Model selection for accurate daily global solar radiation prediction in China

Shaban G. Gouda[a,b,] [*], Zakia Hussein[a,b], Shuai Luo[a,c], Qiaoxia Yuan[a, **]

[a] College of Engineering, Huazhong Agricultural University, Shizishan Street, Hongshan District, Wuhan 430070, PR China
[b] Agricultural and Biosystems Engineering Department, Faculty of Agriculture, Benha University, Benha 13736, Egypt
[c] Shandong Academy of Agricultural Machinery Sciences, Jinan 250100, China

Abstract

Solar energy can share with a big percentage in solving in the environmental pollution and energy crisis in China. For the most application of solar energy, an accurate information of solar radiation amount received by a horizontal surface is the first step. In this study, nine-day of the year based models are calibrated and evaluated to estimate the daily global solar radiation (DGSR) on a horizontal surface using long-term data of 84 stations all over China. After dividing China into five solar climatic zones, the highest performance model for each zone is chosen to be a general day of the year based models and then calibrate and evaluated at each solar zone. Based on statistical indicator; the root mean square error (RMSE), the mean absolute bias error (MABE), coefficient of determination (R^2) and the correlation coefficient (r) and the uncertainty based on a 95% confidence level (U_{95}) and Taylor diagram, results show that the models perform best in the solar zones that receive a high amount of solar radiation and the performance decreases with decreasing of the daily solar radiation incidence. The hybrid sine and cosine wave day of the year based models exhibit the best performance, model D7 is the best for zone I and model D8 for zones II and IV. For zones II and V, the Gaussian form model (D3 model) and the 4th order polynomial model (D4 model) yield high performance, respectively. Whereas, the two Gaussian form model is the worth performance for all zones. The developed general day of the year based models for each zone in this study can be used with high accuracy to estimate the (DGSR) on a horizontal surface in areas where there is no data or the data is missing all over China.

*Keywords***:** China; Day of the year; General model; Global solar radiation; Solar climate zones; Statistical evaluation

* Corresponding authors: Q. Yuan (yc1group@126.com & qxyuan@mail.hzau.edu.cn); S.G. Gouda (shaban.gouda@fagr.bu.edu.eg)

1. Introduction

Renewable energy is the best solution for problems of the energy crisis and environmental pollution. China consumed 3.25 billion tons of cool equivalent of total energy and the coal production and consumption represented 67% in 2010 according to National Bureau of Statistics of People's Republic of China (http://www.stats.gov.cn/tjsj/ndsj/2010/indexch.htm), which illustrated the significant environmental pollution problem that faces China. Solar energy is clean, safe and environmental friendly shares with a big percentage in solving these problems. However, the main challenge to use solar energy is the intermittency issue. Therefore, the solar energy systems need to be coupled with energy storage technologies which are under development now (Dizaji and Hosseini, 2018).

China has thousands of towns and hundreds of cities, and the different cities have different daily irradiations parameters and ideal obliquities depending on their different latitudes. Over 90% of China's territory receives annual solar irradiation above 4500 MJ/m^2, comprising China's abundant solar energy resources (Liu et al., 2010; Xie et al., 2012). There is a need for accurate solar radiation information to assist in system design and simulations of different topographic climates (Gouda et al., 2018). This information is also beneficial to agriculturists, architects, urban planners, manufacturers and other field engineers (Cao et al., 2017; Lemos et al., 2017; Yao et al., 2017;). Despite the huge investment and priority of the Chinese government, only 122 of more than 2000 meteorological stations measure solar radiation (Chen and Li, 2013). Therefore, it is essential to find an accurate method for estimating the daily global solar radiation (DGSR) on a horizontal surface, especially, in the locations that do not contain on any tools to measure it.

For this purpose, several empirical models have been investigated to estimate the DGSR depending on meteorological and geographical data of the site of interest, such as sunshine duration (Angstrom, 1924; Bahel et al., 1987; Jemaa et al., 2013; Ögelman et al., 1984; Prescott, 1940; Rietveld, 1978), air temperature (Allen, 1997; Chen and Li, 2013; Chen et al., 2004; Li et al., 2014; Li et al., 2010b), precipitation (Li et al., 2010b), relative humidity, cloud, latitude longitude, altitude (Chen et al., 2006; Li et al., 2010b; Liu et al., 2009; Zhao et al., 2013; Zou et al., 2016; Goudaet al., 2018). All these models are dependable, hence, to estimate the DGSR need to have any meteorological data that require tools to measure which make it difficult and more cost in some locations all over China.

The independent models such as day of the year based model can solve this problem because there is no dependency to any specific input meteorological data and is not need to any pre-calculation analysis (Mohammadi et al., 2015). However, a few day of the year based models have been previously developed for estimating the DGSR on a horizontal surface that found in the literature. Four different models were proposed by Al-Salaymeh (2006) to predict the DGSR in Amman city, Jordan, and the sine wave form yielded the best among others. Two trigonometric day of the year based models were developed, the first one is sine wave form which was proposed by Bulut and Büyükalaca (2007), the obtained results showed a good accuracy for estimating the DGSR in Istanbul city, Turkey. The second model is cosine wave function which was investigated by Kaplanis and Kaplani (2007). The performance of this model was evaluated at six climate zones in Greece and yielded high performance. Li et al. (2010a) evaluated three existed day of the year based model to estimate the DGSR on a horizontal surface using long-term data of 79 stations all over China, and proposed new model that consists

of a hybrid sine and cosine wave, the results indicated to the new proposed model exhibited a better performance than others.

Additionally, Zang et al. (2012) examined two existed models and proposed a new one which is similar to the model of Ref. (Li et al., 2010a), but in a different form. They proposed a new model and evaluated it beside another two existed models using data of 35 stations in China. The new model performed well across all zones in China hence, there is a need to use it on a larger scale. In four Iranian cities of Bandarabass, Isfahan, Kerman and Tabass, six-day of the year based models were examined for estimating the DGSR and determine the best performance for each city and regions with similar climates (Khorasanizadeh and Mohammadi, 2013). A universal model for estimating the DGSR based on the day of the year was developed by Kaplanis et al. (2016). The model is two cosine wave functions that can be used in the Southern and Northern hemispheres with higher accuracy for most locations located between 0° to 71° from north to south and for most all longitudes.

Recently, seven existing day of the year based models were examined and evaluated to predict the DGSR in ten cities in Egypt by Hassan et al. (2016). The results showed that the hybrid sine and cosine wave and 4th order polynomial models performed best and therefore, a general model was developed for all country. A new day of the year based DGSR models in form of two Gaussians function was proposed by Quej et al. (2017) for prediction in six cities in the Yucatan Peninsula, Mexico, also they compared the results with other four existing day of the year based models. The results indicated that the new model had a higher accuracy than others, also in the seasonal analysis in all seasons. Aoun and Bouchouicha (2017) examined six existed day of the year based models for estimating the DGSR at 21 cities, Algeria, and developed a general model

for all country using 21-year of DGSR data. The results showed that the hybrid sine and cosine wave model yielded the best performance for all Algeria.

All developed general models to predict the DGSR on a horizontal surface in China are dependent models which need to observed meteorological data and most of them are not cover all China mainland for example Li et al. (2013b) collected a data of solar radiation, sunshine duration, relative humidity, temperature from 83 stations across China to develop a general model to estimate DGSR for different solar radiation zones in China based on latitude, longitude and altitude and site-specific models. They found that the developed general model based on sunshine duration, temperature, relative humidity, and geographical information for each zone can be used with good accuracy. Li et al. (2013a) estimated the DGSR on a horizontal surface in Southwest China by two models based on maximum and minimum temperatures. The results showed that these two models performed best.

It is noticed that from the literature, models' performance is not the same, different from site to site and from mathematical form to other form. Meanwhile, some studies have been carried out to study the day of the year based models and focused on a few stations. Additionally, there are no independent general models were developed for China based on solar climatic zones. Therefore, this study aims to examined and evaluate the existed day of the year based models in the literature to estimate the DGSR on horizontal surface using a data from 84 stations across China, to classify China to different zones using the best solar climatic zoning and select the model with the highest performance in each solar zone and then determine and evaluate the general model to estimate the DGSR for different solar climatic zones.

2. Materials and methods

2.1. Data used and study area

The daily solar radiation data are taken from DAM (http://dam.itpcas.ac.cn/data/daily_solar_radiation_dataset_over_China.zip). These data include 716 China Meteorological Administration (CMA) stations and cover a 50-year period from 1961 to 2010. It obtained by combining two datasets. The first one is the hybrid daily solar radiation model stations developed by Yang et al., (2001) and Yang et al., (2006) estimate at 716 stations and the second one is the artificial neural network-based model estimate at 96 radiation stations, which has higher accuracy. These data were used in several studies such as Tang et al., (2013); Tang et al., (2010) and Zou et al., (2016).

China is a large country with an area of about 9.6 million km^2 (Lam et al., 2005). China includes a vast territory and adjoins the Eurasian continent in the north and the Pacific Ocean in the east. Topographic elevations generally decrease from west to east. Those factors result in markedly different climatic conditions, rainfall generally decreases from the southeast coast to the northwest inland, and solar radiation is unevenly distributed in China (Ouyang et al., 2017). Numerous ways are used to classify different climatic zones according to different criteria using different climatic variables and indices. This depends largely on the purpose of establishing such classification (Domrös and Gongbing, 1988).

