST2_V9.1 model: Angsrom's wavelength exponents (*alfa1* for wavelength above 0.7 nm, *alfa2* for wavelength above 0.7 nm), elevation (*Altit*), AOD at 550 nm (*AOD550*), air pressure (*p*), relative humidity (*RH*), regional ground albedo (*rog*), air temperature (*T*), reduced NO2 vertical pathlength (*un*), reduced ozone

vertical pathlength (u0), precipitable water vapor (*w*), year, month, day, solar zenith angle and time zones. These parameters could be formed into 32 types of input parameter combinations. Considering the data availability, hourly reanalysis meteorological records including *AOD550*, *rog*, *p* and *w* derived from MERRA-2 data set during 1981–2014 were used as model inputs for REST2_V9.1 model. The spatial resolution of the MERRA-2 data set that were used in this study is 0.50° (lat) *0.625° (lon).

2.3. Data quality control

Five main quality check processes including the climate limit value or allowable value check, station extreme value check, internal consistency check among timing value, the time consistency check for daily average value and daily extreme value, and spatial consistency check have been done to ensure quality of the sunshine duration measurements. Then, the sunshine duration measurements were marked with 0, 1, 2 and 8, which represent correct data, suspicious data, wrong data and missing data, respectively. These work have been done by the China Meterological Administration. Meanwhile, the quality of DNI measurements was also conducted following the rule that each measured DNI value should not exceed the global solar irradiance at the top of the atmosphere (G0) at the same geographical location, otherwise it will be directly deleted in the DNI measurements. For DNI estimates, the estimated DNI value that exceeded the G0 will be replaced by the mean values in the nearby days. G0 could be calculated as following equation:

$$G_0 = 24 S_0 L_0 \left[\left(\frac{\pi}{180} \right) \gamma (sin\delta sin \varnothing) + (cos\delta cos \varnothing sin\gamma) \right] \pi \qquad (2)$$

2.4 where γ is the sunrise hour angle; S_0 represent the solar constant (1367 W m^{-2}); L_0 is the earth orbit correction factor; δ is the solar declination angle; $\varnothing$ is the geographic latitude. *Anusplin.*

The sunshine durations during 1981–2014 were routinely measured
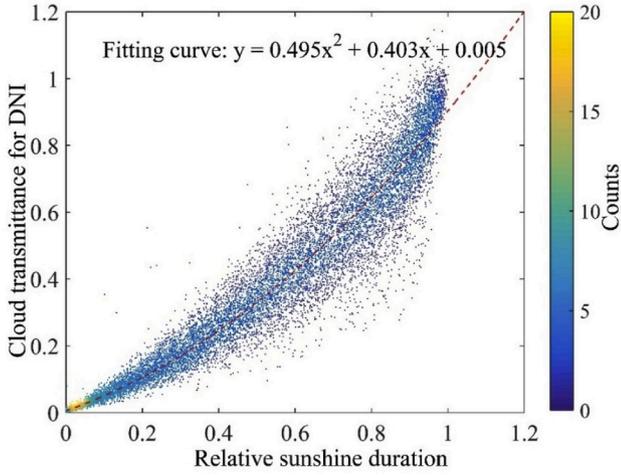
Fig. 4. The relationship between the relative sunshine duration and cloud transmittance for *DNI*.
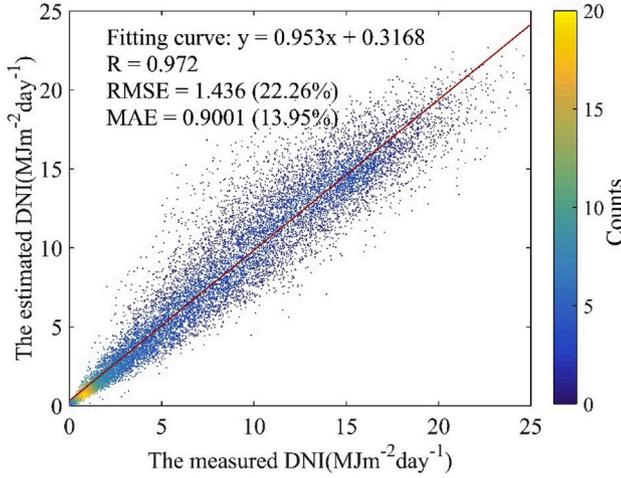
Fig. 5. Comparison between the estimated and measured *DNI* in all-sky condition at BSRN stations.

at 2474 CMA stations over mainland China. However, these stations were still too sparse for the *DNI* estimation in this study, thus, we produced the gridded sunshine duration data (0.50° (lat) *0.625° (lon)) using these sunshine durations measurements at CMA stations based on Anusplin tool. The Anusplin tool was widely used in climate data interpolation [51]. Detail description of the Anusplin tool could be found in Refs. [52]. The main interpolation process in Anusplin tool was shown in Fig. 3.

2.4. Comparisons of measures of fit

The measures of fit-used in the present study include the root mean square error (RMSE, MJm^{-2}day^{-1}), the mean absolute bias error (MAE, MJm^{-2}day^{-1}), the relatively root mean square error (RMSER, %), the relatively mean absolute bias error (MAER, %) and the correlation coefficient (R), which can be expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Xm_i - Xo_i)^2} \quad (3)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|Xm_i - Xo_i| \quad (4)$$

$$RMSER = \frac{100}{\overline{Xo}} \times \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Xm_i - Xo_i)^2} \quad (5)$$

$$MAER = \frac{100}{\overline{Xo}} \times \frac{1}{N}\sum_{i=1}^{n}|Xm_i - Xo_i| \quad (6)$$

$$R = \frac{\sum_{i=1}^{n}(Xm_i - \overline{Xm})(Xo_i - \overline{Xo})}{\sqrt{\sum_{i=1}^{n}(Xm_i - \overline{Xm})^2}\sqrt{\sum_{i=1}^{n}(Xo_i - \overline{Xo})^2}} \quad (7)$$

where *N* and bar respectively indicate the number of data and mean of the variables; *Xm* and *Xo* are the modeled and observed *DNI*; $\overline{Xm}$ and $\overline{Xo}$ are the mean value of the observed DNI value.

Although five indicators were introduced to reveal the accuracy of the estimated DNI values, each indicator could not represent the overall accuracy of the estimated DNI values. Therefore, a global performance indicator (GPI) was used in this study to access the overall model accuracy of DNI values. The GPI could be calculated as following equation:

$$GPI_i = \sum_{j=1}^{n}A_j(\check{Y} - Y_{ij}) \quad (8)$$

where $\check{Y}$ is the median of the scaled values of indicator j, Y_{ij} is the scaled value of indicator j for model i, and n is the number (5) of indicators. A_j equals −1 for R, and equals 1 for other indicators. The greater the accuracy of the estimated DNI values, the higher the value of the GPI.

