Atmospheric Research 254 (2021) 105505

Contents lists available at ScienceDirect

Atmospheric Research

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

The improvement and comparison of diffuse radiation models in different climatic zones of China

Tingting Zhu [a,b], Jun Li [a,*], Liang He [c], Dingrong Wu [d], Xiaojuan Tong [e], Qingyun Mu [a], Qiang Yu [f,g]

[a] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[b] International Institute for Earth System Sciences, Nanjing University, Nanjing 210023, China
[c] National Meteorological Center, Beijing 100081, China
[d] State Key Laboratory of Severe Weather (LASW), Chinese Academy of Meteorological Sciences, Beijing 100081, China
[e] College of Forestry, Beijing Forestry University, Beijing 100083, China
[f] State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A& F University, Yangling 712100, China
[g] School of Life Sciences, University of Technology Sydney, NSW 2007, Australia

ARTICLE INFO

Keywords:
Clearness index
Climatic zones
Cloud cover
Diffuse radiation
Neural network
Sunshine duration

ABSTRACT

Diffuse radiation is vital for climatology, sustainable energy, agriculture and biological activities. It is often estimated based on some meteorological factors but seldom takes total cloud cover into consideration. In this study, cloud cover data are used to establish models, compared with other regular models for estimating diffuse radiation in different climatic zones of China during 1992–2015. The results showed that the performance of the new model based on sunshine fraction (n/N), clearness index (k_t), total cloud cover (CI), air temperature (T_a), relative humidity (RH), wind speed and the day of the year outperformed other models and further improved the predictive accuracy. Generally, the k_d models based on the single k_t was generally more precise compared with single n/N-based or single CI-based models. And models on the basis of multiple factors performed better than the single factor-based model. Comparisons between these models indicated that neural network models (with the largest R and smallest RMSD values) provided better overall accuracy than other models. The parameter of CI as the co-factor can improve the prediction of diffuse radiation. The models developed and evaluated in this study can contribute to developing and utilizing solar energy in China, especially in areas without diffuse radiation records.

1. Introduction

The increasing price of fossil fuel and global demand for energy are focusing the attention of more countries on the development of renewable energy, such as wind and solar energy. Reliable measurements of solar radiation are necessary for predicting the long-term changes and the dynamic processes of radiation resources due to the fact that solar radiation is the main source of life on earth (Beer et al., 2010; Kanniah et al., 2012). Diffuse radiation (H_d) as a component of global solar radiation (H) reaching the surface of the earth is correlated with solar altitude, cloud cover, aerosols, water droplets, and so on. Diffuse radiation has a great effect on the utilization of regional radiation, vegetation photosynthesis, and ecosystem carbon uptake (Kanniah et al., 2013). However, the diffuse component is not readily available for most locations on earth due to the high cost of installing and maintaining the measuring instruments. In China, there are 726 weather stations and 98 global solar radiation sites, but only 17 of them measure diffuse radiation (Wang et al., 2019). Therefore, different methods have been developed to estimate diffuse radiation, such as empirical models and machine learning models.

Based on the availability of meteorological data, different categories of empirical diffuse radiation models have been proposed, such as global solar radiation-based models (González and Calbó, 1999; Dervishi and Mahdavi, 2012; Jacovides et al., 2006; Reindl et al., 1990; Ridley et al., 2010; Torres et al., 2010; Wang et al., 2019), sunshine duration-based models (Fan et al., 2019a, 2019b; Feng et al., 2018; Li et al., 2012), and cloud cover-based models (Furlan et al., 2012; Nunez and Li, 2008;

* Corresponding author.
E-mail address: lijun@igsnrr.ac.cn (J. Li).

https://doi.org/10.1016/j.atmosres.2021.105505
Received 7 September 2020; Received in revised form 25 January 2021; Accepted 31 January 2021
Available online 2 February 2021

Nomenclature

H	Global solar radiation (MJ m^{-2} day^{-1})
H_d	Diffuse solar radiation (MJ m^{-2} day^{-1})
H_0	Extra-terrestrial solar radiation (MJ m^{-2} day^{-1})
I_0	Solar constant (1367 W m^{-2})
n	Daily sunshine duration hours (h)
N	Daily maximum possible sunshine hours (h)
CI	Total cloud cover
k_t	Clearness index
k_d	Diffuse fraction
RH	Daily average relative humidity (%)
T_a	Daily average air temperature (°C)
win	Wind speed (m s^{-1})

Greek letters

d_r	Correction factor of the Earth's orbit
ω_s	Sunset hour angle (°)
δ	Solar declination angle (°)
φ	Latitude of the location (°)

Abbreviations

ANN	Artificial neural networks
GABP	Genetic algorithm-optimized neutral networks
GPI	Global performance index
MBE	Mean bias error (MJ m^{-2} day^{-1})
MPZ	Mountain plateau climate zone
NRMSE	Normalized root mean squared error (%)
R	Correlation coefficient
RMSE	Root mean square error (MJ m^{-2} day^{-1})
SMZ	Subtropical monsoon climatic zone
TCZ	Temperate continental climatic zone
TMZ	Temperate monsoon climatic zone
TMPZ	Tropical monsoon climatic zone

Stephens, 2005), in conjunction with the use of other meteorological variables such as temperature (T_a) and relative humidity (RH). Early studies reported that the diffuse fraction (k_d) decreased with increasing clearness index (k_t) (Liu and Jordan, 1960). Orgill and Hollands (1977), Erbs et al. (1982) and Reindl et al. (1990)obtained a relationship between k_d and k_t by the piecewise linear method. Ridley et al. (2010) estimated H_d based on k_t, solar altitude, and the apparent time. The result showed that the performance of the hourly-scale diffuse radiation model was very good in Europe, Africa, and Asia, and the model was suggested to be applicable globally. Tapakis et al. (2016) used three methods to predict H_d which were based on k_t and solar altitude. They found that H_d was significantly affected by solar altitude. Wang et al. (2019) compared 96 models (employing 11 input parameters) at the 17 diffuse radiation stations of China. They found that a more accurate model performance came with a greater number of input parameters.

However, even though the k_t-based method was extensively studied and showed good performance, its application was limited by the general unavailability of global solar radiation data. Therefore, several models were developed for estimating diffuse solar radiation based on sunshine fraction (n/N). Iqbal (1979) developed a model that correlated k_d as a function of n/N. Li et al. (2012) calculated diffuse solar radiation from n/N, T_a, and RH, and validated the method in China. Feng et al. (2018) evaluated 15 typical sunshine-based models for predicting diffuse solar radiation in different climatic zones of China and found that the second-order polynomial model was performed better than other models. Sunshine-based empirical models are simple and easy to use in practice, but they are not necessarily able to estimate H_d accurately under climatic conditions different from those under which they were developed.

Cloud-cover index-based models have been studied less than the above two types of models because cloud-cover data have been difficult to obtain and relative information has been incomplete. It is indisputable that water droplets in clouds and scattering aerosols from pollutants have had a great effect on diffuse radiation(Furlan et al., 2012). Therefore, cloud cover and cloud types should be considered as the main input parameters for estimating diffuse radiation. However, some studies have indicated that variations in diffuse solar radiation cannot be explained by k_t, n/N, or cloud cover alone, and that a combination of these parameters should be used to estimate diffuse solar radiation (Fan et al., 2019b; Jamil and Akhtar, 2017). Meteorological data (e.g., T_a and RH) have also been incorporated to further improve the estimation accuracy of diffuse solar radiation. Furlan et al. (2012) developed a new model based on k_t and included the effects of clouds (cloud cover and cloud type), traditional meteorological variables (T_a, RH, atmospheric pressure observed at the surface), and air pollution (concentration of particulate matter observed at the surface). The results showed that the new model was capable of predicting hourly values of diffuse solar radiation better than previously developed models.