The solar radiation zoning in China are the prerequisites for the assessment the global solar radiation models. There have been a number of works on climate classification since the 1930s (Lam et al., 2005) e.g., Lau et al. (2007) and Cao et al. (2017). In the present study, the classification by Liu et al. (2017) is used, they employed a data from 660 meteorological stations for classifying China to five solar radiation zones using the k-means clustering and Support

Vector Machine-Genetic Algorithm, based on the monthly mean daily solar radiation as a zoning index (Gouda and Yuan, 2018a, b, c; Liu et al., 2018). **Fig. 1.** illustrates the distribution of 84 stations with corresponding solar climatic zones across China. The geographical information and the data records of 84 stations distributed in 5 solar zones across China are listed in **Table 1**.
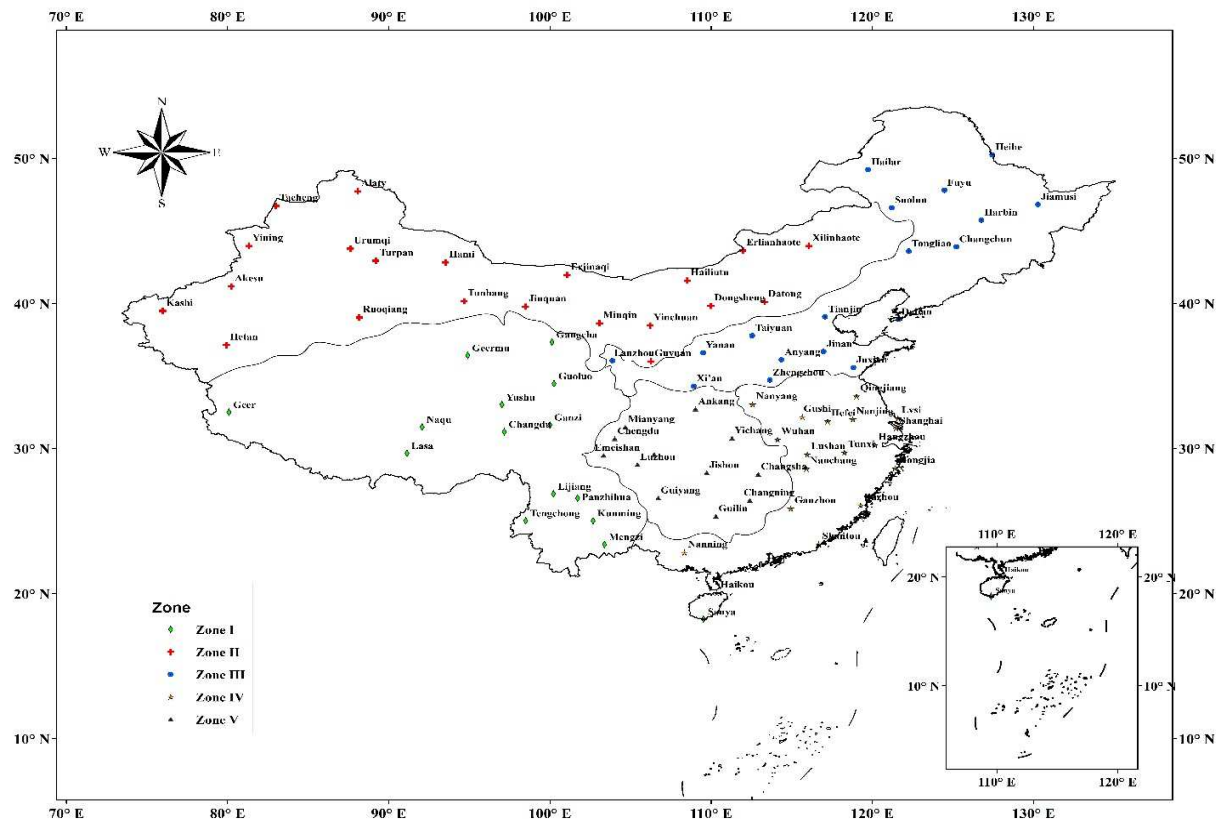

Fig. 1. A map of China solar zones with corresponding stations.

Table 1

Geographical locations and data records of the selected stations with corresponding solar climatic zones across China.

Zone	Stations	Longitude (° E)	Latitude (° N)	Altitude (m)	Period Calibration	Validation	Zone	Stations	Longitude (° E)	Latitude (° N)	Altitude (m)	Period Calibration	Validation
I	Gangcha	100.13	37.33	3301.5	1961-1995	1996-2006	III	Lanzhou	103.88	36.05	1517.2	1961-1993	1994-2004
	Geermu	94.9	36.42	2807.6	1961-1995	1996-2006		Taiyuan	112.55	37.78	778.3	1961-1995	1996-2006
	Geer	80.08	32.5	4278	1961-1995	1996-2006		Yanan	109.5	36.6	958.5	1961-1995	1996-2006

	Station	Lon	Lat	Elev	Period 1	Period 2
	Naqu	92.07	31.48	4507	1961-1995	1996-2006
	Lasa	91.13	29.67	3648.7	1969-1995	1996-2006
	Yushu	97.02	33.02	3681.2	1961-1995	1996-2006
	Guoluo	100.25	34.47	3719	1991-2000	2001-2006
	Changdu	97.17	31.15	3306	1961-1995	1996-2006
	Ganzi	100	31.62	3393.5	1961-1995	1996-2006
	Lijiang	100.22	26.87	2392.4	1961-1995	1996-2006
	Panzhihua	101.72	26.58	1190.1	1988-2001	2002-2006
	Tengchong	98.5	25.02	1654.6	1961-1995	1996-2006
	Kunming	102.68	25.02	1892.4	1969-1995	1996-2006
	Mengzi	103.38	23.38	1300.7	1961-1995	1996-2006
	Sanya	109.52	18.23	5.9	1961-1995	1996-2006
II	Yining	81.33	43.95	662.5	1961-1995	1996-2006
	Urumqi	87.62	43.78	917.9	1961-1995	1996-2006
	Turpan	89.2	42.93	34.5	1961-1995	1996-2006
	Akesu	80.23	41.17	1103.8	1961-1995	1996-2006
	Alaty	88.08	47.73	735.3	1961-1995	1996-2006
	Tacheng	83	46.73	534.9	1961-1995	1996-2006
	Kashi	75.98	39.47	1288.7	1968-1995	1996-2006
	Ruoqiang	88.17	39.03	888.3	1961-1995	1996-2006
	Hetan	79.93	37.13	1374.5	1961-1995	1996-2006
	Hami	93.52	42.82	737.2	1968-1995	1996-2006
	Erjinaqi	101.07	41.95	940.5	1961-1995	1996-2006
	Tunhang	94.68	40.15	1139	1969-1995	1996-2006
	Jiuquan	98.48	39.77	1477.2	1961-1995	1996-2006
	Minqin	103.08	38.63	1367	1961-1995	1996-2006
	Erlianhaote	111.97	43.65	964.7	1961-1995	1996-2006
	Hailiutu	108.52	41.57	1288	1962-1995	1996-2006
	Datong	113.33	40.1	1067.2	1961-1995	1996-2006
	Dongsheng	109.98	39.83	1460.4	1961-1995	1996-2006
	Yinchuan	106.22	38.48	1111.4	1961-1995	1996-2006
	Guyuan	106.27	36	1753	1961-1995	1996-2006
	Xilinhaote	116.07	43.95	989.5	1961-1995	1996-2006
	Heihe	127.45	50.25	166.4	1961-1995	1996-2006
	Hailar	119.75	49.22	610.2	1961-1995	1996-2006
	Fuyu	124.48	47.8	162.7	1961-1995	1996-2006
	Suolun	121.22	46.6	499.7	1961-1995	1996-2006
	Jiamusi	130.28	46.82	81.2	1961-1995	1996-2006
III	Harbin	126.77	45.75	142.3	1961-1995	1996-2006
	Anyang	114.37	36.12	75.5	1961-1995	1996-2006
	Tongliao	122.27	43.6	178.5	1961-1995	1996-2006
	Changchun	125.22	43.9	236.8	1961-1995	1996-2006
	Tianjin	117.07	39.08	2.5	1961-1995	1996-2006
	Dalian	121.63	38.9	91.5	1961-1995	1996-2006
	Jinan	116.98	36.68	51.6	1961-1995	1996-2006
	Juxian	118.83	35.58	107.4	1961-1989	1990-1997
	Xi'an	108.93	34.3	397.5	1961-1995	1996-2006
	Zhengzhou	113.65	34.72	110.4	1961-1995	1996-2006
IV	Nanyang	112.58	33.03	129.2	1961-1995	1996-2006
	Wuhan	114.13	30.62	23.1	1961-1995	1996-2006
	Ganzhou	114.95	25.85	123.8	1961-1995	1996-2006
	Qingjiang	119.03	33.6	17.5	1961-1991	1992-2001
	Gushi	115.67	32.17	57.1	1961-1995	1996-2006
	Nanjing	118.8	32	8.9	1961-1995	1996-2006
	Lvsi	121.6	32.07	5.5	1961-1995	1996-2006
	Hefei	117.23	31.87	27.9	1961-1995	1996-2006
	Shanghai	121.48	31.4	6	1991-2001	2002-2006
	Hangzhou	120.17	30.23	41.7	1961-1995	1996-2006
	Lushan	115.98	29.58	1164.5	1961-1995	1996-2006
	Tunxi	118.28	29.72	142.7	1961-1995	1996-2006
	Nanchang	115.92	28.6	46.7	1961-1995	1996-2006
	Hongjia	121.42	28.62	1.3	1969-1995	1996-2006
	Fuzhou	119.28	26.08	84	1961-1995	1996-2006
	Shantou	116.68	23.4	1.1	1961-1995	1996-2006
	Nanning	108.35	22.82	73.1	1961-1995	1996-2006
	Haikou	110.35	20.03	13.9	1961-1995	1996-2006
	Mianyang	104.68	31.47	470.8	1963-1995	1996-2006
V	Chengdu	104.02	30.67	506.1	1961-1993	1994-2003
	Emeishan	103.33	29.52	3047.4	1961-1995	1996-2006
	Ankang	109.03	32.72	290.8	1961-1995	1996-2006
	Yichang	111.3	30.7	133.1	1961-1995	1996-2006
	Chongqing	106.47	29.58	259.1	1961-1995	1996-2006
	Luzhou	105.43	28.88	334.8	1961-1995	1996-2006
	Jishou	109.73	28.32	208.4	1961-1994	1995-2005
	Changsha	112.92	28.22	68	1987-2000	2001-2006
	Guiyang	106.72	26.58	1074.3	1961-1995	1996-2006
	Changning	112.4	26.42	116.6	1987-2000	2001-2005
	Guilin	110.3	25.32	164.4	1969-1995	1996-2006