3. Result and discussion

3.1. Cloud transmittance for direct normal irradiance

Because of the variability of cloud's shapes, types and stages, cloud has always been considered as the most uncertain factor for the estimation of surface solar radiation. In this study, the relative sunshine duration defined as the ratio between the measured sunshine durations (n, the length of time for which solar direct normal irradiance exceeds a threshold value of 120 W m^{-2}) and the maximum possible sunshine durations (N) was introduced to correct the cloud effect on direct normal irradiance.

Following the example of the Ångström-Prescott equation, we parameterized the cloud transmittance (τc) as a function of the relative sunshine duration (n/N), and the formula form was a quadratic polynomial formulation as follows:

$$\tau_c = \frac{R_b}{R_{b,clr}} = a + b\left(\frac{n}{N}\right) + c\left(\frac{n}{N}\right)^2 \quad (9)$$

where R_b and $R_{b,clr}$ were the daily all-sky *DNI* and clear-sky *DNI*, respectively; n and N are the sunshine duration and the maximum possible sunshine duration, respectively.

The BSRN stations could provide hourly and daily DNI measurements in all-sky conditions, but did not provide DNI measurements under clear-sky conditions. The clear sky detection method (CSD) proposed by Inman et al. [53] were used to detect periods of clear sky and reconstruct hourly and daily DNI datasets under clear-sky conditions using meteorological measurements at BSRN stations. Detail descriptions about CSD could be found in Ref. [53]. Finally, 17,365 samples were randomly selected to fit the parameter in formula 7.

Fig. 4 showed the relationship between the cloud transmittance for daily *DNI* and the relative sunshine duration. It should be stated that τ_c could not be fully parameterized with $\frac{n}{N}$, as shown in Fig. 4, the data points do not cluster tightly around the fitting curve. The uncertainty of the cloud transmittances needs to be discussed, which may be the most uncertain factor for the *DNI* estimations in all-sky conditions. The calibrated cloud transmittance for *DNI* was shown as following equations:
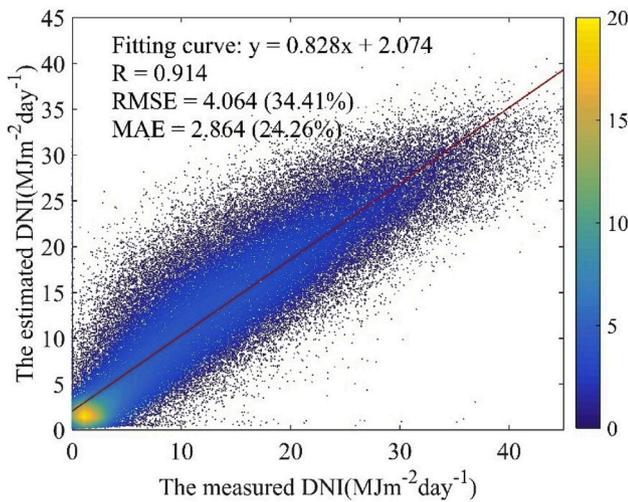
Fig. 6. Validation of the estimated *DNI* at 17 CMA stations.

$$\tau_c = 0.005 + 0.403\left(\frac{n}{N}\right) + 0.495\left(\frac{n}{N}\right)^2 \qquad (10)$$

Fig. 5 illustrated the validation result of the estimated *DNI* based on formula 8 using the 17,365 samples at BSRN stations that were used in this study. The estimated *DNI* showed good agreements with the *DNI*

measurements with RMSE, MAE, RMSER, MAER and R of 1.436 MJm^{-2}day^{-1} and 0.900 MJm^{-2}day^{-1}, 22.26%, 13.95%, and 0.972, respectively.

3.2. Validation of the estimated DNI at CMA stations

As mentioned above, *DNI* measurements (129,662 samples) during 1993–2014 at 17 CMA stations were used to validate the accuracy of daily estimated *DNI* measurements by REST2_V9.1 model. Fig. 6 indicated the validation results of the estimated *DNI* at these CMA stations under all-sky conditions. The daily *DNI* estimations showed high correlation with the ground *DNI* measurements with RMSE, MAE, RMSER, MAER and R of 4.064 MJm^{-2}day^{-1}, 2.864 MJm^{-2}day^{-1}, 34.41%, 24.26%, and 0.914, respectively.

The model deviations were mainly caused by four factors. First, there are too many low *DNI* values in Fig. 6, because low *DNI* values generally correspond to cloudy sky conditions which bring great uncertainties to *DNI* estimates. To verify this conclusion, we divided the samples into *DNI* below 2.700 (the first four percentile of the measured *DNI* in the samples) and *DNI* above 2.700. The result showed that the model deviations for *DNI* records below 2.7 (RMSER = 323.438%, MAER = 154.407%, R = 0.384) were significantly larger than that for *DNI* records above 3.850 (RMSER = 29.08%, MAER = 22.33%, R = 0.860). Secondly, the gridded sunshine duration data used in this study were interpolated from sunshine duration measurements at 2474 CMA stations over mainland China, which were the only available sd datasets

Table 2
The statistical indicators representing the model performance at 17 CMA stations.

