

Evaluation of the performance of different models for predicting direct normal solar irradiance

Danny H W Li^a, Wenqiang Chen^{a*}, Shuyang Li^a

^a Building Energy Research Group, Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China

Abstract

Solar energy is considered as a clear and sustainable energy resource, and the application of solar energy includes electric power generation and solar concentrators. The precise estimation of solar irradiance plays an important role in evaluating the performance of active solar energy utilizations such as concentrator photovoltaic systems. While global solar irradiance received by a horizontal surface can be easily measured, the availability of direct normal irradiance (DNI) is quite limited. Many models for predicting DNI have been developed and a number of them provided a satisfactory performance. However, it may be difficult for users to efficiently pick up the appropriate models that can be applied to their projects. This study analyses the solar irradiance data in Hong Kong based on continuous measurements and evaluates the performance of three empirical and machine learning models. The accuracy of individual approaches was evaluated using measured Hong Kong data. The results would be helpful to select suitable DNI prediction models for various applications.

Keywords: Direct normal irradiance (DNI), Prediction models, Model validation

Nomenclature

α	Solar altitude angle, degrees	K_d	Diffuse component, dimensionless
cl	Cloud amount (%)	K_t	Clearness index, dimensionless
DHB	Direct solar irradiance on a horizontal surface, W/m^2	K_i'	Zenith angle-dependent clearness index, dimensionless
DHI	Diffuse solar irradiance on a horizontal surface, W/m^2	T_a	Air temperature, °C
DNI	Direct normal solar irradiance, W/m^2	vis	Visibility, m
DNI_{disc}	Direct normal solar irradiance estimated by the DISC model, W/m^2	X	Coefficient function in Perez model, dimensionless
GHI	Global solar irradiance on a horizontal surface, W/m^2	Z	Solar zenith angle, degrees
GHE	Extraterrestrial solar irradiance on a horizontal surface, W/m^2	$\Delta K_t'$	Stability index, dimensionless
K	Hourly diffuse fraction, dimensionless	ψ	Persistence predictor in BRL model, dimensionless

1. Introduction

Solar energy has been considered as the most popular renewable energy resource for generating power and collecting thermal resources and is applied in many areas worldwide compared to others. The widespread concern is how to efficiently utilize the solar energy and accurately estimate the amount of solar radiation that falls on the solar equipment. The design and development of solar photovoltaic (PV) systems require the knowledge of variation and maximum utilization of solar radiation falling on it [1]. For instance, the design and analysis of concentrating photovoltaic (CPV) systems require knowledge of direct normal solar radiation [2]. Thus, solar radiation data are necessary to the design and evaluation of

* Manuscript received March 12, 2018; revised January 15, 2019.

Corresponding author. Tel.: +852-6737-5667; E-mail address: chenwenqiang123@sina.cn .

doi: 10.12720/sgce.8.2.231-238

active solar energy facilities and passive energy-efficient building designs [3].

Although long-term data measurement is an effective and accurate way of setting up databases of solar radiation, ground-based measurements of these data are available for not many locations. For example, there are 756 meteorological stations altogether in China, among which only 122 of them have records of global solar radiation [4], and fewer of them have records of direct or diffuse component. Besides, the solar radiation station is quite expensive to build and the maintenance of the devices is time-consuming. Consequently, predicting solar radiation data by correlating it with other easily measured meteorological parameters, such as solar altitude, ambient temperature, cloud cover, humidity et al., is an alternative method to obtain desired solar radiation data. Predicting solar irradiance has been an important topic in renewable energy application fields.

The radiation arriving on the ground directly in line from the solar disk is called direct normal or beam radiation, and its instantaneous flux received by unit area is direct normal irradiance (DNI). Its horizontal component is combined with the scattered diffuse irradiance (DHI) forms the global solar irradiance (GHI) on the horizontal plane. The relationship of these three components can be described by Equation (1):

$$\text{GHI} = \text{DNI} * \sin(\alpha) + \text{DHI} \quad (1)$$

where ‘ α ’ is the solar altitude. While direct and diffuse irradiance is needed in the calculations of heating and cooling loads and the design of flat-plate collectors, only a knowledge of direct irradiance is needed in designing many concentrating systems [5]. A lot of models for predicting the global solar radiation or direct/diffuse radiation on a horizontal surface have been developed by many researchers using empirical methods [6] and machine learning methods which developed rapidly in recent years [7]. Although many models have been provided as a good performance in original research, it is difficult for others to choose which model would be appropriate for their projects. Three models namely the Perez model, the BRL model and the LOU model are assessed in this paper by evaluating their accuracy for predicting DNI against measured solar irradiance data of Hong Kong. The results would be helpful for selecting suitable DNI prediction models for various applications.