2.2. Models used

Nine models using only the day of the year without using any other parameters existed in the literature for estimating the DGSR on a horizontal surface are evaluated and examined in this work. A brief introduction is as follows.

D1: Al-Salaymeh model 1

Al-Salaymeh (2006) proposed four different form models to estimate the DGSR on a horizontal surface in Amman city, Jordan. The first form is a sine wave correlation:

$$R_g = a_1 + a_2 \sin\left(\frac{2\pi N}{a_3} + a_4\right) \tag{1}$$

Where R_g is the DGSR on a horizontal surface, MJ/m^2 day, N is the number of the day of the year and a_1, a_2, a_3, and a_4 are the amplitude, the phase shift, the wavelength, and the intercept, respectively. empirical coefficients.

D2: Al-Salaymeh model 2

The second model is the following Lorentzin correlation model (Al-Salaymeh, 2006):

$$R_g = \frac{a_1}{1 + \left(\dfrac{N - a_2}{a_3}\right)^2} \tag{2}$$

Where a_1, a_2, and a_3 are the amplitude, the center, and the width, respectively.

D3: Al-Salaymeh model 3

The suggested Gaussian form is the third, presented as (Al-Salaymeh, 2006):

$$R_g = a_1 \exp\left[-0.5 \times \left(\frac{N - a_2}{a_3}\right)^2\right] \tag{3}$$

Where a_1, a_2, and a_3 are the amplitude standing, the center, and the width, respectively.

D4: Al-Salaymeh model 4

A 4th order polynomial degree model is finally proposed (Al-Salaymeh, 2006):

$$R_g = a_1 + a_2 N + a_3 N^2 + a_4 N^3 + a_5 N^4 \tag{4}$$

Where a_1-a_4 are the empirical coefficients.

D5: Bulut and Büyükalaca model

Bulut and Büyükalaca (2007) presented a simple model depending only on the day of the year to predict the DGSR on a horizontal surface using a sine wave equation.

$$R_g = a_1 + a_2 \left| \sin\left(\frac{\pi(N+5)}{365} \right) \right|^{1.5} \tag{5}$$

D6: Kaplanis and Kaplani

Kaplanis and Kaplani (2007) suggested the following cosine wave form to predict the DGSR based on the day of the year:

$$R_g = a_1 + a_2 \cos\left(\frac{2\pi N}{365} + a_3 \right) \tag{6}$$

D7: Li model

Li et al. (2010a) proposed a trigonometric model in conjunction with sine and cosine waves to estimate the DGSR by using the daily measured global solar radiation data from 79 meteorological stations covering all China:

$$R_g = a_1 + a_2 \sin\left(\frac{2\pi N}{365} a_3 + a_4 \right) + a_5 \cos\left(\frac{2\pi N}{365} a_6 + a_7 \right) \tag{7}$$

Where a_1-a_7 are the empirical coefficients.

D8: Zang model

Zang et al. (2012) developed a novel global solar radiation model in conjunction with sine and cosine wave correlations to simulate the long-term measured data over at least 10 years at 35 stations:

$$R_g = a_1 + a_2 \sin\left(\frac{2\pi N}{365} + a_3\right) + a_4 \cos\left(\frac{2\pi N}{365} + a_5\right) \tag{8}$$

D9: Quej model

Quej et al. (2017) proposed a new day of the year based model using a sum of two Gaussian correlation formulas to estimate the DGSR on a horizontal surface of Yucatan Peninsula in Mexico:

$$R_g = a_1 + a_2 \exp\left[-0.5 \times \left(\frac{N - a_3}{a_4}\right)^2\right] + a_5 \exp\left[-0.5 \times \left(\frac{N - a_6}{a_7}\right)^2\right] \tag{9}$$

CurveExpert software (http://www.curveexpert.net) is employed to obtain the empirical coefficients of used models, using nonlinear regression method for all equations except Eq. (4), the polynomial regression method is used.

2.2. Model performance evaluation

In this study, the accuracy and suitability of the models were assessed based on common statistical indicators, namely, the root mean square error (RMSE), the mean absolute bias error (MABE), the coefficient of determination (R^2) and the correlation coefficient (r). These indicators are described as follows:

$$RMSE = \sqrt{\frac{1}{j} \sum_{i=1}^{j} \left(\overline{R}_{i,m} - \overline{R}_{i,c}\right)^2} \tag{10}$$

$$MABE = \frac{1}{j} \sum_{i=1}^{j} \left| \left(\overline{R}_{i,m} - \overline{R}_{i,c} \right) \right| \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{j} \left(\overline{R}_{i,m} - \overline{R}_{i,c} \right)^2}{\sum_{i=1}^{j} \left(\overline{R}_{i,m} - \overline{R}_{m,avg} \right)^2} \tag{12}$$

$$r = \frac{\sum_{i=1}^{j} \left(\overline{R}_{i,c} - \overline{R}_{c,avg} \right) \left(\overline{R}_{i,m} - \overline{R}_{m,avg} \right)}{\left(\sum_{i=1}^{j} \left(\overline{R}_{i,c} - \overline{R}_{c,avg} \right)^2 \sum_{i=1}^{j} \left(\overline{R}_{i,m} - \overline{R}_{m,avg} \right)^2 \right)^{0.5}} \tag{13}$$

where j is the number of values, the subscript i, represents the ith value of the solar radiation; c and m represent calculated and measured solar radiation values, respectively; $\overline{R}_{c,avg}$ and $\overline{R}_{m,avg}$ are the mean calculated and measured solar radiation values, respectively; and R^2 varies between 0 and 1 (a value of R^2 closer to 1 reflects better performance). This statistic provides information about how well a model can estimate a set of measured data. A model exhibits better performance when the RMSE is closer to zero. Additionally, r ranges from $-\infty$ to 1 and a model with good performance will have an r value close to one. The MABE gives the absolute value of mean bias error which provide information on the performance and low value is desired, hence MABE's values close to zero reflect high performance.

Furthermore, the uncertainty based on a 95% confidence level (U_{95}) was used to provide more information about model deviation (Gueymard, 2014):

$$U_{95} = 1.96 \left(SD^2 + RMSE^2 \right)^{0.5} \tag{14}$$

where 1.96 is the coverage factor corresponding to the 95% confidence level and SD is the standard deviation of the difference between the calculated and measured values.

3. Results and Discussions

In this study, nine models based on day of the year are evaluated for their accuracy for estimating DGSR at 84 stations all over China, after classifying China into five solar zones according to the DGSR data.

The measured data of DGSR at 84 stations across China are divided into two datasets. The first dataset is employed to obtain the empirical coefficients of the models based on multiple linear and nonlinear regression methods. The coefficient of determination (R^2) indicator is used to determine how well the regression line approximates the actual data points. The second dataset is employed to validate the models. **Table S1 (supplementary data)** presents the regression coefficients of the 9 models with statistical indices at 84 location distributed under 5 solar zones across China. The performance of tested models varies from site to site accordingly, from zone to zone. The best model is highlighted in bold.

In the following subsections, the performance of different models is assessed based on statistical indices and Taylor diagram and compared with observed data after a brief climatic discerption for each zone **(Table 2)**. The best model is selected and a representative station is also selected to illustrate the differences between the used models and the performance of the best model for each station.