Currently, artificial intelligence methodologies, such as kernel-based algorithms, tree-based ensemble models, and artificial neural networks (ANN), have been popular and applied widely in various fields of scientific research due to the negligible correlation between input parameters (Behrang et al., 2010). Alam et al. (2009) developed an ANN model to estimate monthly mean hourly and daily diffuse solar radiation. They concluded that the ANN model was more accurate and versatile than other models. Shamshirband et al. (2016) developed a coupled model for estimating daily horizontal diffuse radiation by integrating support vector machine with wavelet transform algorithm (SVM-WT) and found that estimated diffuse solar radiation values by SVM–WT were in favorable agreement with measured data. Feng et al. (2017) compared arificial intelligence models with empirical models and found that the extreme learning machine, genetic algorithm-optimized neutral networks, random forest, and generalized regression neural networks all performed much better than the empirical models. Xue (2017) developed genetic algorithm and particle swarm optimization to improve the efficiency and generalization ability of a back-propagation neural network model for predicting daily diffuse solar radiation. Fan et al. (2019b) compared empirical and machine learning models for predicting solar radiation in China and concluded that the machine learning models were superior to the empirical models.

The objective of this study was to develop, compare and evaluate empirical models, mainly based on k_t, n/N, total cloud cover (CI) and neural network models for predicting daily diffuse irradiation in different climatic zones of China (including tropical, subtropical, warm-temperate, mid-temperate, cold-temperate, and Qinghai-Tibet Plateau climatic zones) using long-term daily surface observations of meteorological data and radiation data during 1992–2015. The results may be helpful in selecting the most appropriate diffuse radiation models for solar energy applications.

2. Materials and methods

2.1. Sites and data processing

Due to its vast territory, China is encountering cold temperate, mid-temperate, warm temperate, subtropical and tropics zones from north to south. Among them, the Qinghai-Tibet Plateau has an average altitude of over 4000 m. The typical mountain plateau zone (MPZ, including Geermu and Lhasa) is characterized by cold winters and large temperature differences. The northwestern region is characterized by low

precipitation and has an arid climate in the temperate continental zone (TCZ, including Urumqi, Ejinaqi, Kashi, and Lanzhou,). The Northeast Plain and the Huanghuaihai Plain belong to the temperate monsoon climatic zone (TMZ, including Mohe, Harbin, Shenyang, Beijing, and Zhengzhou), and have hot and rainy summers and cold and dry winters. Plains of the Middle and Lower Reaches of the Yangtze River, the Southeast Hills, the Sichuan Basin, and the Yunnan-Guizhou Plateau belong to the subtropical monsoon climatic zone (SMZ) and rainless winters. Sanya is in the tropical monsoon climatic zone (TMPZ), having no winters throughout the year, high temperatures, and frequent rains. The distribution of the diffuse radiation stations and climatic zones are shown in Fig. 1. Relevant details for the 17 radiation stations studied in this paper are listed in Table 1.

Continuous long-term time series of daily meteorological data, total cloud cover data, and diffuse radiation data on a horizontal surface during 1992–2015 were collected from 17 diffuse radiation stations across different climatic zones of China. These data were obtained from the National Meteorological Information Center (NMIC) of the China Meteorological Administration (CMA), which has checked the data quality. In every climatic zone, one station (we choose Lhasa, Urumqi, Beijing and Wuhan in MPZ, TCZ, TMZ and SMZ, due to their good data, respectively.) was used as training data and the other stations were used for model cross-validation (Table 1). In theory, $n/N \leq 1$, $H/H_0 \leq 1$, $H_d/H \leq 1$. These data were further filtered based on the following rules: when $k_t < 0.2$, then $k_d<0.8$; when $k_t>0.6$, then $k_d>0.8$ (Reindl et al., 1990; Ridley et al., 2010; Tapakis et al., 2016). In order to avoid seasonal bias, the years of observed data for all stations must be continuous.

Table 1
Geographical locations for the 17 diffuse radiation stations in China. It is noted that the first station is used as the training data and other stations are used for cross-validation in every climatic zone.

Station name	Latitude (N)	Longitude (E)	Altitude (m)	Climatic zone	Time period
Lhasa	29.67	91.13	3648.70	MPZ	1992–2015
Geermu	36.42	94.92	2807.60	MPZ	1992–2013
Urumqi	43.78	87.65	935.00	TCZ	1992–2015
Kashi	39.47	75.98	1288.70	TCZ	1992–2013
Ejinaqi	41.95	101.07	940.50	TCZ	1992–2013
Lanzhou	35.87	104.15	1874.4	TCZ	2005–2013
Beijing	39.80	116.47	31.30	TMZ	1992–2015
Mohe	53.47	122.52	433.00	TMZ	1997–2013
Harbin	45.75	126.77	142.30	TMZ	1992–2015
Shengyang	41.73	123.45	44.70	TMZ	1992–2015
Zhengzhou	34.72	113.65	110.40	TMZ	1992–2015
Wuhan	30.62	114.13	23.10	SMZ	1992–2015
Chengdu	30.75	103.87	547.7	SMZ	2004–2013
Kunming	25.02	102.68	1892.40	SMZ	1992–2013
Shanghai	31.40	121.48	6.00	SMZ	1992–2015
Guangzhou	23.17	113.33	41.00	SMZ	1992–2015
Sanya	18.23	109.52	5.90	TMPZ	1992–2015

2.2. Calculating solar radiation

k_t is the ratio of global solar radiation (H) to the extraterrestrial solar radiation (H_0). The calculation of k_t is given as follows:

$$k_t = \mathrm{H}/H_0 \tag{1}$$

$$H_0 = \frac{24}{\pi} I_0 d_r(\omega_s sin\varphi sin\delta + cos\varphi cos\delta sin\omega_s) \tag{2}$$

$$\delta = 23.45 sin\left(360\frac{J+284}{365}\right) \tag{3}$$

$$\omega_s = arcos\left(-\frac{tan\varphi}{tan\delta}\right) \tag{4}$$

where I_0 is the solar constant (1367 W m^{-2}), d_r represents correction factor of the Earth's orbit, ω_s is the sunset hour angle, δ is solar declination, φ is the latitude, J is day of the year starting from January 1st.

2.3. Diffuse radiation empirical models

H_d is estimated in terms of k_d; k_d is the ratio of H_d to H. k_d can be calculated on the basis of a single variable (k_t or n/N), and the expression can be linear, quadratic, cubic, exponential, inverse proportional, logarithmic, and so on. However, the use of a single variable results in

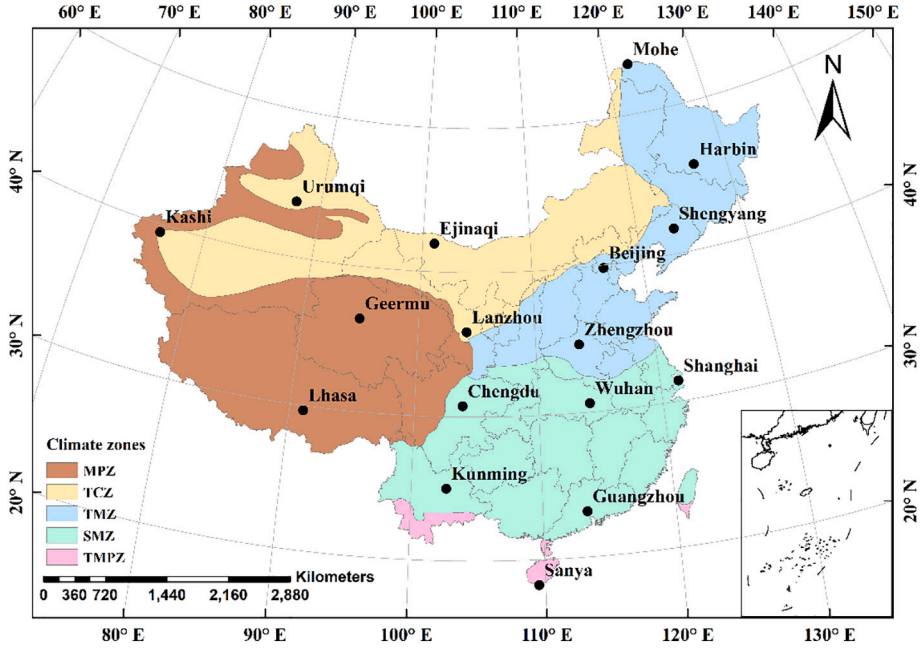

Fig. 1. Distribution of the 5 climatic zones across China and the geographical locations of the 17 diffuse radiation stations used in this study. MPZ, mountain plateau climatic zone; TCZ, temperate continental climatic zone; TMZ, temperate monsoon climatic zone; SMZ, subtropical monsoon climatic zone; TMPZ, tropical monsoon climatic zone.

poorer prediction of the diffuse radiation than when multiple variables (such as T_a, RH, and CI) are taken into consideration. All of the models used for predicting diffuse solar radiation were classified into several categories in this study. Detailed classification information for the regression models is shown in Table 2.