2. Solar radiation Data of Hong Kong

2.1. Monthly average daily and hourly radiation

With a subtropical and lighter polluted climate, the availability of solar energy in Hong Kong is very sufficient in spite of the high density of urban area. The positive utilization of solar energy plays an important role in the sustainable development of the city. The Hong Kong Observation (HKO) has measured the hourly global solar radiation for almost 40 years, but no records of direct normal and diffuse radiation until 2008. A station was established to measure the horizontal global, sky diffuse and direct normal solar radiation in Kau Sai Chau, Hong Kong (22.3° N, 114.3° E). The data collection started just before sunrise and ended after sunset each day. The measurements were recorded hourly in local civil time. The hourly solar irradiance data from 2012 Jan. 1st to 2014 Dec. 31th was analyzed in this study and was used to evaluate the performance of the three mentioned models.

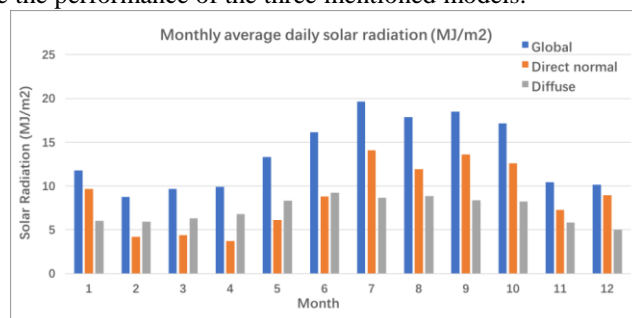


Fig. 1. Monthly average daily direct normal, global and diffuse solar radiation.

Fig. 1 presents the monthly average daily direct normal, global and diffuse solar radiation according to the statistical hourly solar radiation data of Year 2012 to 2014. Totally, the daily global solar radiation received on a horizontal surface fluctuates from month to month, and the amount ranging from 8.77 MJ/m² in February to 19.62MJ/m² in December. The seasonal fluctuation can be easily detected from the figure, which is in line with the actual climate condition of Hong Kong. The global radiation observed in June to October are higher than other months, which mainly affected by the high solar altitude and long daytime. The low values recorded from December to April are mainly caused by the low solar altitude and the humid and cloudy weather. The similar trends can also be found in direct normal and diffuse solar radiation. It can be seen that the direct normal exceeds the diffuse component in January and July to December. Particularly, the direct normal radiation was almost as twice as the diffuse radiation in December, and it reaches about 90% of the global radiation at a value of 9.0MJ/m². It can be concluded that solar radiation availability depends on the solar altitude and prevailing weather conditions.

To further examine the direct normal solar radiation data in December, daily average hourly solar irradiance (averaged data of 2012-2014) is analyzed as shown in Fig. 2. While monthly average data are not very precise records for solar radiation, hourly data would be more appropriate for solar heating, solar power system and building designs (e.g. cooling loads). The daily average hourly DNI, GHI, DHI plotted in Fig. 2 gives the detailed change of the intensity of solar irradiance during a day. The DNI largely exceeded the DHI for all hours and was close to the GHI. Specially, the DNI exceeded the GHI at around 8 - 10:00 and 16 - 18:00 (local civil time). The large components of DNI indicate that the climate conditions in December of 2012 to 2014 were mainly of cloudless days constitute a major part of the month. It also implies the possibility of utilizing sun-track devices in solar systems to maximize the output of thermal or power, especially in locations with dry and cool climate conditions.

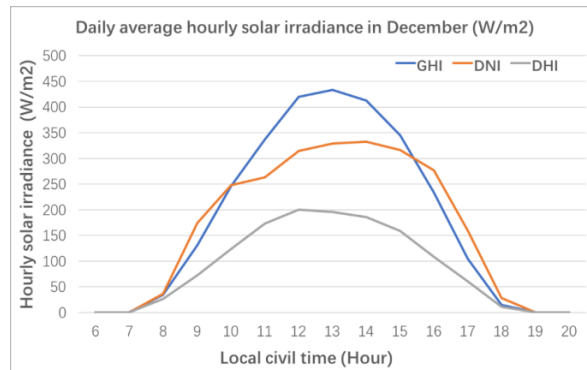


Fig. 2. Daily average hourly solar irradiance in Decembe

2.2. Climate indicators

Fraction expressions of the relationship of direct, diffuse and global solar irradiance are commonly used in solar energy studies and they can be found in many models which have been tested by practice. There are mainly three climate indicators namely: clearness index - K_t , diffuse fraction - K , and diffuse coefficient - K_d . The criteria for determining sky clearness can be illustrated by the solar irradiance correlations between these three indicators [8]. Their expressions of these three indicators are shown in Equation (2-4),