Station Code	Name	RMSE	MAE	RMSER	MAER	R	GPI	Temperate zones	Humidity zones
MH	Mohe	3.991	2.887	30.45	22.02	0.915	−0.212	I	A
HEB	Herbin	3.845	2.883	33.78	25.33	0.910	−0.212	II	A
WUR	Wurmuqi	4.456	3.307	27.97	20.76	0.919	9.150	II	D
KAS	Kashi	5.891	4.712	35.15	28.11	0.891	−8.253	III	D
EJN	Ejinaqi	4.452	3.501	23.55	18.52	0.897	15.595	II	D
GEM	Germu	4.466	3.428	22.89	17.57	0.903	17.269	HII	D
LAZ	Lanzhou	6.202	4.698	40.42	30.62	0.839	−16.380	III	C
SHY	Shengyang	4.213	2.984	35.38	25.06	0.873	−2.044	II	A
BEJ	Beijing	3.367	2.447	25.72	18.69	0.940	15.437	III	B
LAS	Lasha	5.337	4.177	25.91	20.28	0.863	9.887	HII	C
CHD	Chengdu	3.504	2.240	58.19	37.20	0.872	−35.533	V	A
KUM	Kunming	4.014	2.910	28.52	20.68	0.931	9.534	V	A
ZHZ	Zhengzhou	2.736	1.958	29.62	21.20	0.931	10.132	III	B
WUH	Wuhan	3.819	2.685	37.62	26.45	0.908	−4.946	IV	A