2.4. Artificial neural network (ANN)

The neural network model is a computational program model that imitates biologic neural networks. Due to the high efficiency and accuracy of the results, it is widely used to solve different large-scale complex problems. The neural network model mainly includes an input layer, a hidden layer, and an output layer. The basic structure includes neurons and nodes. The function of the input layer is to input variables. The hidden layer processes the input data by allocating weights to obtain an output result. In this paper, the temperature, relative humidity, wind speed, sunshine duration, clearness index, total cloud cover, and the day of the year are used as the input data for the input layer. The basic structure of the neural network is shown in Fig. 2, and is expressed as follows (Deo and Şahin, 2017):

Table 2
Different categories of existing and developed empirical models used in this study for estimating daily diffuse solar radiation. k_d, diffuse fraction; k_t,clearness index; n/N,sunshine fraction; CI, total cloud cover; T_a, air temperature; RH, relative humidity; J, the day of the year; win, wind speed; a-j, equation coefficients.

Model ID	Equation	References
Category I: $k_d = f(k_t)$		
M1	$k_d = ak_t + b$	Lebaron and Dirmhirn (1983)
M2	$k_d = ak_t^2 + bk_t + c$	Jacovides et al. (2006)
M3	$k_d = ak_t^3 + bk_t^2 + ck_t + d$	Jacovides et al. (2006)
M4	$k_d = 1/(1 + e^{ak_t+b})$	Boland et al. (2001)
M5	$k_d = ak_t + bT_a + cRH + d\sin(2\pi J/365 + e) + fwin + g$	Cao et al. (2017)
Category II: $k_d = f(n/N)$		
M6	$k_d = an/N + b$	Jamil and Akhtar (2017)
M7	$k_d = an/N^2 + bn/N + c$	Bailek et al. (2018)
M8	$k_d = an/N^3 + bn/N^2 + cn/N + d$	Feng et al., 2018
M9	$k_d = 1/(1 + e^{an/N+b})$	This study
M10	$k_d = an/N + bT_a + cRH + d\sin(2\pi J/365 + e) + fwin + g$	This study
Category III: $k_d = f(CI)$		
M11	$k_d = aCI + b$	This study
M12	$k_d = aCI^2 + bCI + c$	This study
M13	$k_d = aCI^3 + bCI^2 + cCI + d$	This study
M14	$k_d = 1/(1 + e^{aCI+b})$	This study
M15	$k_d = aCI + bT_a + cRH + d\sin(2\pi J/365 + e) + fwin + g$	This study
Category IV: $k_d = f(k_t, n/N, CI)$		
M16	$k_d = ak_t + bn/N + c$	Jamil and Akhtar (2017)
M17	$k_d = ak_t + bn/N + cCI + d$	This study
M18	$k_d = ak_t + bk_t^2 + cn/N + dn/N^2 + e$	Jamil and Akhtar (2017)
M19	$k_d = ak_t + bk_t^2 + cn/N + dn/N^2 + eCI + fCI^2 + g$	This study
M20	$k_d = ak_t + bn/N + cRH + dT_a + e\sin(2\pi Jf/365 + g) + hwin + i$	This study
M21	$k_d = ak_t + bn/N + cCI + dRH + eT_a + f\sin(2\pi Jg/365 + h) + iwin + j$	This study
Category V: k_d based on ANN		
M22	GABP based on the input parameters of M20	This study
M23	GABP based on the input parameters of M21	This study

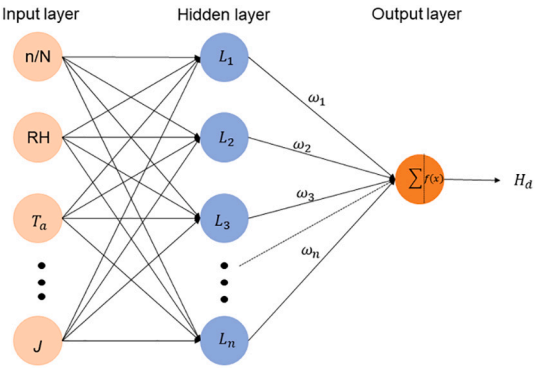

Fig. 2. Artificial neural network architecture to predict diffuse solar radiation. n/N, sunshine fraction; T_a, air temperature; RH, relative humidity; J, the day of the year; H_d,diffuse radiation; L_i,hidden layer at i^{th} layer; ω_i, the weight; $f(x)$, the hidden transfer function.

$$y(x) = F\left[\sum_{i=1}^{n} w_i(t)x_i(t) + b\right] \tag{5}$$

where $y(x)$ is an output variable, $x_i(t)$ is the input variables in discrete time space t, F is the hidden transfer function, n is the hidden neurons, $w_i(t)$ is the weight, and b is the bias.

The back propagation neural network (BPNN) is the most widely used ANN training method. BP training is a gradient descent algorithm. Its main features are signal forward transmission and error backward propagation. By constantly adjusting the network weights, the value is such that the final output of the network is as close as possible to the desired output for the training purposes. In order to improve the accuracy and application of the BP neural network model prediction, this paper uses a genetic algorithm (GA) to optimize BPNN to estimate diffuse radiation (Fan et al., 2019b).

2.5. Comparison of models and statistical error analysis

The accuracy and performance of the studied models to estimate daily diffuse solar radiation were evaluated and compared using six commonly used statistical indicators including the root mean square error (RMSE) (Eq. (6)), the mean bias error (MBE) (Eq. (7)), normalized root mean squared error (NRMSE) (Eq. (8)), the determination coefficient (R^2) (Eq. (9)), and the centered root mean square difference (RMSD) (Eq. (10)). The equations for calculating these statistical indicators are:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(H_{ci} - H_{mi})^2}{n}} \tag{6}$$

$$\mathrm{MBE} = \frac{\sum_{i}^{n}(H_{ci} - H_{mi})}{n} \tag{7}$$

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(H_{ci} - H_{mi})^2}{n}}}{H_{ma}} * 100\% \tag{8}$$

$$R^2 = \frac{\left(\sum_{i=1}^{n}(H_{ci} - H_{ca})(H_{mi} - H_{ma})\right)^2}{\sum_{i=1}^{n}(H_{ci} - H_{ca})^2 \sum_{i=1}^{n}(H_{mi} - H_{ma})^2} \tag{9}$$

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{n}\left((H_{ci} - H_{ca}) - (H_{mi} - H_{ma})\right)^2}{n}} \quad (10)$$

where H_{ci} and H_{ca} indicate the individual and average values estimated by the models, respectively. H_{mi} and H_{ma} indicate the individual and average measured values obtained by observation, respectively. n indicates the number of data points.