Table 2

Climatic description of the five solar zones across China.

Zone	Main location	Hottest month	Monthly mean daily temperature (°C)	Annual mean daily relative humidity (%)	Annual mean daily sunshine duration (h)	Annual mean DGSR (MJ/m² day)
I	Tibet, Qinghai, Ningxia, Yunnan and Hainan	July	16.1	59.5	6.8	17.0

II	Xinjiang, Inner Mongolia and Gansu	August	22.8	51.2	8.0	15.9
III	East and Central China	August	23.4	64.3	6.7	13.9
IV	Eastern and southern coastal areas of China	July	27.4	77.5	4.9	12.8
V	Sichuan and Guiyang	July	26.2	78.6	3.6	10.4

3.1. Zone I

For this zone, model D7 exhibits the best performance among the nine models, and its performance is more stable through all stations in zone I. For this model, r ranges between 0.755 and 0.986. RMSE, MABE, U_{95} and R^2 values vary between 0.933 to 1.646 MJ/m^2 day, 0.751 to 1.308 MJ/m^2 day, 2.575 to 4.550 and 0.799 to 0.983, respectively. In addition, Model D4 yields good and stable performance all over zone I, but less than model D7. It has ranges for r, RMSE, MABE, U95 and R^2 of 0.668-0.985, 0.953-1.861 MJ/m^2 day, 0.755-1.456 MJ/m^2 day, 2.637-5.157 and 0.667-0.970, respectively (**Table S1**).

Model D8 performs best, furthermore, models D2 and D3 have less performance than model D8. Whereas, model D1 yields the worst performance followed by model D9, where they have lowest r values (0.328 and 0, respectively). Model D7 presents superior performance among others; hence, the daily data of 18 stations are averaged and employed to fit model D7 as the best model for this zone. The results listed in **Table 3** show that the general model for zone I yields an acceptable performance. For this model, r, RMSE, MABE, U_{95} and R^2 values are 0.984, 0.780 MJ/m^2 day, 0.586 MJ/m^2 day, 2.062 and 0.976, respectively.

Naque station is selected to represent zone I, **Fig. 2** compares the DGSR data and the predicted data of different models and the best model for this zone (GM).
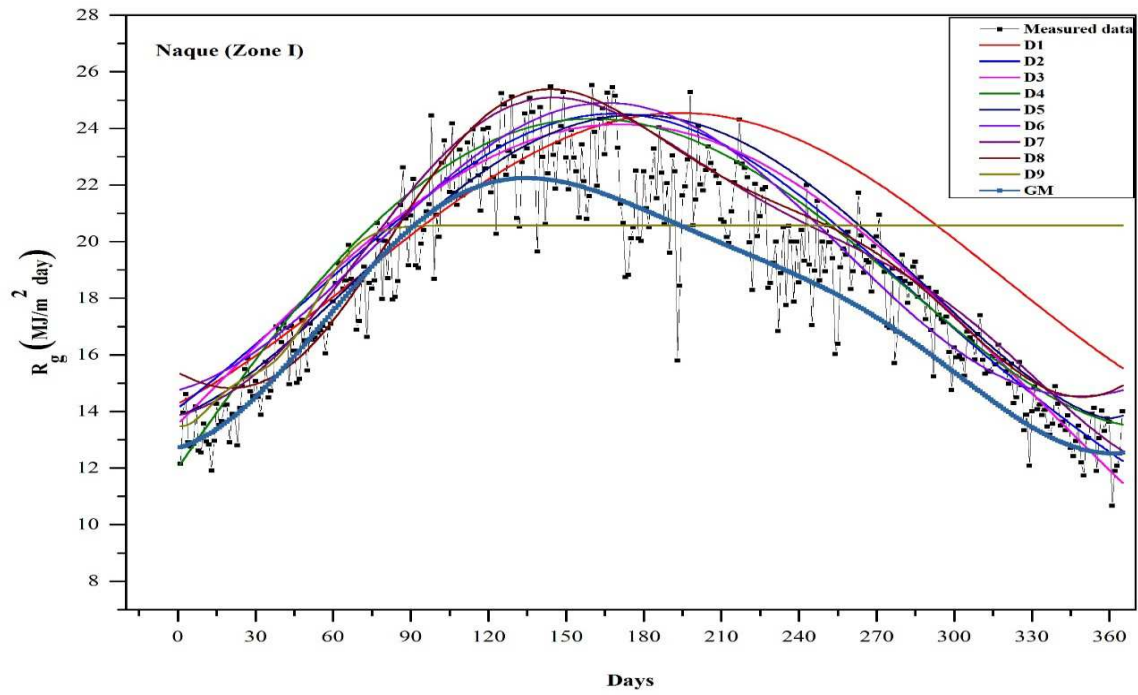

Fig. 2. The observed data compared with the predicted data by the used model and the best model (GM) in Naque (zone I).
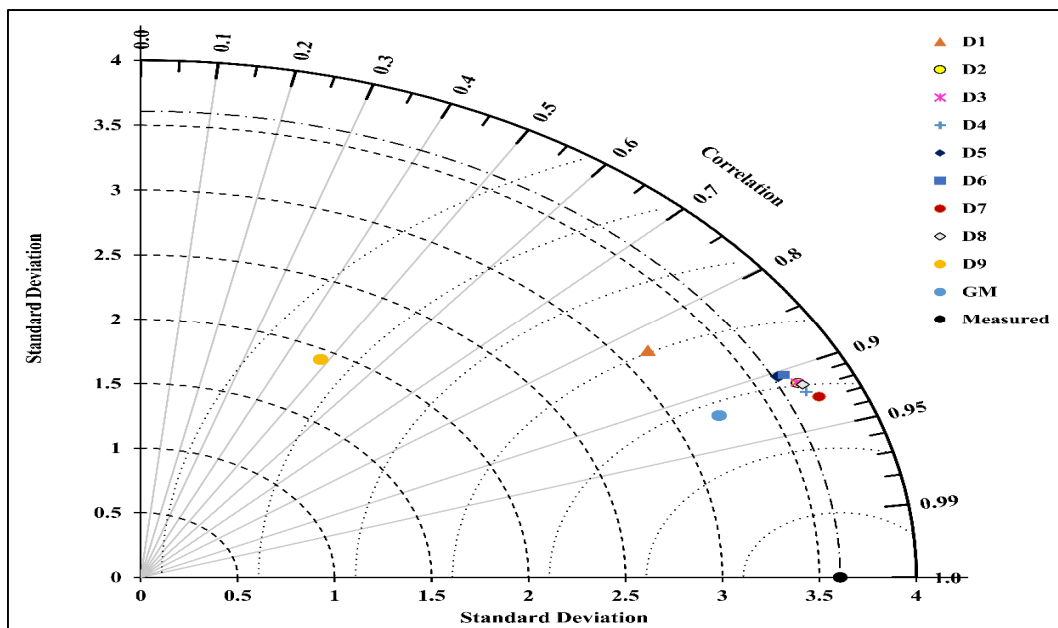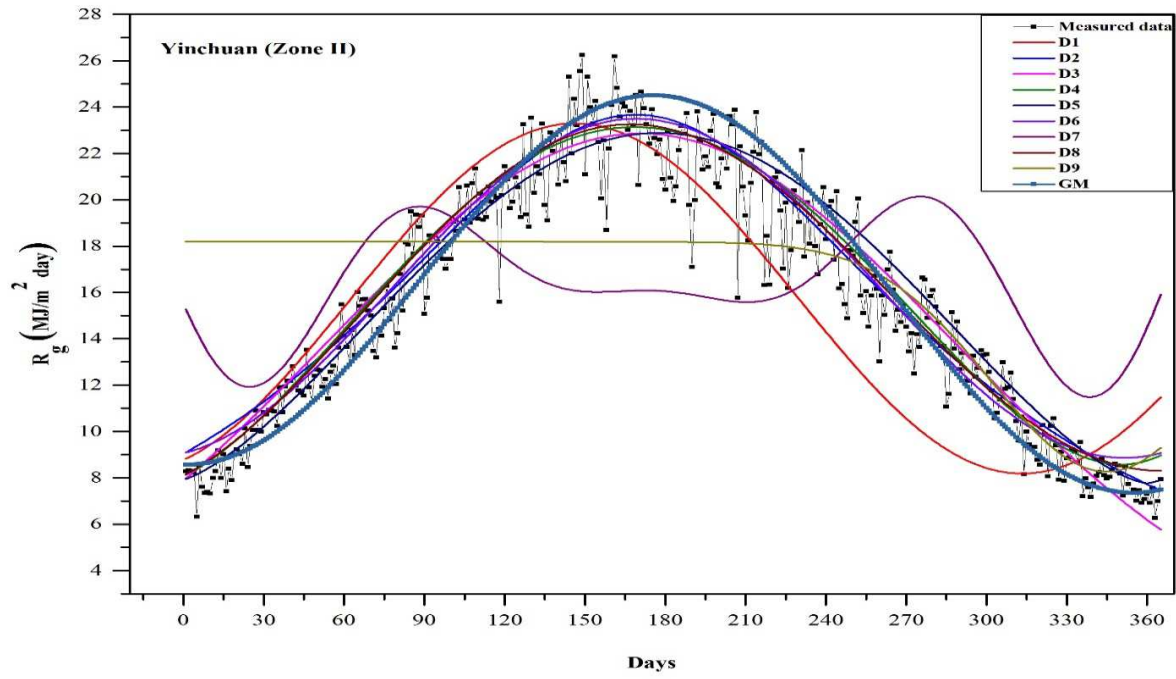
Fig. 3. Taylor diagram for the used models and the general model (GM) of DGSR at Naque (zone I).