SHH	Shanghai	2.948	2.106	33.68	24.07	0.924	2.844	IV	A
GUZ	Guangzhou	2.808	1.951	35.71	24.81	0.921	0.365	VI	A
SAY	Sanya	4.056	2.988	32.86	24.21	0.879	1.490	IIE	A
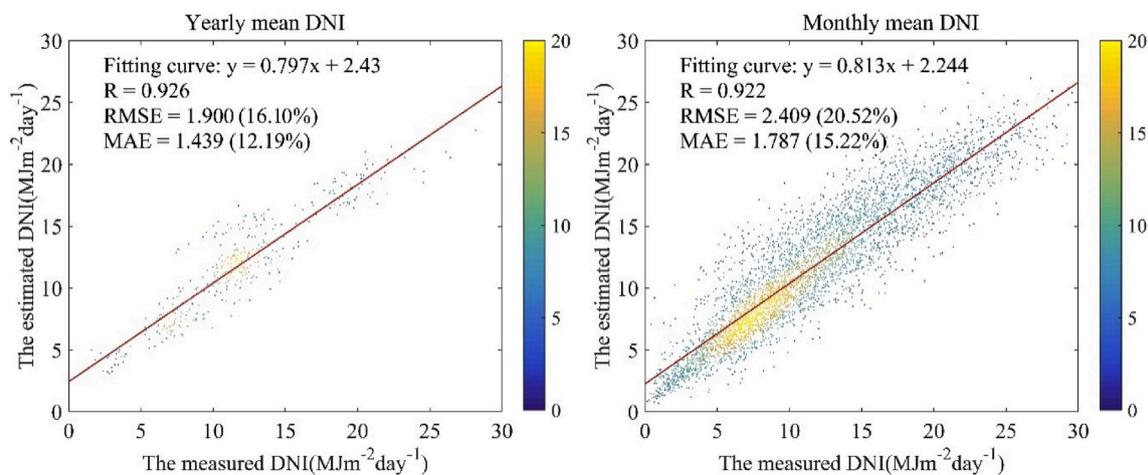
Fig. 7. Validation of the yearly and monthly mean *DNI* values at CMA stations.
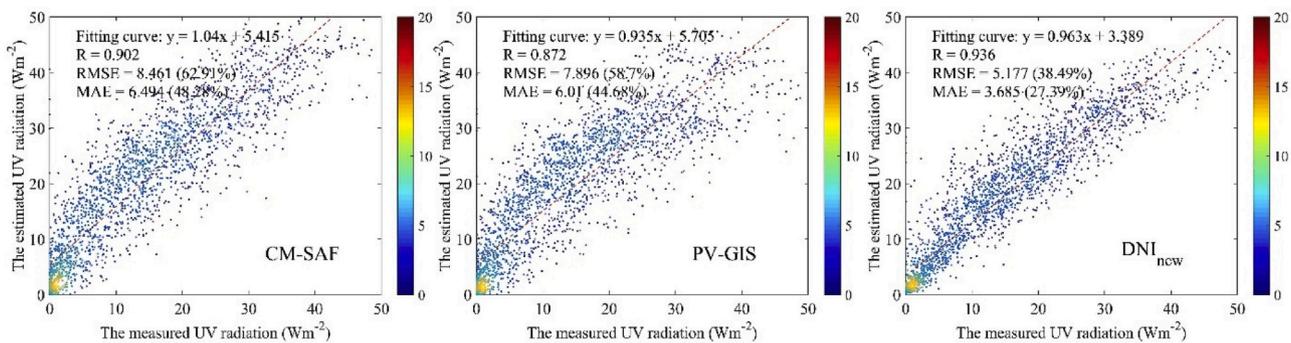
Fig. 8. Comparison of the accuracy of the daily mean DNI values derived from CM-SAF, PVGIS and DNI data in this study in 2007 at CMA stations.
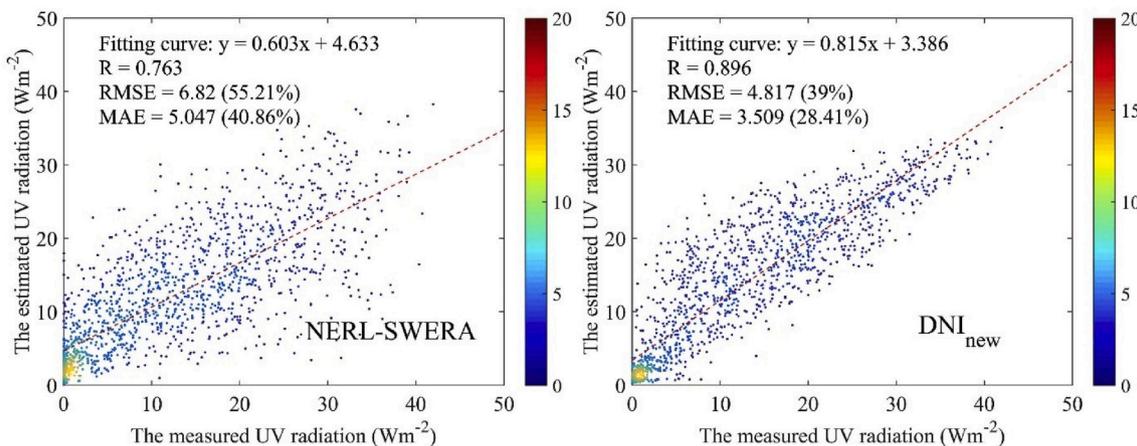
Fig. 9. Comparison of the accuracy of the daily mean DNI values derived from NERL-SWERA and DNI data in this study in 2002 at CMA stations.
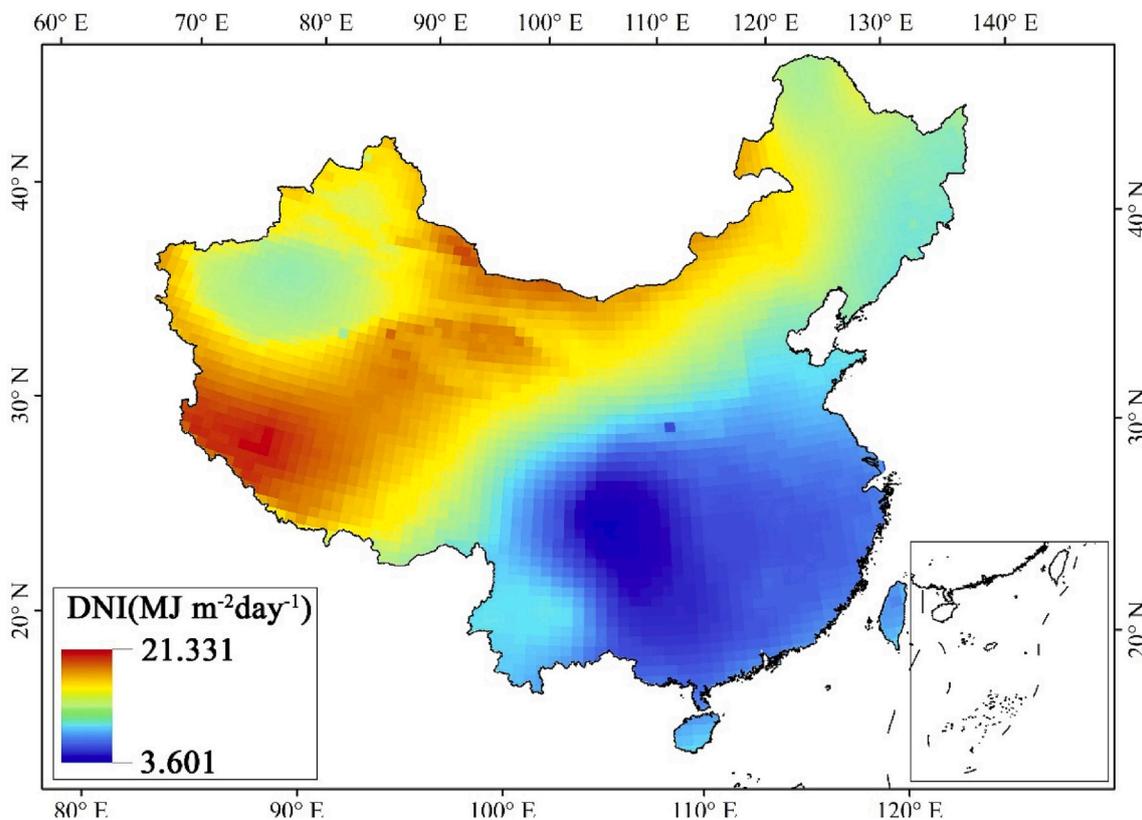
Fig. 10. Spatial distribution of *DNI* over mainland China.
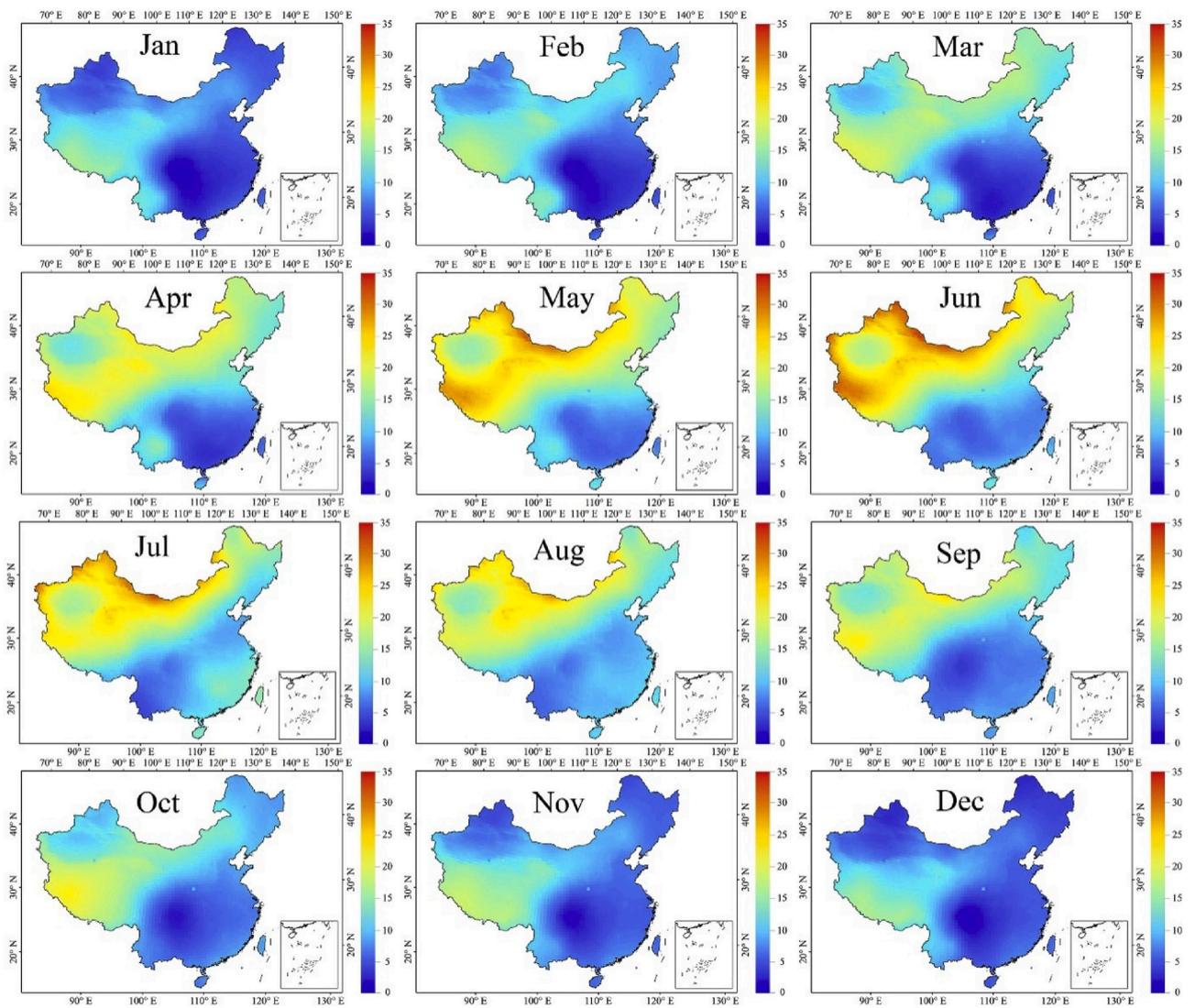
Fig. 11. Spatial and temporal distributions of *DNI* over mainland China.