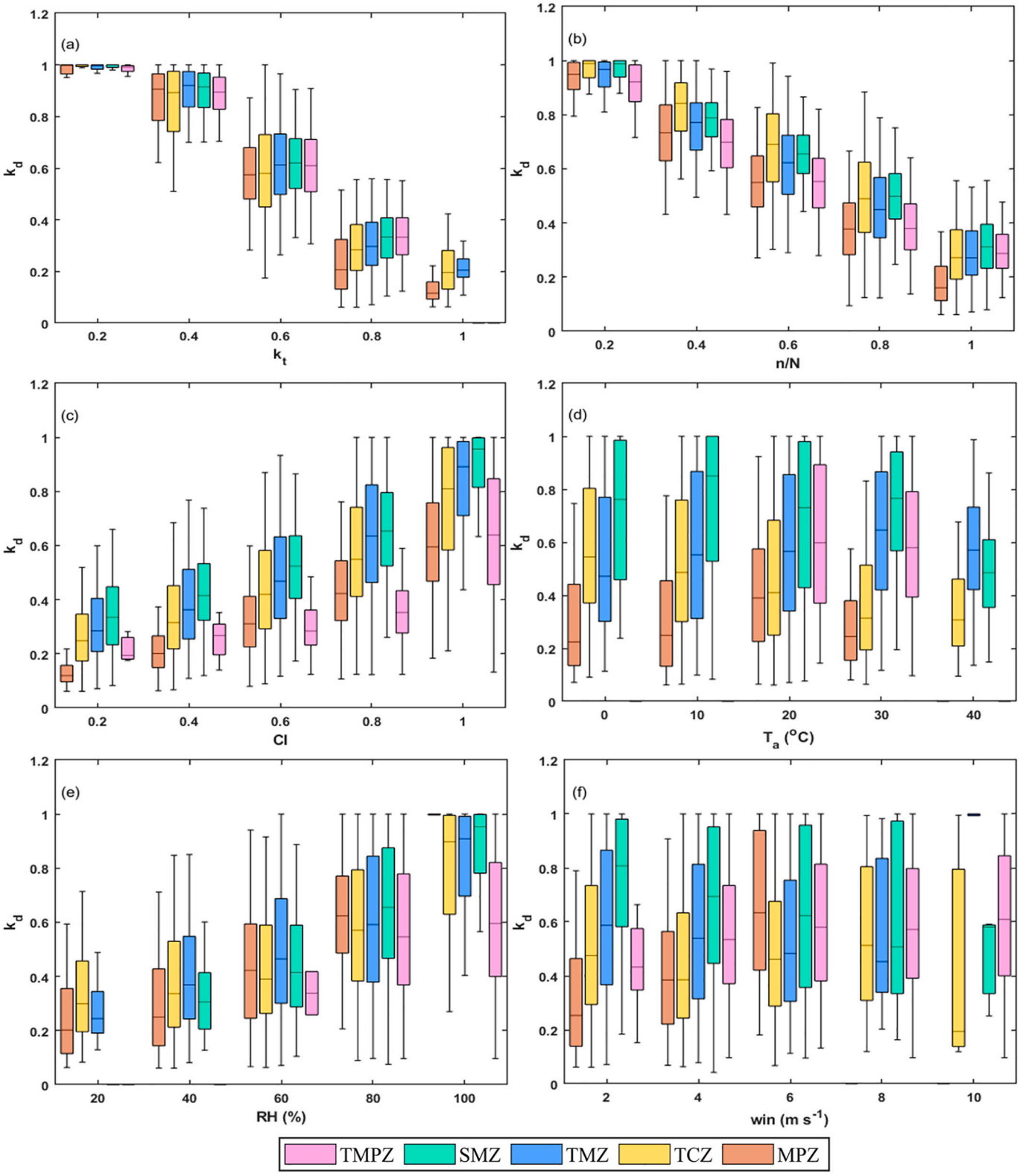

Fig. 3. Boxplot between diffuse fraction (k_d) and meteorological parameters in five climatic zones of China. Some zones have no values in relative interval. (a)-(c) for 0.2 bin; (d) for 10 °C bin; (e) for 20% bin; (f) for 2 m s^{-1} bin. The lower and upper box boundaries of the box plots represent the 25th and 75th percentiles; the bar in the box represents medium value; the lower and upper whiskers represent the 5th and 95th percentiles.

However, the above statistical measures are not fully suitable for revealing the accuracies of different models using different statistical factors due to the huge volume of data, and thus it was difficult to determine which model performed better. Therefore, the global performance indicator (GPI) was proposed by Despotovic et al. (2015).

$$GPI_j = \sum_{i=1}^{5} \alpha_i (y_i - y_{ij}) \quad (11)$$

where y_i is the median of scaled values of the i_{th} statistical indicator, y_{ij} is the scaled value of the i_{th} statistical indicator for the j_{th} model, α_i is equal to -1 for $j = 4$ (R), and equal to 1 for the other statistical indicators. Higher GPI values indicate better model performance.

In addition, Taylor diagrams have been useful in combining the relative skills of many different models, and have been widely applied in various fields for model comparisons. The correlation coefficient (R), standard deviation (STD), and centered root mean square difference (RMSD) between measured and estimated values were combined at a single point in a two-dimensional polar diagram, which graphically displayed how close the estimated results of a particular model were to the observed data. In order to compare the performances of different models, all models were evaluated by graphical visualization based on these statistical factors.

3. Results and discussion

3.1. The effect of meteorological factors on diffuse radiation fraction

Diffuse solar radiation is not only closely related to sunshine duration and global solar radiation, but also affected by other meteorological factors, such as total cloud cover, temperature, and relative humidity. Fig. 3 shows that k_d decreased with increasing k_t and n/N, and increased with increasing CI and RH. In SMZ and TMPZ, k_d was greater than 0.2 when k_t was below 0.8, which showed that the rainy weather for these two regions through the entire year may have resulted in the larger diffuse radiation. Values of k_d were generally greater in TCZ than in other zones at a given level of sunshine fraction, probably due to the more arid conditions in TCZ. A significant positive relationship was observed between k_d and CI because the increased number of water droplets in polluted clouds and the spread of scattering aerosols affected diffuse solar radiation (Feng et al., 2018). T_a had a slightly negative effect on diffuse radiation in TCZ, which may be caused by the much lower air temperature in winter (Fan et al., 2019a). And RH positively affected diffuse radiation in all five climate zones, k_d had weak significant relationship with wind speed (Fig. 3,Table 3). It was apparent that models that only depended on k_t or n/N would be unable to predict daily values of k_d with great accuracy, and that more information about sunshine duration, cloud cover, temperature, and relative humidity would be necessary to improve model accuracy. However, even though cloud is one of the most frequent and important atmospheric phenomena, acting as a critical factor in the climate system and in climate change processes (Stephens, 2005), cloud information available as easily as other traditional meteorological variables is not often considered in models. This fact makes it difficult to explicitly incorporate cloud data in empirical models (Feng et al., 2018). In this study, the cloud cover data were obtained by CMA and were incorporated into diffuse radiation models to improve the simulations of diffuse radiation.

Table 3
Correlation coefficient between k_d and k_t, n/N, CI, T_a, RH. All correlation coefficients show $P < 0.01$.

Correlation coefficient	k_t	n/N	Cl	T_a	RH	win
k_d_MPZ	−0.911	−0.878	0.790	0.149	0.461	0.280
k_d_TCZ	−0.884	−0.853	0.712	−0.222	0.432	−0.060
k_d_TMZ	−0.901	−0.902	0.751	0.132	0.466	−0.063
k_d_SMZ	−0.923	−0.933	0.818	−0.031	0.600	−0.097
k_d_TMPZ	−0.906	−0.884	0.57	−0.253	0.235	0.176

3.2. Model performance

The accuracy and performance of 23 models for modeling daily H_d were compared. Seven meteorological parameters from 17 sites during 1992–2015 were used as model input elements for estimating diffuse solar radiation in China. The regression coefficients for the 23 models from the 5 stations in order to cross validation are given in Supplemental Table S1. Figs. 4(a-p) is a Taylor diagram for visualizing the model accuracies at the 17 radiation stations.