A Taylor diagram is used to compare the performance of the models. The main advantage of presenting the performance of the models in a Taylor diagram is that the models are clustered according to their performance. In agreement with the results in **Table S1**, models D9 and D1 exhibit the furthest representative points from the reference point (Measured), which indicates that these performed poorer than other models. Otherwise, models D1, D2, D3, D4, D5, D6, and D8 are the closest models to the reference point (Measured). Hence, their representative points were close to each other, which suggests that their performance was similar. Additionally, Model D7 is the nearest point to the measured point, therefore, it has the best performance among others, and the general model (GM) for zone I yielded an acceptable performance, where it is representative point closes to the measured point.

3.2. Zone II

For zone II, model D8 exhibits the best performance, it has r values vary from 0.935-0.989, RMSE values are between 1.071 to 1.688 MJ/m^2 day, MABE values are between 0.820 to 1.336 MJ/m^2 day, U_{95} ranges between 2.968 to 4.666 and R^2 vary between 0.948 to 0.991. Model D6 performs best slightly decrease compared to model D8. For D6 model, the values of r, RMSE, MABE, U95 and R^2 range between 0.934-0.989, 1.087-1.811 MJ/m^2 day, 0.835-1.455 MJ/m^2 day, 3.014-4.959 and 0.948-0.991, respectively. This illustrates that using sine and cosine functions in one model (D8) leads to an improvement as reported by Zang et al. (2012). Model D2 yields good performance followed by models D3, D5, D4, and D1. The worst performance is yielded by D9 followed by D7, where the values of r and R^2 vary between 0 to 0.983 and 0.238 to 0.989; and 0 to 0.983 and 0.59 to 0.992, respectively. These results indicate that D7 and D9 models are

not suitable for estimating DGSR in zone II, and the superior performance is exhibited by model D8 **(Table S1).** Thus, the averaged data of 21 stations in zone II is employed to obtain the empirical coefficients of the best model (D8) for this zone, the empirical coefficients and statistical indices for the general model of zone II (GM) are listed in **Table 3**. It has r, RMSE, MABE, U95 and R^2 values are 0.991, 0.783 MJ/m^2 day, 0.622 MJ/m^2 day, 2.160 and 0.992, respectively.

Furthermore, Yinchuan station is chosen to represent zone II. The observed and predicted daily data are figured **(Fig. 4)** to illustrate the difference between the performance of different models, besides the best model for zone II (GM). It can be seen from the Taylor diagram (**Fig. 5**), models D9, D7 and D1 perform poorer than other models where they have the furthest representative points from the reference point. Otherwise, models D2, D3, D4, D5, D6, and D8 are the closest models to the reference point and close to each other, which suggests that their performance is similar. Model D8 is the nearest point to the measured point, therefore, it has the best performance among others. The general model (GM) for zone II exhibits an acceptable performance, where its representative point is not far from the measured point.

Fig. 4. The observed data compared with the predicted data by the used model and the best model (GM) in Yinchuan station (zone II).
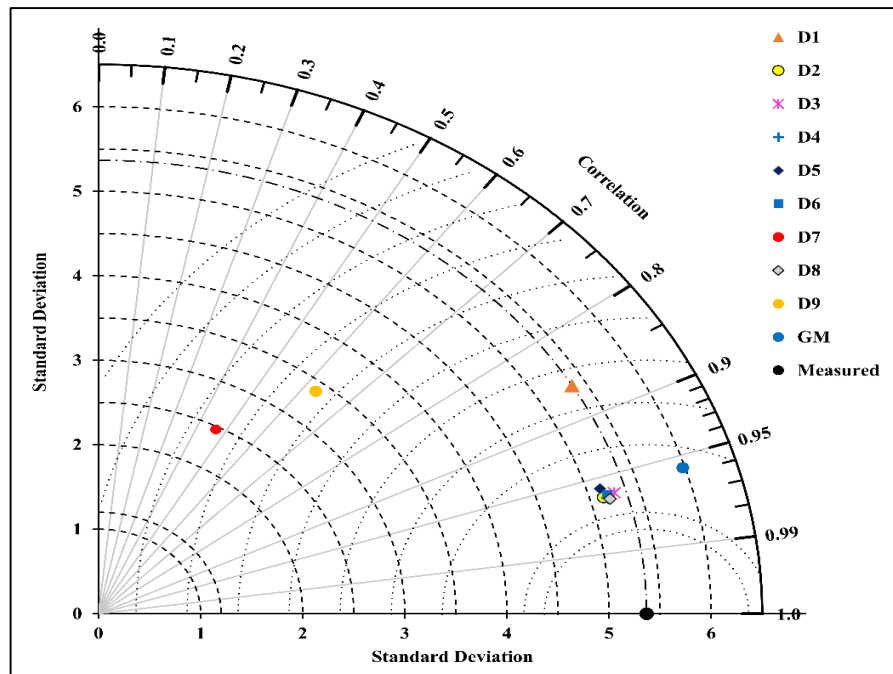

Fig. 5. Taylor diagram for the used models and the general model (GM) of DGSR in Yinchuan station (zone II).

3.3. Zone III

For zone III, the best and more stable performance is yielded by model D3, it has r values vary between 0.885 and 0.975, RMSE values are between 1.648 and 2.304 MJ/m^2 day, MABE values are between 1.233 and 1.764 MJ/m^2 day, U$_{95}$ values range between 4.305 and 6.125, and R^2 values are between 0.917 and 0.983. Models D2, D6, D5, and D1 perform best, whereas models D4, and D8 yields a good performance in all stations except Yanan station for D4 model and Jiamusi station for D8, they have highest values of RMSE, MABE, and U$_{95}$ (7.806 and 10.248 MJ/m^2 day, 7.638 and 10.098 MJ/m^2 day, and 15.623 and 20.376, respectively). The worst performances are yielded by models D9 and D7. For these two models, they have the lowest r and R^2 values; r equals 0 and 0.17 for D9 and D7, respectively, and R^2 equals 0 and 0.027 for D9 and D7, respectively. In this zone, it can notice that from **Table S1,** for the best models' performance, the models perform better in the northeast zone III (Northeast of China) (Hailar and Heihe), and the performance decreases in stations which located in the south of zone III (Juxian and Zhengzhou).

The data of 18 stations are averaged and used for obtaining the empirical coefficients of model D3 which performed the best of zone III, where R^2, RMSE, MABE, U$_{95}$ and r values are 0.980, 1.042 MJ/m^2 day, 0.835 MJ/m^2 day, 2.743 and 0.983, respectively **(Table 3).**

Hailar station represents zone III, and the measured daily data from this station are figured with the calculated data by the used models and GM models as the superior performance among others in this zone (**Fig. 7**). In addition, the Taylor diagram agrees with the presented results in **Table 3.** Models D9 and D1 performed poorer than other models where they have the furthest representative points from the reference point. Otherwise, models D2, D3, D4, D5, D6, D7, and

D8 are the closest models to the reference point, which suggests that their performance is the best. Additionally, the general model (GM) yields best, where its representative point is close to the measured point.
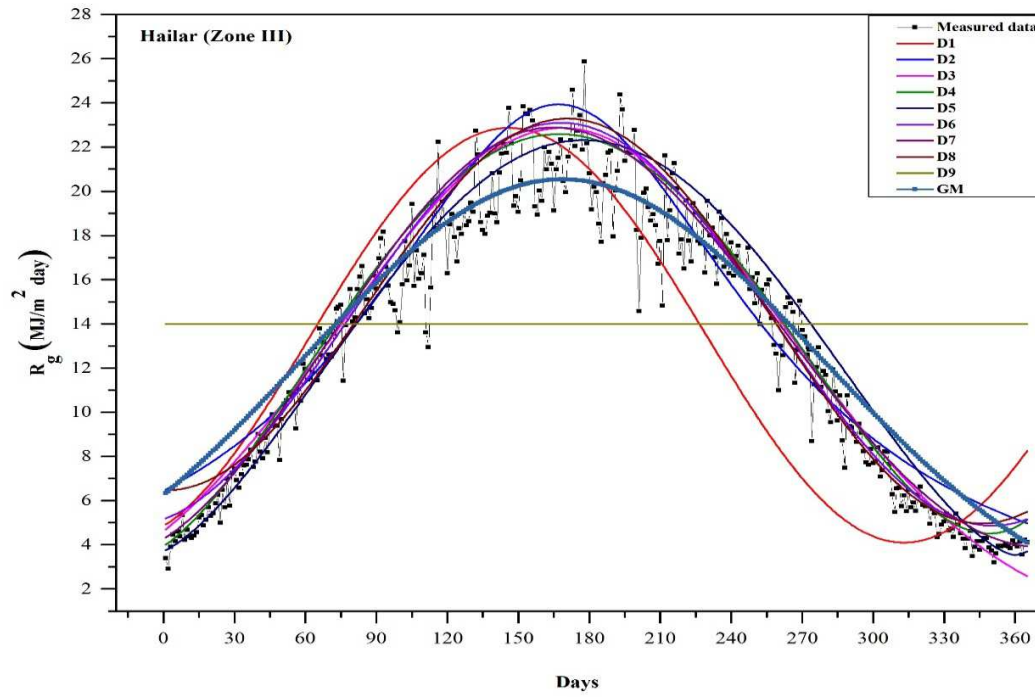