$$K_t = \text{GHI} / \text{GHE} \quad (2)$$

$$K = \text{DHI} / \text{GHI} \quad (3)$$

$$K_d = \text{DHI} / \text{GHE} \quad (4)$$

where GHE is the hourly extraterrestrial solar irradiance received by a horizontal surface.

The clearness index K_t is the most commonly indicators used in both global solar radiation models and direct or diffuse solar irradiance prediction models. It displays the percentage of solar irradiance received by a horizontal surface other than being absorbed or reflected by the atmosphere. The most common approach is to correlate the clearness index K_t with other meteorological data used as predictors in simple regression equation. Fig. 3 depicts the frequency of occurrence of K_t based on the hourly measured global irradiance data from 2012-2014. The frequency of K_t appears evenly distributed between 0-10%, with two peaks centered at 0.12 (9.5%) and 0.78 (0.69%). Low K_t means cloudy skies and high values stand for clear days. Apparently, the accumulated frequency of K_t below 0.5 is larger than that above 0.5. It can be derived from Fig. 3 that the occurrence of partly cloudy skies exceeds the clear skies.

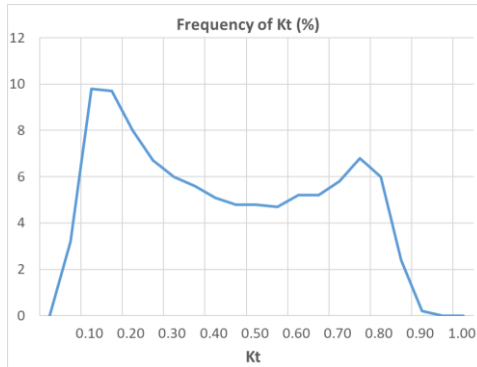


Fig. 3. Frequency of occurrence for K_t

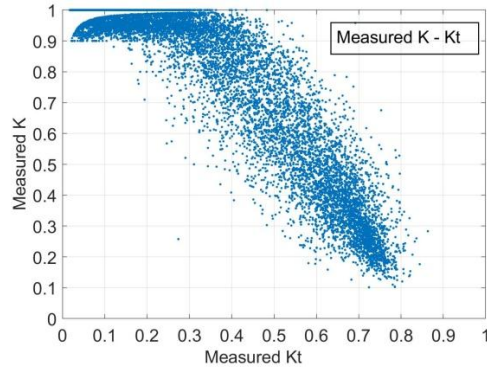


Fig. 4. Relationship between K and K_t

3. Models and Data used for Performance Evaluation

3.1. Perez model

The first model used for comparison is the Perez model. In 1992, Perez et al. [9] developed a direct normal radiation model which is a 4-dimensional coefficient matrix model based on the Maxwell's model [10] which was modified by an insolation-dependent factor X . The Perez model has been recognized as one of the most accurate models for predicting direct normal radiation. It is a dynamic global-to-direct irradiance conversion model which can estimate the DNI from which the diffuse irradiance. It is a 4-dimensional model with the following variables:

$$I = I_{\text{disc}} \cdot X(K_t', Z, W, \Delta K_t') \quad (5)$$

where DNI_{disc} is the direct normal irradiance estimated by the DISC model (It's a function of global irradiance and solar zenith angle) and $X(K_t', Z, W, \Delta K_t')$ is a coefficient function of the four insolation condition parameters, where Z is the solar zenith angle, K_t' is a zenith angle-dependent expression of the clearness Index, $\Delta K_t'$ is a stability index, and W is the atmospheric perception water.