The results indicated that the performance of Category V was the best in all sites, M23 was better than M22. In most sites, the performance of Category IV was second, and then M21 performed best. The performance of Category I was better than Category II except for Mohe, Harbin and Guangzhou stations. In addition, the simulation of Category III was the worst in all sites. For example, M23 performed the best in Mohe and Harbin (Fig. 4(a) and (b)), while the M15 and M2 model were the worst, respectively (R values were 0.768 and 0.764 and RMSD values were 3.192 MJ m^{-2} day^{-1} and 2.426 MJ m^{-2} day^{-1}, respectively). At Eijinaqi, the best performing models were M23, M22, and M19 (in declining order), and the worst model was M15. M23, M22 and M21 produced the best accuracy at Shengyang, with R and RMSD ranges of 0.91–0.94 and 1.26–1.42, respectively. M14 and M15 had the smallest R and largest RMSD at Shenyang. Similarly, the fitting results of M22 and M23 were the closest to the measured values at Zhengzhou. However, M12, M13 and M14 performed the worst at Zhengzhou. The fitting results of M23 and M22 were the closest to the measured values a Wuhan, Shanghai, and Guangzhou. However, M14 performed the worst at Shanghai, M15 at Wuhan, and M14 at Guangzhou.

In order to determine the impacts of the different functional forms and the numbers of input-parameters on the accuracy of the models based on Figs.4, the differences in the models from each category were analyzed and compared. It was found that models based on higher order function performed better than the simple linear function. Generally, the k_d models based on the single k_t was generally more precise compared with single n/N-based or single CI-based models. And models on the basis of multiple factors performed better than the single factor-based model. Comparisons between these models indicated that neural network models (with the largest R and smallest RMSD values) provided better overall accuracy than other models, while the performances of Category III was worse than other models.

3.3. Comparison of five categories of models in different climatic zones

To determine which factors are dominant when estimating H_d, the average statistical indicators (MBE, NRMSE, $R^2$and GPI) for 5 categories models are presented in Figs. 5(a-d) and Supplemental Table S2. In terms of MBE, category I and II obtained the smallest values in TMPZ, category III and V obtained the smallest value in MPZ and category IV in TMZ. In terms of NRMSE, the simulation of 5 categories in SMZ obtained the lowest values but the largest in TCZ. The reason was that the frequent dust occurred in the air in TCZ. The R^2 values performed better in SMZ and TMZ but worse in TMPZ. More total cloud cover was in the east of China and multiple factors was sensitive to diffuse radiation changes. According to the GPI, category I obtained the biggest values in TMZ, category II obtained the highest value in TCZ, category III and IV obtained the highest values in MPZ, category V obtained the greatest value in SMZ. In terms of categories I, II, III model, k_t had a more significant effect than n/N and cloud on estimation of daily diffuse radiation. In addition, the statistical results indicated that the model categories IV and V provided better accuracy for predicting the daily diffuse solar radiation compared with other model categories in term of NRMSE, R^2 and GPI, while the lower prediction accuracy of category III in each zone was calculated. In all zones, the performance was ranked as

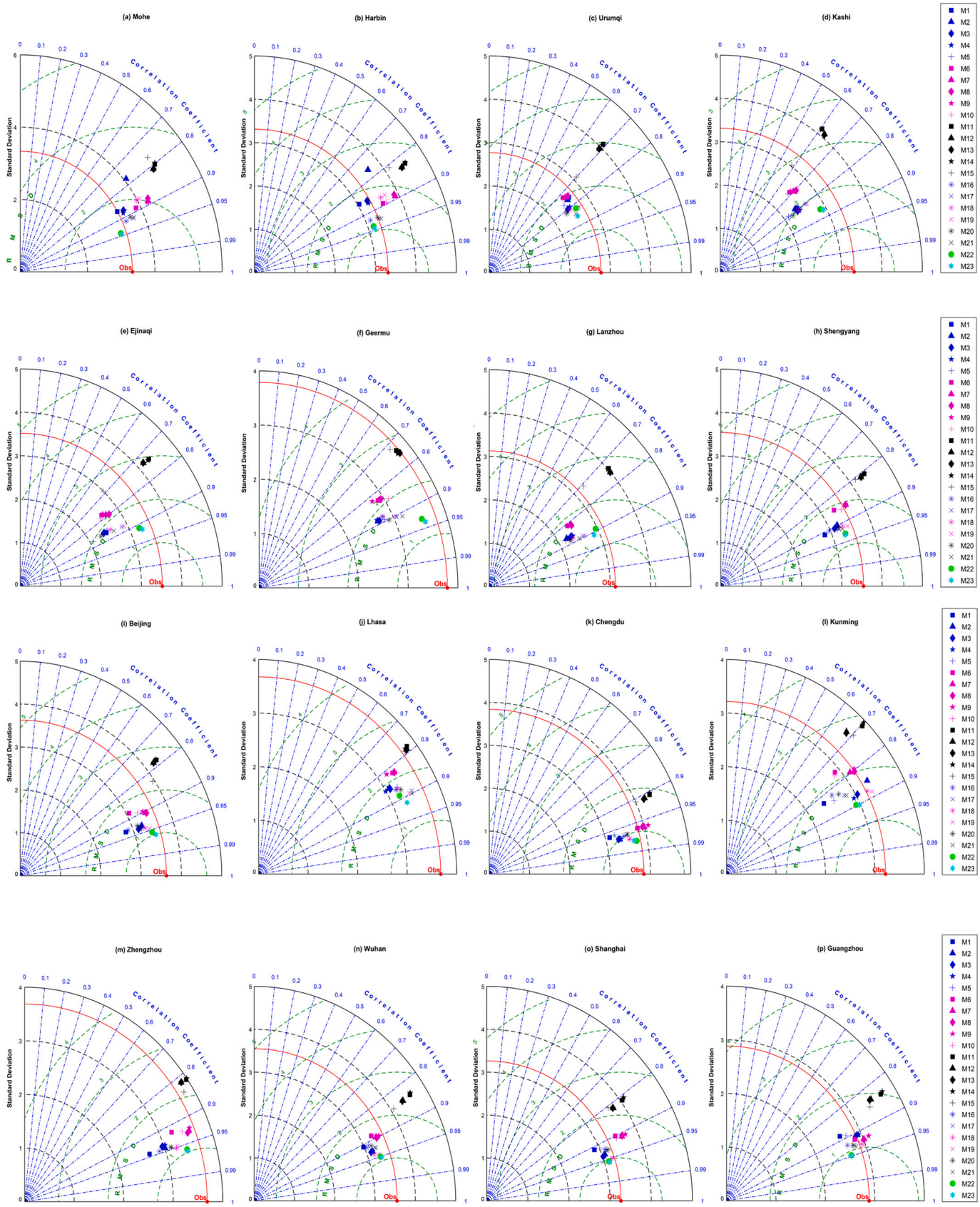

Fig. 4. Taylor diagrams for models applied at the 17 stations in China. Figure (c), (i), (j), (n), (q) were the results of training model and the remaining were of cross-validation.

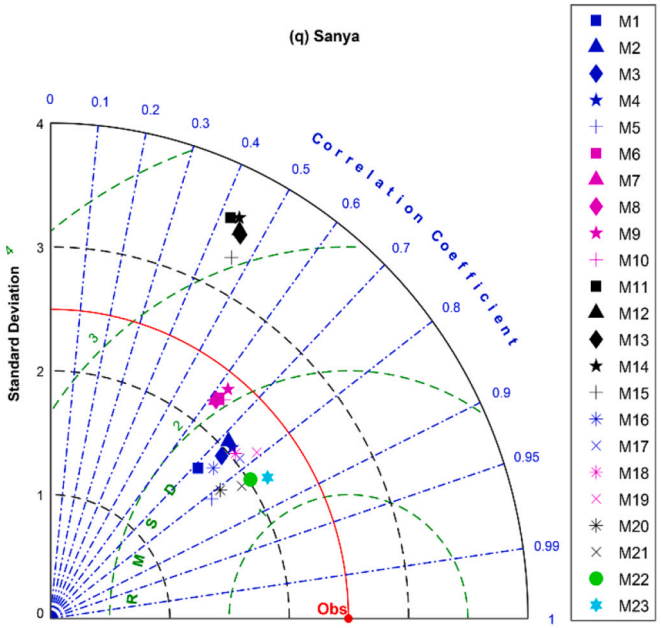

Fig. 4. (*continued*).