with the highest density and accuracy in China. Thirdly, the interpolation method and the point density of the CMA stations may bring certain deviations to the accuracy of the *DNI* datasets. Further study should be conducted to improve the estimation of the cloud transmittances for *DNI*. Lastly, the spatial resolution of the input parameters and the output values were 0.50° (lat) *0.625° (lon), which may also degrade the accuracy of the estimated *DNI*.

Table 2 showed the statistical indicators representing the model deviations at 17 CMA stations. The result showed that the largest model deviation was found in CHD with RMSE, MAE, RMSER, MAER, R and GPI of 3.504 MJm^{-2}day^{-1}, 2.240 MJm^{-2}day^{-1}, 58.19%, 37.20%, 0.872 and −35.533, because the abundant precipitable water vapor and frequent cloud occurrences there would affect the model accuracy. Relatively larger estimation errors were also found in arid zones, which may be due to the severe dusty air conditions and poor quality of radiation measurements since solar radiation measurements would be seriously affected in dusty air conditions. For example, the RMSE, MAE, RMSER, MAER, R and GPI for LAZ were 6.202 MJm^{-2}day^{-1}, 4.698 MJm^{-2}day^{-1}, 40.42%, 30.62%, 0.839, −16.380, respectively. The relatively lower estimation errors were found in Plateau areas, because of the clear sky conditions there, for example, the RMSE, MAE, RMSER, MAER, R and GPI for GEM were 4.466 MJm^{-2}day^{-1}, 3.428 MJm^{-2}day^{-1}, 22.89%, 17.57%, 0.903 and 17.269, respectively; the RMSE, MAE, RMSER, MAER, R and GPI for LAS were 5.337

MJm^{-2}day^{-1}, 4.177 MJm^{-2}day^{-1}, 25.91%, 20.28%, 0.863 and 9.887, respectively.

Whether the accuracy of the estimated *DNI* could be applied for CST applications in China is worth of discussion. The yearly and monthly mean *DNI* records may be enough to meet the requirements of CST applications. Therefore, the yearly and monthly mean estimated *DNI* values were compared with the *DNI* measurements at the CMA stations. Fig. 7 showed the validation results of the yearly and monthly mean *DNI* values by REST2_V9.1 model. The result indicated that the yearly and monthly mean estimated *DNI* values showed higher accuracy with measured *DNI* values than that of the daily mean estimated *DNI* values. Generally, the *DNI* datasets generated in this study could be used for the potential solar energy estimations and the proper installations of solar power plants using CST.

To further validate the applicability of the DNI data generated in this study, the DNI data were compared with previous DNI products. The Climate Monitoring Satellite Application Facility (CM-SAF), the JRC Photovoltaic Geographical Information System (PVGIS), the Solar and Wind Energy Resource Assessment (SWERA) supported by the National Renewable Energy Laboratory (NREL) could provide estimates of direct normal irradiance at the earth surface in China. The daily mean DNI records derived from CM-SAF, PVGIS during 2007 and NREL-SWERA during 2002 were compared with the daily mean DNI data generated in this study at 7 CMA radiation stations (KAS, EJN, LAZ, LAS, CHD,

8
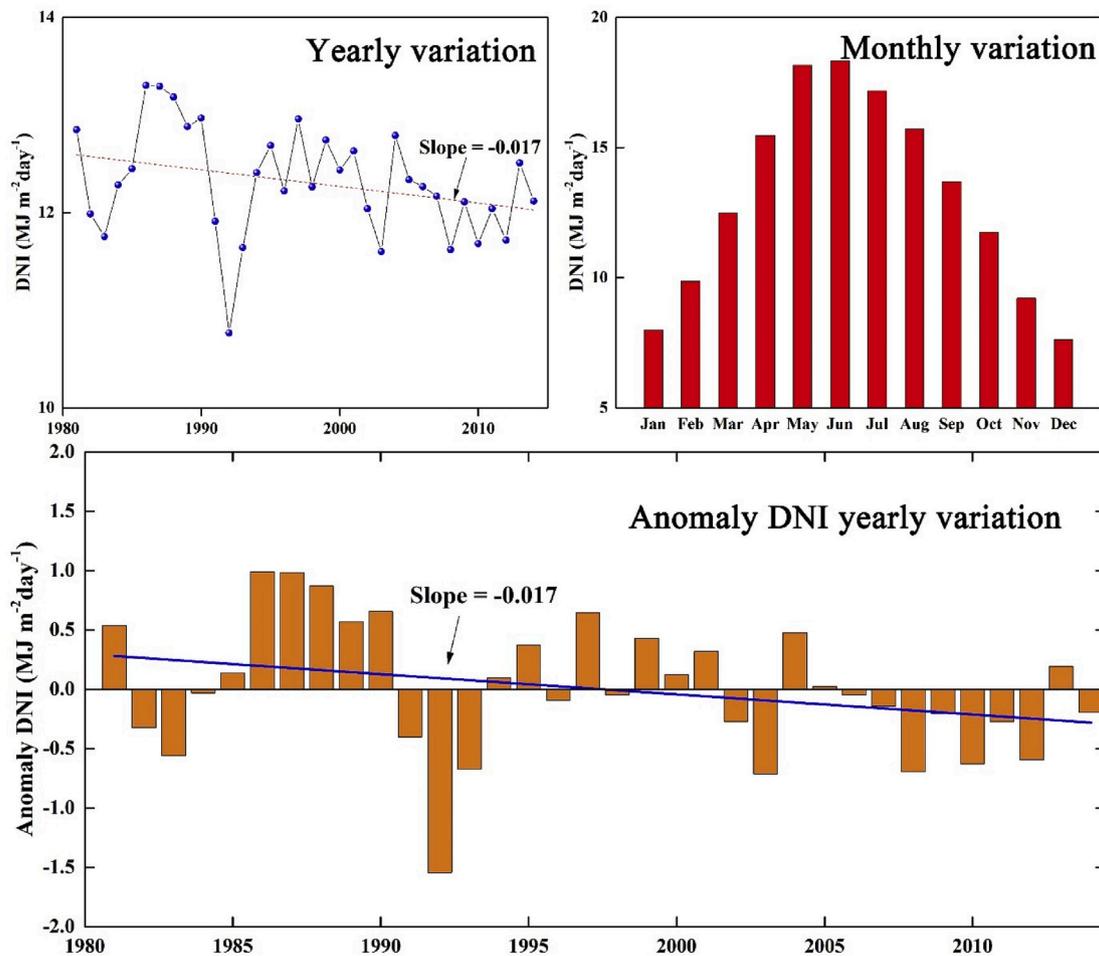
Fig. 12. The annually and seasonal variations of *DNI* throughout China.

KUM and WUH).

Fig. 8 and Fig. 9 showed the validation results of the DNI data by CM-SAF, PV-GIS, NERL-SWERA and the data generated in this study. The result indicated that the DNI data in this study showed better agreement with DNI measurement than that for CM-SAF, PV-GIS and NERL-SWERA. The RMSE, MAE, RMSER, MAER and R for the DNI data in 2007 in this study were 5.177 MJm^{-2}day^{-1}, 3.685 MJm^{-2}day^{-1}, 38.49%, 27.39% and 0.936, respectively. The RMSE, MAE, RMSER, MAER and R for the DNI data in 2002 in this study were 4.817 MJm^{-2}day^{-1}, 3.509 MJm^{-2}day^{-1}, 39.00%, 28.41% and 0.896, respectively.