Fig. 6. The observed data compared with the predicted data by the used model and the best model (GM) in Hailar (zone III).
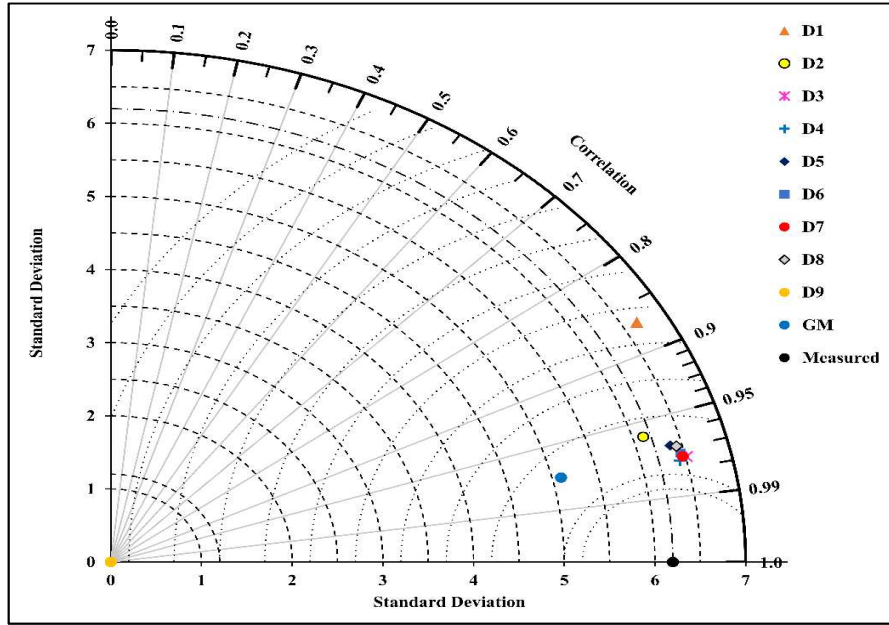
Fig. 7. Taylor diagram for the used models and the general model (GM) of DGSR in Hailar (zone III).

3.4. Zone IV

For zone IV, D8 model performed best followed by model D4. For these two models, the values of r, RMSE, MABE, U_{95} and R^2 range between 0.789-0.916 and 0.790-0.914, 2.015-2.651 and 2.035-2.644 MJ/m^2 day, 1.615-2.069 and 1.639-2.061 MJ/m^2 day, 5.115-7.348 and 5.163-7.330, and 0.740-0.924 and 0.739-0.921, respectively. Models D3, D2, D6, and D5 exhibit acceptable and approximately similar performance. On other hands, model D9 followed by models D1 and D7 give the worst performance among others, where the values of r are between 0-0.907, 0.158-0.904 and 0.299-0.909, respectively. R^2 values range between 0-0.931, 0.119-0.938 and 0.024-0.927, respectively **(Table S1)**. It is noticeable that the model D8 is the superior performance among others in zone IV. Therefore, the averaged data of 18 station are employed to fit this model and the empirical coefficients and statistical indicators are presented in **Table 3**. For

21

general model, R^2, RMSE, MABE, U_{95} and r values are 0.950, 1.241 MJ/m^2 day, 1.0 MJ/m^2 day, 3.366 and 0.947, respectively.

The measured daily solar radiation data of Lvsi station as a representative station of zone IV are employed to show the difference between the predicted data by different models and the general model (GM) for this zone **(Fig. 8)**. Furthermore, the performance of different models and the general models is illustrated in Taylor diagram **(Fig. 9)**, the furthest representative points from the reference point are points of D9, D8, and D1 models, which suggests that they have a poor performance agreed with the results in **Table S1.** Otherwise, the nearest points to the reference point are the representative points of models D2, D3, D4, D5, D6, and GM, which means a good and similar performance is performed by these models.
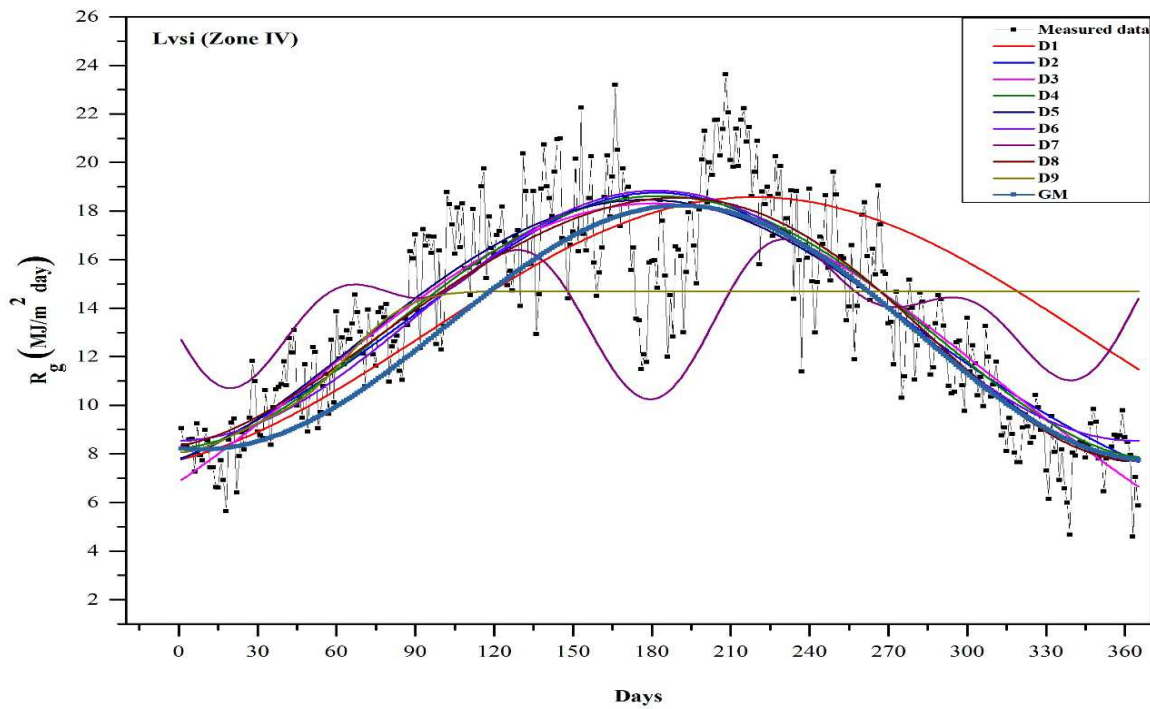

Fig. 8. The observed data compared with the predicted data by the used model and the best model (GM) in Lvsi (zone IV).
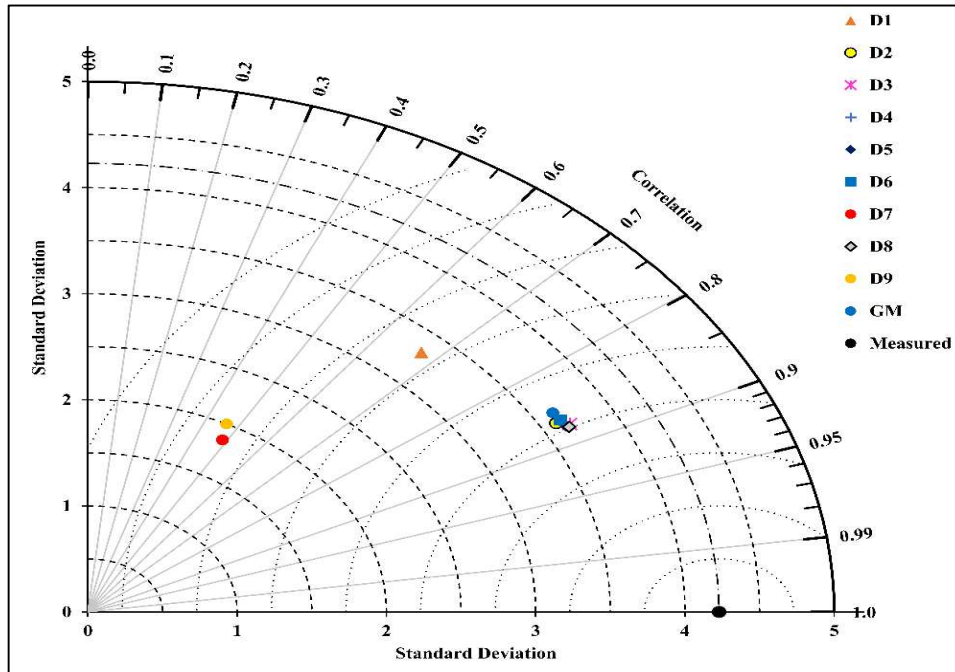
Fig. 9. Taylor diagram for the used models and the general model (GM) of DGSR in Lvsi (Zone IV).