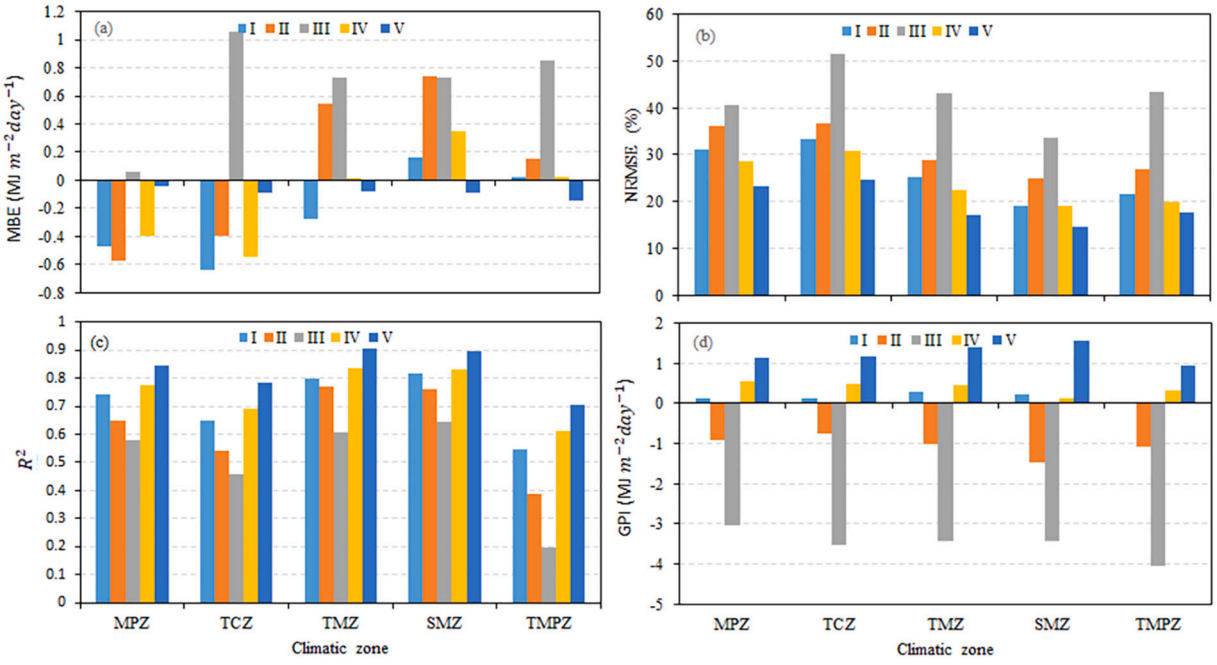

Fig. 5. The mean statistical indicators of 5 categories of diffuse radiation models in the 5 climatic zones of China: (a) MBE; (b) NRMSE; (c) R^2; (d) GPI.

followed: V > IV > I. > II > III. Thus, diffuse solar radiation is generally correlated with multiple meteorological variables. It is reasonable that a combination of a greater number of influencing factors results in better prediction of diffuse solar radiation.

3.4. The effect of cloud on diffuse radiation in different climatic zones

Considering the above analysis that category V (M22 and M23, called ANN models) of diffuse radiation model performed best, In the section, the two models were taken for example to illustrate the effect of cloud on diffuse radiation. It was obvious in Fig. 6 and Table 4 that the simulation result of diffuse radiation was close to the measured diffuse radiation. When CI was taken into consideration, the R^2 of MPZ, TCZ, TMZ, SMZ and TMPZ were 0.859, 0.799 0.908, 0.896 and 0.718, respectively. And their slopes were also close to 1. R^2 and slopes were higher significantly than those without the parameter of CI, especially in MPZ and TCZ. The estimation of diffuse radiation in TMZ and SMZ outperformed other climatic zones, due to their higher $R^2$and slope. In terms of the lower slope and the smaller R^2 in TMPZ, it can be attributed that only one station was analyzed in TMPZ. The overall statistical performance of M23 in all zones was better than M22. As shown in Fig. 7, there was a significant correlation between observation and simulation based on M23 with a slope of 0.998 across different zones. Among zones, the largest diffuse radiation was found in TMPZ (7.6 ± 3 MJ m^{-2} day^{-1}), and the smallest (6.1 ± 3 MJ m^{-2} day^{-1}) in TCZ. Therefore, cloud data as the co-factor can improve the accuracy of diffuse radiation model.