3.3. Spatial and temporal variations of DNI

By applying REST2_V9.1 model, we constructed a gridded *DNI* dataset (0.50° (lat) *0.625° (lon)) over mainland China that covers the period of 1981–2014. This dataset has been packed into NetCDF files. Each data file is named as MERRA2_direct_normal_irrdiance_YYYY_MM_DD.nc (eg. MERRA2_ direct_normal_irrdiance_2002_01_01. nc), where YYYY is the four-digital year; MM is the two-digital month; DD is the two-digital day. The suffix ".nc" indicates that the data were stored as NetCDF format. More detail information about NetCDF could be found on the official website of UCAR (http://www.unidata.ucar. edu/software/netcdf). This DNI dataset generated in this study could be obtained by contacting the corresponding author of this article. Meanwhile, we will share this DNI dataset on scientific data sharing platform.

Fig. 10 and Fig. 11 showed the spatial variations of the annual mean estimated *DNI (ADNI)* values during 1981–2014 over mainland China.

Generally, *DNI* was higher in Western China than that in Southeastern China, due to the relatively dry air conditions in Western China. The Qinghai Tibetan Plateau has always been an area with the highest *DNI* (21.331 MJm^{-2}day^{-1}) across China due to the weak atmospheric extinction effects there. The Tarim Basin was also an area with high *DNI* values, but the *DNI* there showed strong seasonal variabity because of the seasonal changes of the incoming solar radiations in the atmosphere. In contrast, the Sichuan Basin had always been an area with the lowest *DNI* (3.601 MJm^{-2}day^{-1}), due to the perennial cloudy weather and strong atmospheric extinctions.

Fig. 12 illustrated the annual and monthly variations of *DNI* during 1981–2014 over mianland China. The result showed that the *ADNI* throughout mainland China has been gradually decreased during 1981–2014, which may be due to the growing aerosol radiative forcing effects throughout China in the recent decades [54]. The lowest *ADNI* value was observed in 1992 (10.767 MJm^{-2}day^{-1}), corresponding to the explosion growth of aerosol particles in the air throughout East Asia in 1992 (the Pinatubo Volcano eruption in 1992) [55]. In terms of the monthly variation of *DNI*, the monthly mean *DNI* values always gradually increased from January (7.995 MJm^{-2}day^{-1}) to June (18.337 MJm^{-2}day^{-1}), then gradually decreased from July (17.167 MJm^{-2}day^{-1}) to December (7.610 MJm^{-2}day^{-1}), owing to the monthly variations of the annual cycle of solar zenith and the maximum sunshine duration in China.

Fig. 13 presented the annual mean and monthly mean *DNI* values in different climatic zones. It was obvious that *DNI* were significantly negtively related to the humidity due to the complicated and strong atmospheric extinction processes in the humid zones. The mean *ADNI* values in the humid area, semi-humid area, semi-arid area and arid area
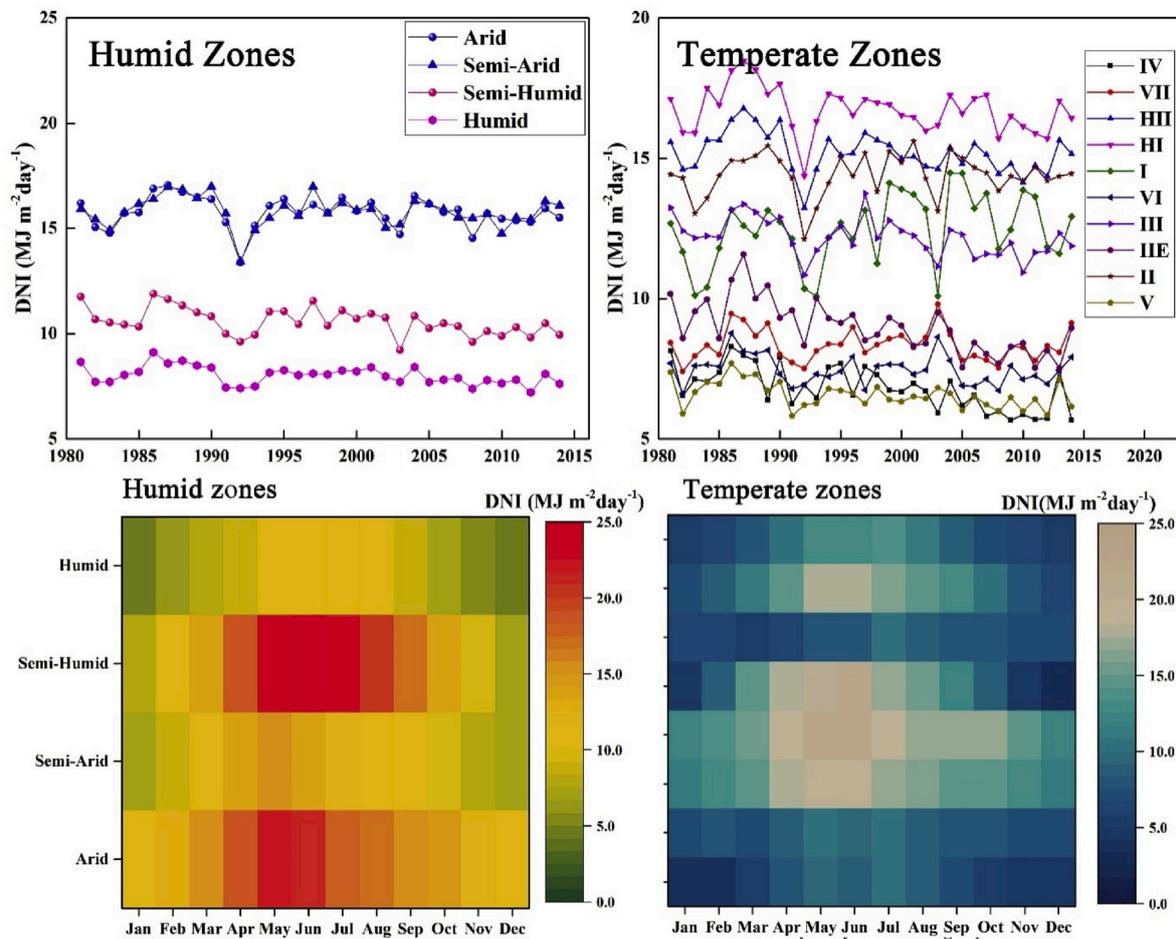
Fig. 13. Yearly and monthly variations of DNI in different climate zones.

were 8.011, 10.567, 15.762 and 15.799 MJm^{-2}day^{-1}, respectively. In terms of the temperate zones, the relatively larger *ADNI* values were found in Plateau areas, which may be due to relatively weak radiation dumping processes there, for example the mean value of *ADNI* for HI and HII were 16.724 MJm^{-2}day^{-1} and 15.197 MJm^{-2}day^{-1}, respectively. In other temperate zones except HI and HII, *ADNI* were approximately proportional to the temperature. The smallest *ADNI* were found in cold temperate zone (6.563 MJm^{-2}day^{-1}), due to the extremely lower temperature than other areas. As for the monthly variations of DNI in climate zones, the monthly mean DNI in arid areas were always higher than that in humid area, but with stronger seasonal variations than humid area. The plateau areas has always been the area with the highest DNI in different humidity and temperate zones throughout the year.