3.2. Boland-Ridley-Lauret (BRL) model

The BRL model given by Ridley et al. [11] determines the diffuse fraction K based on K_t and other four possible predictors, which can be shown in Equation (6). Although the model was originally to predict the diffuse fraction, the authors provided the evidence that using the BRL model to estimate diffuse solar fraction, and then calculate DNI [12]. To obtain the DNI from the BRL model, the steps are applied as follows. Firstly, use the BRL model to obtain the diffuse fraction - K . Afterwards, using Equation (7), the direct normal irradiance DNI can be obtained. The original BRL model uses values of the lagged clearness index ' ψ ' as a predictor. Since the model is used to predict DNI, it takes the 'persistence' (ψ) as an extra predictor which is an average of both a lag and a lead of the clearness index,

the persistence predictor is shown in Equation (8).

$$K = \frac{1}{1 + e^{-5.38 + 6.63k_t + 0.006AST - 0.007\alpha + 1.75K_t + 1.31\psi}} \tag{6}$$

$$DNI = \frac{GHI - (K \cdot GHI)}{\sin(\alpha)} \tag{7}$$

$$\psi = \begin{cases} \frac{k_{t-1} + k_{t+1}}{2}, & \text{sunrise} < t < \text{sunset} \\ k_{t+1}, & t = \text{sunrise} \\ k_{t-1}, & t = \text{sunset} \end{cases} \tag{8}$$

3.3. LOU model

LOU et al. [13] used the Boosted Regression Tree (BRT) algorithm to generate an ensemble of Regression Trees (RT) that correlates hourly diffuse fraction K with the relevant predictors. A simple equation that is readily accessible to use was given by applying the technique of logistic regression as shown in Equation (9). The predictors used in this model are listed in Table 1.

$$K = \frac{1}{1 + \exp[f_0 + \sum_{i=1}^n (f_i X_i)]} \tag{9}$$

Table 1. Predictors being used in the LOU model

	Variable	Name
X1	K_t	Clearness index
X2	$\sin(\alpha)$	Sine of solar altitude angle: $\sin(\alpha)$
X3	$(T_a - 25)/5$	Air temperature
X4 (in %)	<i>cld</i>	Cloud amount (%)
X5	$(vis - 14,575)/7653$	Visibility

3.4. Data quality control

The hourly solar irradiance data from 2012 Jan. 1st to 2014 Dec. 31th were used for the study. To eliminate spurious data and erroneous measurements, quality-controls tests based on the CIE guidance [14] were adopted. Table 2 summarizes the tests and results. Under Level 0 test, 1390 hourly irradiance data with a solar altitude of less than 4° or the GHI less than 20 W/m² were rejected. Then, 4 and 1224 hourly data were excluded under Level 1 and Level 2 tests, respectively. After all the tests, 10608 hourly solar irradiance data sets were retained for the analysis and Table 2 summarizes the number of accepted and rejected data for each process. Fig. 5 shows the relationship between the measured global irradiance (GHI) against direct and diffuse (DHB+DHI) irradiance data after the quality control process.

Table 2. Quantity control process according to CIE test levels (2012-2014 HKO data).

CIE test level	Criterion	Number of data record (Hour)	
		Rejected	Accepted
Total	Solar angle > 0°		13226
Level 0	Solar angle > 4° ; GHI > 20 W/m ²	1390	11836
Level 1	0 < GHI < 1.2*GHE; 0 < DHI < 0.8*GHE; 0 ≤ DNI < GHE	4	11832
Level 2	0.9 * (GHI - DNI*cosZ) < GHI < 1.1*(GHI - DNI*cosZ); DHI ≤ GHI	1224	10608

4. Results

The performances of each model for predicting hourly DNI were assessed by statistical indicators: the mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE) and coefficient of determination (R^2). The four error indicators (with percentage MAE, MBE and RMSE) were calculated and the results are summarized in Table 3.

Fig. 6 gives the scatter plot of the predicted DNI against measured DNI. It can be seen that the Perez model and BRL model are more scattered than LOU model, which can also be derived from the MAE and RMSE values. The Perez model has the highest MAE and RMSE values of 44.75 and 73.53 W/m^2 , which means it has the largest deviation. According to Fig. 6 (a), it can be seen that the model performs better when DNI is higher than 600 W/m^2 than when DNI is lower than that, indicating that Perez model may be more suitable for predicting DNI in clear days. The MBEs of all three models are positive and are more than 14 W/m^2 , indicating that they all overestimate the DNI, and in other words, underestimating the amount of DHI.