3.5. Zone V

For zone V, through twelve stations in this zone, model D4 is more accurate and stable. For this model, r values are between 0.692 and 0.934, RMSE values vary between 1.850 and 2.116 MJ/m^2 day, MABE range between 1.455 to 1.708 MJ/m^2 day, U95 are 5.089 to 5.865 and R^2 are 0.646 to 0.907. Models D6, D8, D2 exhibit good and similar performance followed by model D5. Whereas models D9 and D1 performed a worst performance where their r and R^2 values are 0 to 0.933 and 0 to 0.911 for D9 model, respectively, and 0.384 to 0.926 and 0.165 to 0.903 for D1 model, respectively. Although, model D7 performs the best in most stations, his performance is dropped down in Luzhou station, where r and R^2 values are 0.292 and 0.241, respectively. Hence, model D4 is selected as the best model to estimate DGSR in this zone. For this purpose, daily data of 12 stations are averaged and employed to obtain the coefficients of D4 model as listed in

Table 3. It can be seen that the general model performs best, it has R^2 of 0.912, RMSE of 1.309 MJ/m^2 day, MABE of 1.030 MJ/m^2 day, U_{95} of 3.557 and r of 0.944.

To show the difference between measured and predicted data, Luzhou station is selected for this purpose as shown in **Fig. 9**. Taylor diagram **(Fig. 10)** approve that D9, D7, and D1 models are the furthest representative points from the measured point, which illustrates that it cannot be dependable for estimating the DGSR in this zone especially D9 model. On other hand, models D3, D4, D5, D6, and D8 have a similar and good performance, which their representative points lie in the same position and close to the measured point.
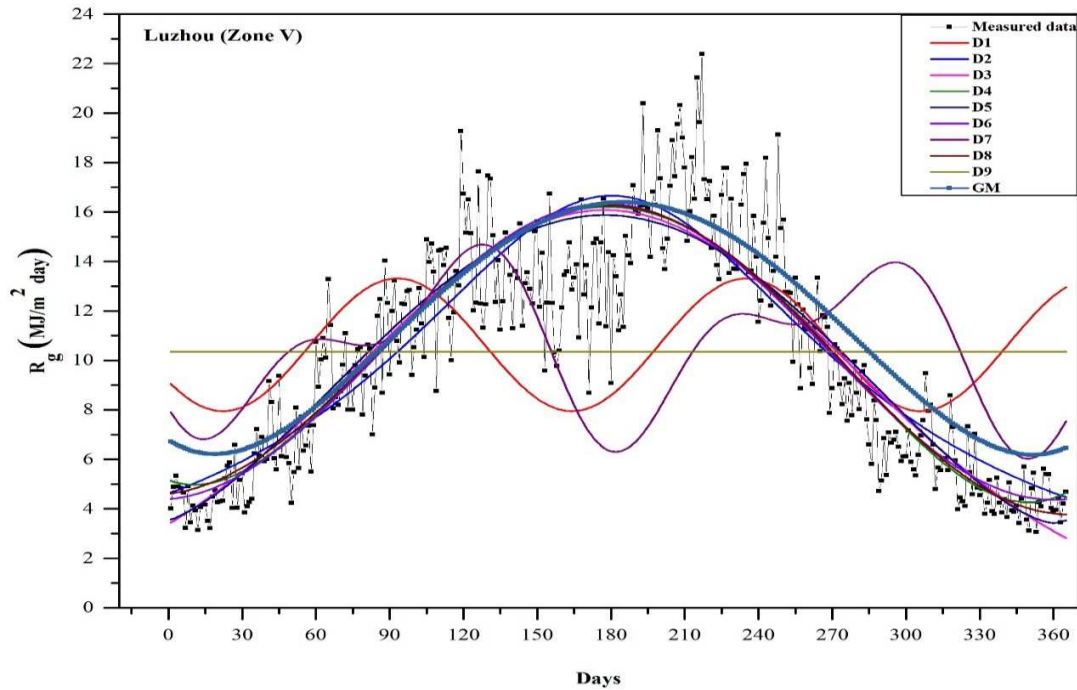

Fig. 10. The observed data compared with the predicted data by the used model and the best model (GM) in Luzhou (Zone V).
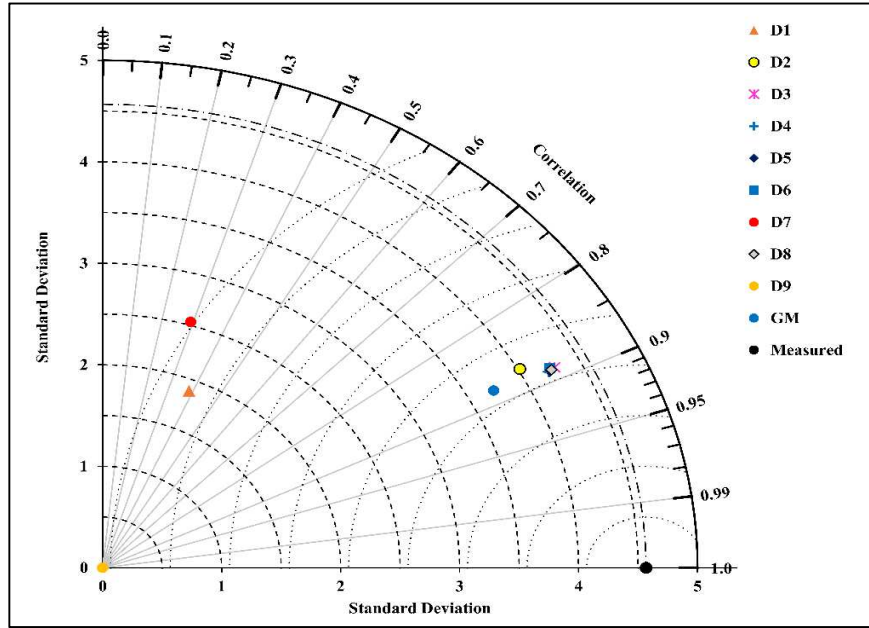
Fig. 11. Taylor diagram for the used models and the general model (GM) of DGSR in Luzhou (zone V).

Table 3

The empirical coefficients and statistical indices of general models for each solar zone across China.

Zone	General Model (GM)	a_1	a_2	a_3	a_4	a_5	a_6	a_7	R^2	RMSE	MABE	U_{95}	r
I	D7	17.743	4.551	-0.976	-14.597	-0.982	1.952	-0.17	0.984	0.780	0.586	2.062	0.976
II	D8	16.841	-49.939	0.004	-8.271	1.032	-	-	0.991	0.783	0.622	2.160	0.992
III	D3	20.538	168.631	109.313	-	-	-	-	0.980	1.042	0.835	2.743	0.983
IV	D8	13.009	-0.523	1.269	-4.779	.977	-	-	0.950	1.241	1.0	3.366	0.947
V	D4	6.791	-0.065	1.999×10^{-3}	-9.850×10^{-6}	1.330×10^{-8}	-	-	0.912	1.309	1.030	3.557	0.944

4. Conclusion

In this study, a DGSR data of 84 zones across China are employed to divided China into five solar climatic zones and evaluate the performance of eight day of the year based model, existed in the literature, for estimating DGSR across China, and developing a general model for each

25

solar zone in China. Polynomial and nonlinear regression methods are used for generating the empirical coefficients of used models. For evaluating the performance statistical indicators are employed such as regression coefficient (R^2), correlation coefficient (r), root mean square error (RMSE), mean absolute bias error (MABE) and the uncertainty based on a 95% confidence level (U_{95}), and Taylor diagram. The results indicate for zone I, the trigonometric model, hybrid sine, and cosine wave model (D7) yields the best and it can be the general model for this zone. For zones II and IV, the hybrid sine and cosine wave model (D8) is the superior performance among others, therefore, it is identified to be the general model for these two zones. The Gaussian model (D3) is selected to be the general model for zone III due to its good and stable performance across 18 stations in this zone. For the last zone, zone V, the fourth polynomial degree model (D4) yields best performance all over zone V, hence, it can be used as a general model to predict the DGSR in this zone.

The performance of all used models is higher in zones which receive high values of DGSR (zone I) and lower in zones which receive a lower amount of solar radiation (zone II). The general day of the year based models which developed and evaluated in this study can be employed to estimate the daily global solar radiation in China with a good accuracy without needing any meteorological data. Studying the changing in the daily solar radiation data obtained from the models for each site is not considered in this type of models, that need to be studied in the future. Therefore, using the geographical parameters and/or the daily extraterrestrial solar radiation on a horizontal surface of the site of interest may be improve model performance and makes it changeable with the site and year.

References

Al-Salaymeh, A., 2006. Modelling of global daily solar radiation on horizontal surfaces for Amman city. Emirates Journal for Engineering Research 11 49-56.