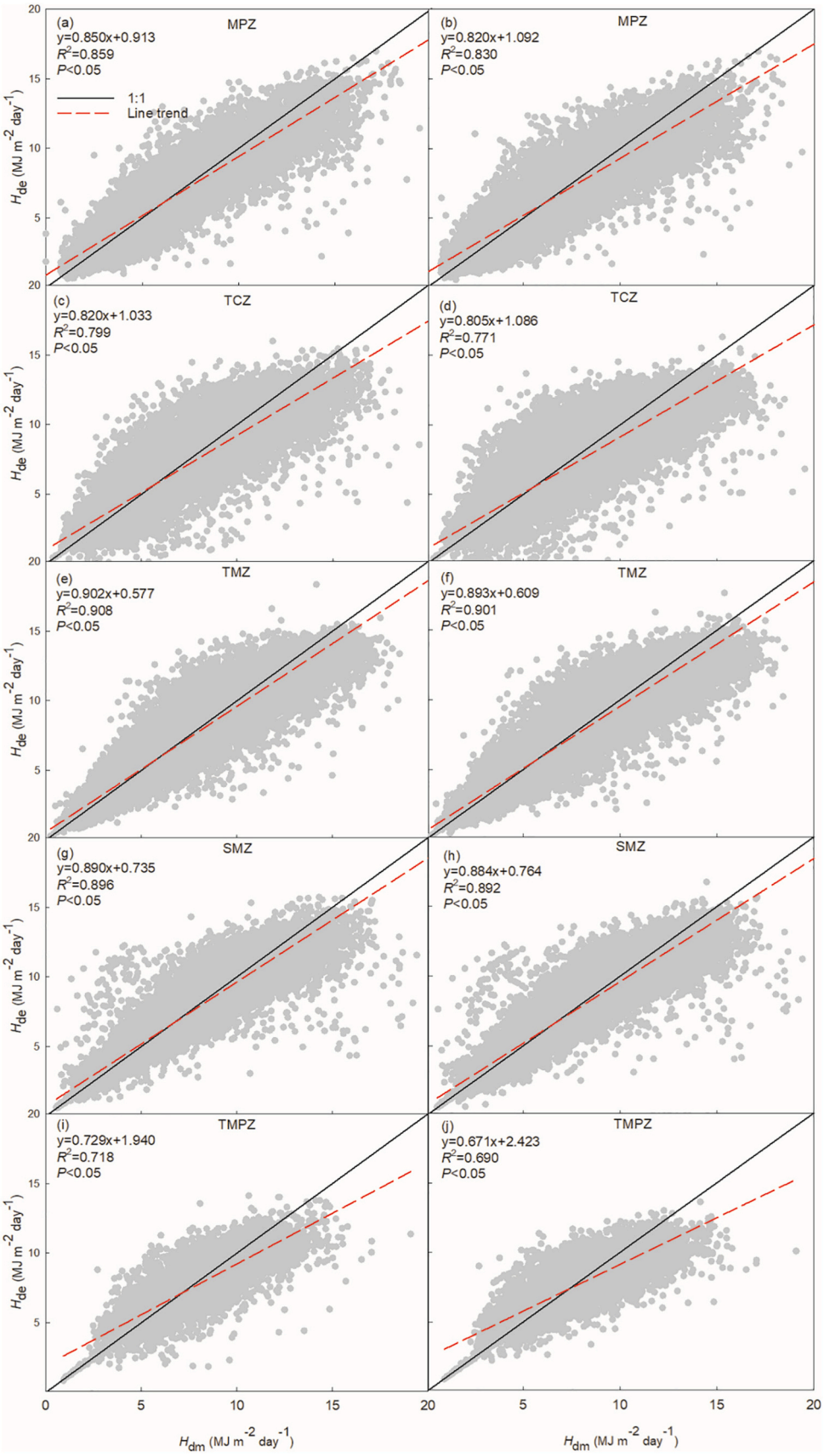

Fig. 6. Comparison between measured diffuse radiation (H_{dm}) vs. estimated values (H_{de}) based on M22 and M23. On the left panel, H_{de} was estimated based on M23 with cloud data. On the right panel, H_{de} was obtained based on M22 without cloud data.

Table 4
Statistical performance from the model category V in the 5 climatic zones of China.

Climatic Zone	Model ID	MBE (MJ m^{-2} day^{-1})	RMSE (MJ m^{-2} day^{-1})	NRMSE (%)	R^2	STD (MJ m^{-2} day^{-1})	RMSD (MJ m^{-2} day^{-1})	GPI (MJ m^{-2} day^{-1})
MPZ	M22	−0.048	1.541	24.300	0.830	3.363	1.540	0.760
	M23	−0.038	1.403	22.126	0.859	3.425	1.402	1.185
TCZ	M22	−0.111	1.566	25.539	0.771	2.985	1.562	0.963
	M23	−0.070	1.466	23.903	0.799	2.990	1.464	1.186
TMZ	M22	−0.090	1.143	17.558	0.901	3.404	1.140	1.219
	M23	−0.060	1.099	16.869	0.908	3.426	1.097	1.298
SMZ	M22	−0.093	1.091	14.760	0.892	3.089	1.087	1.764
	M23	−0.079	1.068	14.456	0.896	3.102	1.065	1.809
TMPZ	M22	−0.128	1.396	18.005	0.690	2.016	1.390	0.968
	M23	−0.165	1.335	17.217	0.718	2.146	1.325	1.142

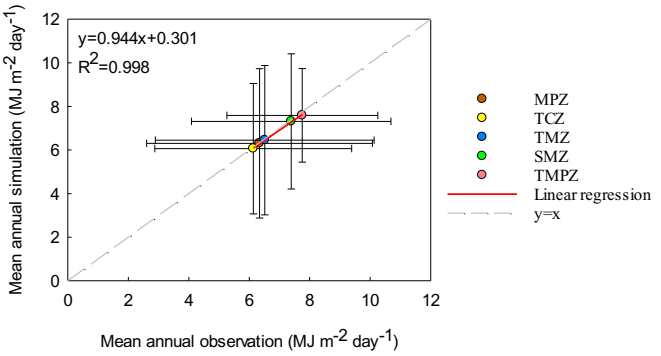

Fig. 7. Comparison between diffuse radiation observation and simulation based on M23.

3.5. Comparison with previous studies

We found the ANN models to be the best of the five categories of models for predicting diffuse radiation from the validation dataset. An ANN model was seen to be more accurate than other models and use of these ANN-type models has already become a promising research area due to their simplicity and high efficiency(Wang et al., 2016). The results shown in Fig. 5 demonstrated that the BPNN model optimized by GA was similar in overall performance and was better than empirical models. These results were consistent with previous comparison studies (Fan et al., 2019b; Feng et al., 2017; Jiang, 2008; Shamshirband et al., 2016; Wang et al., 2016). Empirical models with multiple variables have generally been observed to perform better than those with a single variable (Furlan et al., 2012; Ridley et al., 2010; Wang et al., 2019; Kambezidis et al., 2017), and that finding was consistent with the results of this study. Bailek et al. (2018) indicated that the sunshine-based models were generally more accurate than models based on the clearness index. But, in this study, the clearness index -based models (the model categories I) performed better than the corresponding models based on the sunshine duration and total cloud cover. Few studies have considered cloud data in diffuse radiation models. Furlan et al. (2012) developed a new model whose input parameters included k_t, n/N, RH, T_a, aerosols, cloud type, and precipitation. They found that the model performance was very good. Cao et al. (2017) found that combining sunshine duration and day of the year could establish a group of accurate diffuse solar radiation models in the North China Plain, and that models that included k_t, RH, and T_a performed better. Our results showed that category IV with wind speed in all zones performed better than the model of Cao et al. (2017). It was found that calculating diffuse radiation fraction based on clearness index, sunshine duration, cloud cover, temperature, relative humidity, wind speed and the day of the year was very effective. However, the estimation of diffuse radiation by only considering cloud cover may lead to poor results.

It is known that seasonal and location differences, the solar altitude, atmospheric moisture content and clouds in different climatic conditions exert significant influence on the diffuse solar radiation and can affect the relationship between diffuse solar radiation and influencing factors (Fan et al., 2019a; Feng et al., 2018). For example, the performance of all models showed poor accuracies in TCZ in terms of NRMSE, due to the frequent dust occurrences in the air (Feng et al., 2018). The R^2 of all models were low in TMPZ, because of only one site. The estimation errors in Qinghai-Tibet Plateau may be from the strong heating atmosphere (Feng et al., 2018). The best-performing models were different in different climatic zones. This disagreement is likely due to differences in environmental factors in each region. From the data analyzed in this study, the best-performing models for diffuse solar radiation estimation in different climatic regions were different from those found in previous studies by Chen et al. (2004), Jiang (2009), Feng et al. (2018) and Fan et al. (2019a). This difference can be attributed to the fact that those studies evaluated some empirical models without fully considering cloud data and the day of the year.

In addition, one limitation of this study was that some zones included only one or two stations and thus the models were evaluated as having poor performance. Another limitation of this study was that the climatic zones at high elevation had limited cloud cover, which might cause large errors in NRMSE by neglecting the elevation effects. It is important to consider cloud type and aerosols in the atmosphere and how these parameters influence diffuse radiation should be considered in diffuse radiation models.

4. Conclusions

Empirical and machine learning models for estimating the diffuse radiation (mainly based on clearness index, sunshine fraction, and cloud cover) were comprehensively reviewed in the present study. These models were evaluated and compared in the different climatic zones of China as a case study. The results confirmed that diffuse fraction had a negative correlation with clearness index and sunshine fraction but was significantly positive with total cloud cover and relative humidity. In different climatic zones, it also had a significant relationship with temperature and wind speed. Thus, the models based on the above significant meteorological parameters were established. The neural network models generally outperformed the empirical models but unable to reveal the correlations of the input factors. Among the empirical models, categories IV of models were capable of estimating daily diffuse radiation with good accuracy across China and in the different climatic zones. The model category IV and V which combined clearness index, sunshine fraction, cloud, temperature, relative humidity, wind speed and the day of the year had better performance in each climatic zone than the models in the other categories. The model categories III, mainly based on cloud cover, performed the worst in all climatic zones. This result indicated that more information about clouds (e.g. low and high cloud cover, cloud categories, cloud thick, cloud height and so on) was necessary to improve the accuracy of diffuse radiation models. Overall, in all climatic zones, the performance was ranked as followed: category V > IV > I > II > III. The results of this study allow us to choose appropriate models to estimate diffuse radiation in regions which lack

direct observations of diffuse radiation in different climatic zones. Further investigations should be conducted to further assess the applicability of more machine learning models for diffuse radiation estimation.