4. Conclusion

The applicability of REST2_v9.1 model in modeling direct normal irradiance using meteorological measurements and MERRA2 reanalysis products in China were validated using DNI measurements at 6 BSRN stations (17,365 samples) and 17 first-class CMA radiation stations (129,662 samples). The long-term gridded DNI datasets during 1980–2014 over mainland China were constructed, and the spatial and temporal variations of DNI in China were investigated.

A quadratic polynomial formulation was developed to correct the cloud effect on direct normal irradiance following the Ångström-Prescott equation. The result indicated that the estimated DNI showed high agreements with DNI measurements at BSRN stations (RMSE = 1.436 MJm^{-2}day^{-1}, MAE = 0.900 MJm^{-2}day^{-1}, RMSER = 22.26%, MAER = 13.95%, R = 0.972) and CMA stations (RMSE = 4.064 MJm^{-2}day^{-1},

MAE = 2.864 MJm^{-2}day^{-1}, RMSER = 34.41%, MAER = 24.26%, R = 0.914). Although subject to some limiting factors, such as the input data quality and the interpolated method, the REST2_V9.1 was proved to be able to generate the gridded DNI datasets for solar energy applications with acceptable accuracy. The estimated DNI values by REST2_V9.1 model showed better agreements with DNI measurement than that of PVGIS, CM-SAF and NERL-SWERA products. Then, a gridded DNI dataset (0.50° (lat) *0.625° (lon)) during 1980–2014 covering mainland China was generated using REST2_V9.1 model and packed into NetCDF files. Finally, the spatial and temporal variations of DNI values over mainland China were discussed using the DNI dataset generated in this study. Generally, the DNI has been gradually decreased which may be caused by the stronger and stronger aerosol radiative forcing effects throughout China in the recent decades. The lowest annual DNI values was found in 1992 (10.767 MJm^{-2}day^{-1}), which may be due to the explosion growth of aerosol particles in the air throughout East Asia in 1992. In terms of the monthly variations, the mean *DNI* were higher in June (18.337 MJm^{-2}day^{-1}) than that in other months following with the seasonal variations of solar zenith angle and sunshine durations. The DNI values in arid areas were generally higher than that in humid areas. The mean *ADNI* values in the humid area, semi-humid area, semi-arid area and arid area were 8.011, 10.567, 15.762 and 15.799 MJm^{-2}day^{-1}, respectively. The Qinghai Tibetan Plateau has always been area with high DNI values (clear sky condtion), while the Sichuan Basin has always been an area with the lowest DIN values (cloudy and rainy sky condition).

Certainly, the REST2_V9.1 model should be further validated in other climate zones around the world. Moreover, as described above, the estimation accuracy of the DNI datasets were subjected to some

objective factors such as the accuracy of the interpolated sunshine duration measurements and the relatively coarse resolution of MERRA2 products. Further work should be conducted to improve the accuracy of the DNI datasets generated in this study.

Credit author statement

Wenmin Qin: Data curation, Writing - original draft preparation, revision. Lunche Wang: Conceptualization, Methodology, Software, Reviewing and Editing, Supervision, Project administration, Funding acquisition, Christian A. Gueymard: Methodology, Software, Supervision, Muhammad Bilal: Methodology, Software. Aiwen Lin: Methodology, Software, Supervision. Wei Jing: Visualization, Investigation, Editing. Ming Zhang: Software, Validation, Xuefang Yang: Data curation.

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.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rser.2020.110004.

References

[1] Sweerts B, Pfenninger S, Yang S, Folini D, van der Zwaan B, Wild M. Estimation of losses in solar energy production from air pollution in China since 1960 using surface radiation data. Nature Energy 2019;4:657–63.

[2] Ekins-Daukes N, Kay M. Brighten the dark skies. Nature Energy 2019;4:633–4.

[5] Jacobsson S, Lauber V. The politics and policy of energy system transformation—explaining the German diffusion of renewable energy technology. Energy Pol 2006;34:256–76.

[6] Lewis NS. Research opportunities to advance solar energy utilization. Science 2016;351:d1920.

[7] Yu HJJ, Popiolek N, Geoffron P. Solar photovoltaic energy policy and globalization: a multiperspective approach with case studies of Germany, Japan, and China. Prog Photovoltaics Res Appl 2016;24:458–76.

[8] Zhang S, He Y. Analysis on the development and policy of solar PV power in China. Renew Sustain Energy Rev 2013;21:393–401.

[9] Prasad AA, Taylor RA, Kay M. Assessment of direct normal irradiance and cloud connections using satellite data over Australia. Appl Energy 2015;143:301–11.

[10] Law EW, Prasad AA, Kay M, Taylor RA. Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting–A review. Sol Energy 2014;108:287–307.

[11] Amadei CA, Allesina G, Tartarini P, Yuting W. Simulation of GEMASOLAR-based solar tower plants for the Chinese energy market: influence of plant downsizing and location change. Renew Energy 2013;55:366–73.

[12] Li Y, Liao S, Rao Z, Liu G. A dynamic assessment based feasibility study of concentrating solar power in China. Renew Energy 2014;69:34–42.

[13] Zhao Z, Chen Y, Thomson JD. Levelized cost of energy modeling for concentrated solar power projects: a China study. Energy 2017;120:117–27.

[14] Vieira De Souza LE, Gilmanova Cavalcante AM. Concentrated Solar Power deployment in emerging economies: the cases of China and Brazil. Renew Sustain Energy Rev 2017;72:1094–103.

[15] Blanc P, Espinar B, Geuder N, Gueymard C, Meyer R, Pitz-Paal R, et al. Direct normal irradiance related definitions and applications: the circumsolar issue. Sol Energy 2014;110:561–77.

[16] Boland J, Ridley B, Brown B. Models of diffuse solar radiation. Renew Energy 2008; 33:575–84.

[17] Gueymard CA, Ruiz-Arias JA. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Sol Energy 2016;128:1–30.

[18] Gueymard CA. REST2: high-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation–Validation with a benchmark dataset. Sol Energy 2008;82:272–85.