Allen, R.G., 1997. Self-calibrating method for estimating solar radiation from air temperature. Journal of Hydrologic engineering 2, 56-67.

Angstrom, A., 1924. Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. Quarterly Journal of the Royal Meteorological Society 50, 121-126.

Aoun, N., Bouchouicha, K., 2017. Estimating daily global solar radiation by day of the year in Algeria. The European Physical Journal Plus 132.

Bahel, V., Bakhsh, H., Srinivasan, R., 1987. A correlation for estimation of global solar radiation. Energy 12, 131-135.

Bulut, H., Büyükalaca, O., 2007. Simple model for the generation of daily global solar-radiation data in Turkey. Applied Energy 84, 477-491.

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. Renewable Energy 103, 708-720.

Chen, J.-L., Li, G.-S., 2013. Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. International Journal of Climatology 33, 487-498.

Chen, R., Ersi, K., Yang, J., Lu, S., Zhao, W., 2004. Validation of five global radiation models with measured daily data in China. Energy Conversion and Management 45, 1759-1769.

Chen, R., Kang, E., Lu, S., Yang, J., Ji, X., Zhang, Z., Zhang, J., 2006. New methods to estimate global radiation based on meteorological data in China. Energy Conversion and Management 47, 2991-2998.

Dizaji, H.B., Hosseini, H., 2018. A review of material screening in pure and mixed-metal oxide thermochemical energy storage (TCES) systems for concentrated solar power (CSP) applications. Renewable and Sustainable Energy Reviews 98, 9-26.

Domrös, M., Gongbing, P., 1988. Climate Zones of China, in: Domrös, M., Gongbing, P. (Eds.), The Climate of China. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 258-280.

Gouda, S.G., Hussein, Z., Luo, S., Wang, P., Cao, H., Yuan, Q., 2018. Empirical models for estimating global solar radiation in Wuhan City, China. The European Physical Journal Plus 133, 517.

Gouda, S.G., Yuan, Q., 2018a. Comments on "Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China'', by Liu et al. [Energy Conversion and Management, 2017.10.043.]. Energy Conversion and Management 168, 651-652.

Gouda, S.G., Yuan, Q., 2018b. Corrigendum to 'Comments on "Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in

China", by Liu et al. [Energy Convers. Manage. (2017) 10.043.]' [Energy Convers. Manage. (in press)]. Energy Conversion and Management.

Gouda, S.G., Yuan, Q., 2018c. Reply to "Reply to [Comments on "Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China" [Energy Convers. Manage. (2017) 10.043] [Energy Convers. Manage. (2018) 05.051], by Liu et al.". Energy Conversion and Management.

Gueymard, C.A., 2014. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. Renewable and Sustainable Energy Reviews 39, 1024-1034.

Hassan, G.E., Youssef, M.E., Ali, M.A., Mohamed, Z.E., Shehata, A.I., 2016. Performance assessment of different day-of-the-year-based models for estimating global solar radiation - Case study: Egypt. Journal of Atmospheric and Solar-Terrestrial Physics 149, 69-80.

Jemaa, A.B.E.N., Rafa, S., Essounbouli, N., Hamzaoui, A., Hnaien, F., Yalaoui, F., 2013. Estimation of Global Solar Radiation Using Three Simple Methods. Energy Procedia 42, 406-415.

Kaplanis, S., Kaplani, E., 2007. A model to predict expected mean and stochastic hourly global solar radiation I(h;nj) values. Renewable Energy 32, 1414-1425.

Kaplanis, S., Kumar, J., Kaplani, E., 2016. On a universal model for the prediction of the daily global solar radiation. Renewable Energy 91, 178-188.

Lam, J.C., Tsang, C.L., Yang, L., Li, D.H.W., 2005. Weather data analysis and design implications for different climatic zones in China. Building and Environment 40, 277-296.

Lau, C.C.S., Lam, J.C., Yang, L., 2007. Climate classification and passive solar design implications in China. Energy Conversion and Management 48, 2006-2015.

Lemos, L.F.L., Starke, A.R., Boland, J., Cardemil, J.M., Machado, R.D., Colle, S., 2017. Assessment of solar radiation components in Brazil using the BRL model. Renewable Energy 108, 569-580.

Li, H., Cao, F., Wang, X., Ma, W., 2014. A temperature-based model for estimating monthly average daily global solar radiation in China. The Scientific World Journal 2014, 128754.

Li, H., Ma, W., Lian, Y., Wang, X., 2010a. Estimating daily global solar radiation by day of year in China. Applied Energy 87, 3011-3017.

Li, M.-F., Fan, L., Liu, H.-B., Guo, P.-T., Wu, W., 2013a. A general model for estimation of daily global solar radiation using air temperatures and site geographic parameters in Southwest China. Journal of Atmospheric and Solar-Terrestrial Physics 92, 145-150.

Li, M.-F., Liu, H.-B., Guo, P.-T., Wu, W., 2010b. Estimation of daily solar radiation from routinely observed meteorological data in Chongqing, China. Energy Conversion and Management 51, 2575-2579.

Li, M.-F., Tang, X.-P., Wu, W., Liu, H.-B., 2013b. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Conversion and Management 70, 139-148.

Liu, L.-q., Wang, Z.-x., Zhang, H.-q., Xue, Y.-c., 2010. Solar energy development in China—A review. Renewable and Sustainable Energy Reviews 14, 301-311.

Liu, X., Mei, X., Li, Y., Wang, Q., Zhang, Y., Porter, J.R., 2009. Variation in reference crop evapotranspiration caused by the Ångström–Prescott coefficient: Locally calibrated versus the FAO recommended. Agricultural Water Management 96, 1137-1145.

Liu, Y., Zhou, Y., Wang, D., Wang, Y., Li, Y., Zhu, Y., 2017. Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China. Energy Conversion and Management 154, 168-179.

Liu, Y., Zhou, Y., Wang, D., Wang, Y., Li, Y., Zhu, Y., 2018. Reply to "Comments on ["Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China" [Energy Convers. Manage. 154 (2017) 168–179] by Liu et al.]". Energy Conversion and Management 168, 653-654.

Mohammadi, K., Shamshirband, S., Tong, C.W., Alam, K.A., Petković, D., 2015. Potential of adaptive neuro-fuzzy system for prediction of daily global solar radiation by day of the year. Energy Conversion and Management 93, 406-413.

National Bureau of Statistics of People's Republic of China. Total consumption of energy and its composition, <http://www.stats.gov.cn/tjsj/ndsj/2010/indexch.htm> (accessed 21 January 2019).

Ögelman, H., Ecevit, A., Tasdemiroğlu, E., 1984. A new method for estimating solar radiation from bright sunshine data. Solar Energy 33, 619-625.

Ouyang, L., Sun, L., Yuan, Y., Cao, X., Xiang, B., 2017. Optimum connection modes for photovoltaic thermal collectors in different radiation zones of China. Applied Thermal Engineering 122, 661-672.

Prescott, J., 1940. Evaporation from a water surface in relation to solar radiation. Transactions of the Royal Society of South Australia 64, 114-118.

Quej, V.H., Almorox, J., Ibrakhimov, M., Saito, L., 2017. Estimating daily global solar radiation by day of the year in six cities located in the Yucatán Peninsula, Mexico. Journal of Cleaner Production 141, 75-82.

Rietveld, M., 1978. A new method for estimating the regression coefficients in the formula relating solar radiation to sunshine. Agricultural Meteorology 19, 243-252.

Tang, W., Qin, J., Yang, K., Niu, X., Zhang, X., Yu, Y., Zhu, X., 2013. Reconstruction of daily photosynthetically active radiation and its trends over China. Journal of Geophysical Research: Atmospheres 118, 13,292-213,302.

Tang, W., Yang, K., He, J., Qin, J., 2010. Quality control and estimation of global solar radiation in China. Solar Energy 84, 466-475.

Xie, H., Zhang, C., Hao, B., Liu, S., Zou, K., 2012. Review of solar obligations in China. Renewable and Sustainable Energy Reviews 16, 113-122.

Yang, K., Huang, G.W., Tamai, N., 2001. A hybrid model for estimating global solar radiation. Solar Energy 70, 13-22.

Yang, K., Koike, T., Ye, B., 2006. Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets. Agricultural and Forest Meteorology 137, 43-55.

Yao, W., Zhang, C., Wang, X., Sheng, J., Zhu, Y., Zhang, S., 2017. The research of new daily diffuse solar radiation models modified by air quality index (AQI) in the region with heavy fog and haze. Energy Conversion and Management 139, 140-150.

Zang, H., Xu, Q., Bian, H., 2012. Generation of typical solar radiation data for different climates of China. Energy 38, 236-248.

Zhao, N., Zeng, X., Han, S., 2013. Solar radiation estimation using sunshine hour and air pollution index in China. Energy Conversion and Management 76, 846-851.

Zou, L., Wang, L., Lin, A., Zhu, H., Peng, Y., Zhao, Z., 2016. Estimation of global solar radiation using an artificial neural network based on an interpolation technique in southeast China. Journal of Atmospheric and Solar-Terrestrial Physics 146, 110-122.