The BRL model has the MAE value of 42.75 W/m^2 (21.4%) and the RMSE is 71.72 W/m^2 (35.9%), which performs a little better than Perez model. As BRL and LOU model are originally developed for predicting diffuse fraction K_t , a comparable plot of measured and predicted relationship between K_t and K_t are shown in Fig. 7. It can be used to check whether using a diffuse fraction model is sufficient for estimating DNI from global irradiance through the chain GHI-DHI-DNI. It can be depicted from Fig.7 that the K_t values predicted by BRL model are of less scattered and more compacted in the center of measured K_t - K_t data points, while LOU model performs better as it is close to the overall layout of actual situation. The BRL model may not be suitable for predicting DNI in locations with more changeable weather conditions. Overall, the LOU model has the best performance with the lowest MAE and RMSE values of 35.16 W/m^2 (17.6%) and 56.40 W/m^2 (28.2%) respectively, and with the highest R^2 value of 0.9547.

Table 3. Quantity control process according to CIE test levels (2012-2014 HKO data).

Model	MAE (W/m^2)	%MAE (%)	MBE (W/m^2)	%MBE (%)	RMSE (W/m^2)	%RMSE (%)	R^2
Perez	44.75	22.4%	+21.92	+11.0%	73.53	36.8%	0.9304
BRL	42.75	21.4%	+20.74	+10.4%	71.72	35.9%	0.9349
LOU	35.16	17.6%	+14.87	+7.4%	56.40	28.2%	0.9547

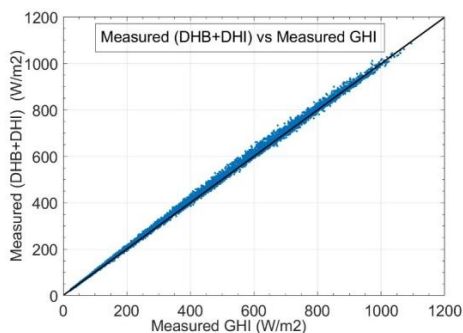


Fig. 5 Measured Irradiance data (W/m^2).

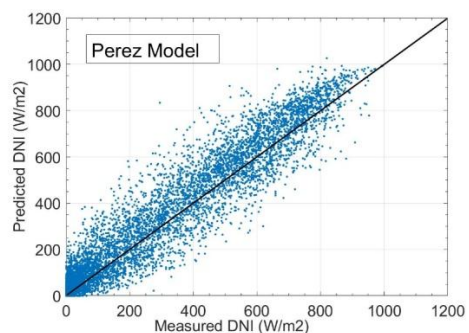


Fig. 6 (a) Predicted DNI by Perez Model (W/m^2);

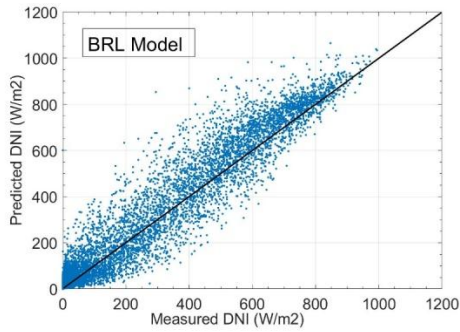
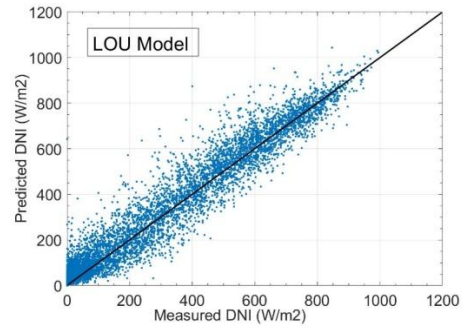
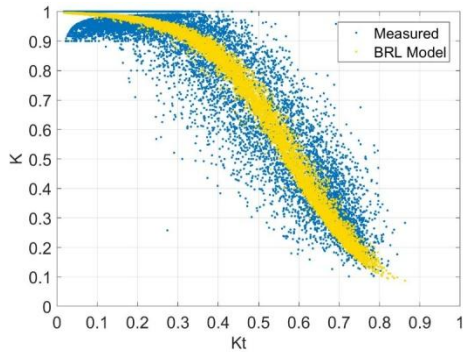
Fig. 6 (b). Predicted DNI by BRL Model (W/m²);Fig. 6 (c). Predicted DNI by LOU Model (W/m²).

Fig. 7 (a).

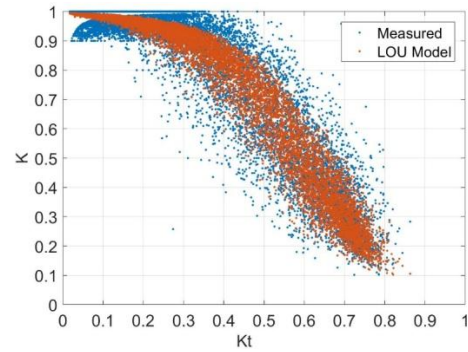


Fig. 7 (b).