Credit author statement

Tingting Zhu: Formal analysis, Methodology, Writing- Original draft preparation;

Jun Li: Conceptualization, Software, Writing – review & editing.

Liang He: Data curation, Validation, Investigation.

Dingrong Wu: Data curation, Validation, Investigation.

Xiaojuan Tong: Supervision, Validation.

Qiang Yu: Supervision, Validation, Funding acquisition;

Qingyun Mu: Resources.

All authors discussed the results and contributed to the final manuscript.

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.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 41730645) and the Strategic Priority Research Program of the Chinese Academy of Sciences, Pan-Third Pole Environment Study for a Green Silk Road (No. XDA20040301). We would like to thank the China Meteorological Administration (CMA) for providing the meteorological, cloud and radiation data. We would also like to thank anonymous reviewers for the comments and suggestions that helped us to improve this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosres.2021.105505.

References

Alam, S., Kaushik, S., Garg, S., 2009. Assessment of diffuse solar energy under general sky condition using artificial neural network. Appl. Energy 86 (4), 554–564.
Bailek, N., Bouchouicha, K., Al-Mostafa, Z., El-Shimy, M., Aoun, N., Slimani, A., Al-Shehri, S., 2018. A new empirical model for forecasting the diffuse solar radiation over Sahara in the Algerian big South. Renew. Energy 117, 530–537.
Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., Papale, D., 2010. Terrestrial Gross Carbon Dioxide Uptake: Global distribution and Covariation with climate. Science 329 (5993), 834.
Behrang, M.A., Assareh, E., Ghanbarzadeh, A., Noghrehabadi, A.R., 2010. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. Sol. Energy 84 (8), 1468–1480.
Boland, J., Scott, L., Luther, M., 2001. Modelling the diffuse fraction of global solar radiation on a horizontal surface. Environmetrics 12 (2), 103–116.
Cao, F., Li, H., Yang, T., Li, Y., Zhu, T., Zhao, L., 2017. Evaluation of diffuse solar radiation models in Northern China: New model establishment and radiation sources comparison. Renew. Energy 103, 708–720.
Chen, R., Kang, E., Yang, J., Lu, S., Zhao, W., Ding, Y., 2004. Estimation of horizontal diffuse solar radiation with measured daily data in China. Renew. Energy 29 (5), 717–726.
Deo, R.C., Şahin, M., 2017. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. Renew. Sust. Energ. Rev. 72, 828–848.
Dervishi, S., Mahdavi, A., 2012. Computing diffuse fraction of global horizontal solar radiation: a model comparison. Sol. Energy 86 (6), 1796–1802.
Despotovic, M., Nedic, V., Despotovic, D., Cvetanovic, S., 2015. Review and statistical analysis of different global solar radiation sunshine models. Renew. Sust. Energ. Rev. 52, 1869–1880.
Erbs, D., Klein, S., Duffie, J., 1982. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Sol.energy 28 (4), 293–302.
Fan, J., Wu, L., Zhang, F., Cai, H., Ma, X., Bai, H., 2019a. Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. Renew. Sust. Energ. Rev. 105, 168–186.
Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X., Zou, H., 2019b. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: a review and case study in China. Renew. Sust. Energ. Rev. 100, 186–212.
Feng, Y., Cui, N., Zhang, Q., Zhao, L., Gong, D., 2017. Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. Int. J. Hydrog. Energy 42 (21), 14418–14428.
Feng, L., Lin, A., Wang, L., Qin, W., Gong, W., 2018. Evaluation of sunshine-based models for predicting diffuse solar radiation in China. Renew. Sust. Energ. Rev. 94, 168–182.
Furlan, C., De Oliveira, A.P., Soares, J., Codato, G., Escobedo, J.F., 2012. The role of clouds in improving the regression model for hourly values of diffuse solar radiation. Appl.Energy 92, 240–254.
González, J., Calbó, J., 1999. Influence of the global radiation variability on the hourly diffuse fraction correlations. Sol. Energy 65 (2), 119–131.
Iqbal, M., 1979. Correlation of average diffuse and beam radiation with hours of bright sunshine. Sol. Energy 23 (2), 169–173.
Jacovides, C.P., Tymvios, F.S., Assimakopoulos, V.D., Kaltsounides, N.A., 2006. Comparative study of various correlations in estimating hourly diffuse fraction of global solar radiation. Renew. Energy 31 (15), 2492–2504.
Jamil, B., Akhtar, N., 2017. Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: a case study. Renew. Sust. Energ. Rev. 78, 329–355.
Jiang, Y., 2008. Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. Energy Policy 36 (10), 3833–3837.
Jiang, Y., 2009. Estimation of monthly mean daily diffuse radiation in China. Appl. Energy 86 (9), 1458–1464.
Kambezidis, H.D., Psiloglou, B.E., Karagiannis, D., Dumka, U.C., Kaskaoutis, D.G., 2017. Meteorological Radiation Model (MRM v6.1): Improvements in diffuse radiation estimates and a new approach for implementation of cloud products, Renew. Sustain. Energy Rev. 74, 616–637.
Kanniah, K.D., Beringer, J., North, P., Hutley, L., 2012. Control of atmospheric particles on diffuse radiation and terrestrial plant productivity: a review. Prog. Phys. Geogr. 36 (2), 209–237.
Kanniah, K.D., Beringer, J., Hutley, L., 2013. Exploring the link between clouds, radiation, and canopy productivity of tropical savannas. Agric. For. Meteorol. 182-183 (12), 304–313.
Lebaron, B., Dirmhirn, I., 1983. Strengths and limitations of the Liu and Jordan model to determine diffuse from global irradiance. Sol.Energy 31 (2), 167–172.
Li, H., Bu, X., Long, Z., Zhao, L., Ma, W., 2012. Calculating the diffuse solar radiation in regions without solar radiation measurements. Energy 44 (1), 611–615.
Liu, B.Y., Jordan, R.C., 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. Sol. Energy 4 (3), 1–19.
Nunez, M., Li, Y., 2008. A cloud-based reconstruction of surface solar radiation trends for Australia. Theor. Appl. Climatol. 91 (1–4), 59–75.
Orgill, J., Hollands, K., 1977. Correlation equation for hourly diffuse radiation on a horizontal surface. Sol. Energy 19 (4), 357–359.
Reindl, D.T., Beckman, W.A., Duffie, J.A., 1990. Diffuse fraction correlations. Sol. Energy 45 (1), 1–7.
Ridley, B., Boland, J., Lauret, P., 2010. Modelling of diffuse solar fraction with multiple predictors. Renew. Energy 35 (2), 478–483.
Shamshirband, S., Mohammadi, K., Khorasanizadeh, H., Yee, L., Lee, M., Petković, D., Zalnezhad, E., 2016. Estimating the diffuse solar radiation using a coupled support vector machine–wavelet transform model. Renew. Sust. Energ. Rev. 56, 428–435.
Stephens, G.L., 2005. Cloud Feedbacks in the climate System: a critical Review. J. Clim. 18 (2), 237–273.
Tapakis, R., Michaelides, S., Charalambides, A.G., 2016. Computations of diffuse fraction of global irradiance: part 1–Analytical modelling. Sol. Energy 139, 711–722.
Torres, J.L., De Blas, M., García, A., de Francisco, A., 2010. Comparative study of various models in estimating hourly diffuse solar irradiance. Renew. Energy 35 (6), 1325–1332.
Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G.A., Zhu, Z., Gong, W., 2016. Solar radiation prediction using different techniques: model evaluation and comparison. Renew. Sust. Energ. Rev. 61, 384–397.
Wang, L., Lu, Y., Zou, L., Feng, L., Wei, J., Qin, W., Niu, Z., 2019. Prediction of diffuse solar radiation based on multiple variables in China. Renew. Sust. Energ. Rev. 103, 151–216.
Xue, X., 2017. Prediction of daily diffuse solar radiation using artificial neural networks. Int. J. Hydrog. Energy 42 (47), 28214–28221.