[19] Ineichen P. A broadband simplified version of the Solis clear sky model. Sol Energy 2008;82:758–62.

[20] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, et al. McClear : a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Measurement Techniques 2013;6:2403–18.

[21] Sengupta M, Habte A, Gueymard C, Wilbert S, Renne D. Best practices handbook for the collection and use of solar resource data for solar energy applications. Golden, CO (United States): National Renewable Energy Lab.(NREL); 2017.

[22] Qin W, Wang L, Lin A, Zhang M, Xia X, Hu B, et al. Comparison of deterministic and data-driven models for solar radiation estimation in China. Renew Sustain Energy Rev 2018;81:579–94.

[23] Gueymard CA, Ruiz-Arias JA. Validation of direct normal irradiance predictions under arid conditions: a review of radiative models and their turbidity-dependent performance. Renew Sustain Energy Rev 2015;45:379–96.

[24] Jin Z, Yezheng W, Gang Y. General formula for estimation of monthly average daily global solar radiation in China. Energy Convers Manag 2005;46:257–68.

[25] Li R, Zhao L, Wu T, Ding Y, Xin Y, Zou D, et al. Temporal and spatial variations of global solar radiation over the Qinghai–Tibetan Plateau during the past 40 years. Theor Appl Climatol 2013;113:573–83.

[26] Meinel AB, Meinel MP. Applied solar energy: an introduction. NASA STI/Recon Technical Report A 1977;77.

[27] Meza F, Varas E. Estimation of mean monthly solar global radiation as a function of temperature. Agric For Meteorol 2000;100:231–41.

[28] Louche A, Notton G, Poggi P, Simonnot G. Correlations for direct normal and global horizontal irradiation on a French Mediterranean site. Sol Energy 1991;46: 261–6.

[29] Tiris M, Tiris C, Ture IE. Correlations of monthly-average daily global, diffuse and beam radiations with hours of bright sunshine in Gebze, Turkey. Energy Convers Manag 1996;37:1417–21.

[30] Benson RB, Paris MV, Sherry JE, Justus CG. Estimation of daily and monthly direct, diffuse and global solar radiation from sunshine duration measurements. Sol Energy 1984;32:523–35.

[31] Wenxian L, Tao L, Shaoxuan P, Wenfeng G, Enrong L. Estimation and distribution patterns of direct solar radiation in Yunnan Province, China. Energy Convers Manag 1996;37:561–7.

[32] Ruiz-Arias JA, Gueymard CA. Worldwide inter-comparison of clear-sky solar radiation models: consensus-based review of direct and global irradiance components simulated at the earth surface. Sol Energy 2018;168:10–29.

[33] Bird RE. A simple, solar spectral model for direct-normal and diffuse horizontal irradiance. Sol Energy 1984;32:461–71.

[34] Gueymard CA. Direct solar transmittance and irradiance predictions with broadband models. Part I: detailed theoretical performance assessment. Sol Energy 2003;74:355–79.

[35] Iqbal M. An introduction to solar radiation. Elsevier; 2012.

[36] Leckner B. The spectral distribution of solar radiation at the earth's surface—elements of a model. Sol Energy 1978;20:143–50.

[37] Yang K, Huang GW, Tamai N. A hybrid model for estimating global solar radiation. Sol Energy 2001;70:13–22.

[38] Yang K, Koike T. A general model to estimate hourly and daily solar radiation for hydrological studies. 2005. p. 41. Water Resources Research.

[39] Yang K, Koike T, Ye B. Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets. Agric For Meteorol 2006;137:43–55.

[40] Gueymard CA. Parameterized transmittance model for direct beam and circumsolar spectral irradiance. Sol Energy 2001;71:325–46.

[41] Gueymard CA, Lara-Fanego V, Sengupta M, Xie Y. Surface albedo and reflectance: review of definitions, angular and spectral effects, and intercomparison of major data sources in support of advanced solar irradiance modeling over the Americas. Sol Energy 2019;182:194–212.

[42] Chen R, Kang E, Ji X, Yang J, Wang J. An hourly solar radiation model under actual weather and terrain conditions: a case study in Heihe river basin. Energy 2007;32: 1148–57.

[43] Qin J, Tang W, Yang K, Lu N, Niu X, Liang S. An efficient physically based parameterization to derive surface solar irradiance based on satellite atmospheric products. J Geophys Res: Atmosphere 2015;120:4975–88.

[44] Tang W, Qin J, Yang K, Liu S, Lu N, Niu X. Retrieving high-resolution surface solar radiation with cloud parameters derived by combining MODIS and MTSAT data. Atmos Chem Phys 2016;16:2543–57.

[45] Xu X, Du H, Zhou G, Mao F, Li P, Fan W, et al. A method for daily global solar radiation estimation from two instantaneous values using MODIS atmospheric products. Energy 2016;111:117–25.

[46] Zhong X, Ruiz-Arias JA, Kleissl J. Dissecting surface clear sky irradiance bias in numerical weather prediction: application and corrections to the New Goddard Shortwave Scheme. Sol Energy 2016;132:103–13.

[47] Tang W, Qin J, Yang K, Niu X, Zhang X, Yu Y, et al. Reconstruction of daily photosynthetically active radiation and its trends over China. Journal of Geophysical Research Atmospheres 2013;118:13292–302.

[48] Tang W, Yang K, He J, Qin J. Quality control and estimation of global solar radiation in China. Sol Energy 2010;84:466–75.

[49] Tang WJ, Yang K, Qin J, Cheng CCK, He J. Solar radiation trend across China in recent decades: a revisit with quality-controlled data. Atmos Chem Phys 2011;11: 393–406.

[50] Gueymard CA. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: improved validation methodology and detailed performance analysis of 18 broadband radiative models. Sol Energy 2012;86: 2145–69.

[51] Xu T, Hutchinson MF. New developments and applications in the ANUCLIMspatial climatic and bioclimatic modelling package. Environ Model Software 2013;40: 267–79.

[52] Hutchinson MF, Xu T. Anusplin version 4.2 user guide. Centre for resource and environmental studies, the Australian national university, vol. 54; 2004. Canberra.

[53] Inman RH, Edson JG, Coimbra CFM. Impact of local broadband turbidity estimation on forecasting of clear sky direct normal irradiance. Sol Energy 2015; 117:125–38.

[54] Qin W, Wang L, Lin A, Zhang M, Bilal M. Improving the estimation of daily aerosol optical depth and aerosol radiative effect using an optimized artificial neural network. Rem Sens 2018;10:1022.

[55] Qin W, Liu Y, Wang L, Lin A, Xia X, Che H, et al. Characteristic and driving factors of aerosol optical depth over mainland China during 1980–2017. Rem Sens 2018; 10:1064.