Fig. 7. Measured and predicted relationship between K and K_t by (a) BRL model; (b) LOU model.

5. Discussion

As most of the existing solar radiation models are mainly for predicting global solar radiation, especially for monthly average or daily average global solar radiation. It seems that more studies for predicting direct normal/beam or diffuse solar radiation in short time scales such as one-hour or 10 minutes should be developed. Many empirical models are developed from data measured at specific locations, the coefficients in the models may be site dependent and thus their usability must be verified. Models need to be developed using latitude, longitude, altitude, extraterrestrial radiation as input parameters, which would be useful for those locations where no meteorological data are available. The machine learning techniques should be widely used to solve real problems in the case where classical methods are insufficient. Solar irradiance prediction is one of the domains of which the techniques were successfully used. Models derived from machine learning techniques tend to perform better than empirical models. More suitable machine learning approaches to predict DNI are required.

6. Conclusion

An analysis of horizontal global, diffuse and direct normal solar radiation was conducted based on the measured data in Hong Kong. The monthly average daily solar radiation and daily average hourly solar irradiance was elaborated. The large components of DNI indicates that the main climate condition in December of 2012 to 2014 were cloudless skies. The climate indicators K and K_t based on the hourly measured irradiance data were also analyzed. Heavily overcast sky conditions were quite common in Hong Kong. The performances of the three models for predicting DNI were evaluated based on four statistical indicators. The models were performed not very accurate for predicting long-term hourly DNI (e.g. 2012-2014), with the %MAEs of more than 17.6% and the %RMSEs exceeding 28.2%. The LOU

model gave the lowest MAE, RMSE and the highest R^2 , indicating that it performs best among all models. The findings would be helpful to select suitable DNI prediction models for various applications, at least for locations with similar latitude and climate as Hong Kong.

Acknowledgements

The work described in this paper was fully supported by a Strategic Research Grant from the City University of Hong Kong (Project no. 7004629) and Wenqiang Chen is supported by a City University of Hong Kong studentship.

References

- [1] Wan C, Zhao J, Song Y, Xu Z, Lin J, Hu Z. Photovoltaic and solar power forecasting for smart grid energy management. *CSEE J Power Energy System*, 2015(1):38–46.
- [2] Vossier A, Chemisana D, Flamant G, Dollet A. Very high fluxes for concentrating photovoltaic: considerations from simple experiments and modeling. *Renewable Energy*, 2012;38:31–9.
- [3] DHW Li, L Yang, JC Lam. Zero energy buildings and sustainable development implications – A review. *Energy* 2013; 54: 1–10.
- [4] Zang H, Xu Q, Bian H. Generation of typical solar radiation data for different climates of China. *Energy* 2012;38:236–48.
- [5] Iqbal M. An introduction to solar radiation. *ACADEMIC PRESS*, 1983.
- [6] Wong LT, Chow WK. Solar radiation mode, *Applied Energy*, 2001; 69: 191–224.
- [7] Mellit A, Kalogirou SA. Artificial intelligence techniques for photovoltaic applications: a review. *Progress in Energy and Combustion Science*, 2008; 34: 547–632.
- [8] Lam JC, Li DHW. Correlation between global solar radiation and its direct and diffuse components. *Build. Environ*, 1996; 31: 527–535.
- [9] Perez RR, Ineichen P, Maxwell EL. Dynamic global-to-direct irradiance conversion models. *ASHRAE Transactions*, 1992;(98):354–369.
- [10] Maxwell EL. A quasi-physical model for converting hourly global horizontal to direct normal insolation. *Technical report. Golden (CO): Solar Energy Research Institute*, 1987. Report No. SERI/TR-215-30871.
- [11] Ridley B, Boland J, Lauret P. Modelling of diffuse solar fraction with multiple predictors. *Renewable Energy* 2010; 35: 478–483.
- [12] John B, Jing H, Barbara R. Decomposing global solar radiation into its direct and diffuse components. *Renewable and Sustainable Energy Reviews* 2013;28:749–756.
- [13] Siwei Lou, Danny H.W. Li, Joseph C. Lam, Wilco W.H. Chan. Prediction of diffuse solar irradiance using machine learning and multivariable regression. *Applied Energy*, 2016;181: 367–374.
- [14] CIE. Guide to recommended practice of daylight measurement, *Technical report. Vienna, Austria: Central Bureau of the CIE*, 1994. Report No. CIE 108-1994.