Validation of empirical models to estimate diffuse radiation at Badajoz (Spain)

G. Sanchez[1], M.L. Cancillo[1], A. Serrano[1], J.A. Garcia[1]

Abstract — To achieve an accurate knowledge of the solar global radiation and its partitioning beam-to-diffuse is very important for the renewable energy sector. Although solar global radiation is frequently measured at most radiometric stations, diffuse radiation measurements are less frequent because the shadow band and sun tracker used to measure it required an economic inversion and a maintenance which is not always possible. In these cases when no measurements of diffuse radiation is available, it must be estimated by means of models. Thus, many authors have developed empirical models that allow the estimation of the diffuse component from different meteorological and radiometric magnitudes. The main objective of this paper is to validate and adapt five empirical models using one-minute data measured at the radiometric station installed in Badajoz (south-west of Spain). Results indicate that all models generally account for the general tendency, being the models proposed by Ruiz-Arias et al. and by Spencer the ones showing the best performance according to the experimental data.

Keywords — diffuse radiation, global radiation, model for diffuse radiation, partitioning direct-diffuse

1 INTRODUCTION

The knowledge of the partitioning of solar radiation into their diffuse and direct components is essential for the design of solar collectors, fixed and mobile, in order to improve the collection of solar energy on tilted surfaces. From this information, the available energy to be used and transformed into electrical or thermal energy can be calculated. There are many meteorological stations around the world where the solar global radiation, as well as its values in different spectral subintervals (ultraviolet, visible and infrared), is measured. However, its directional components (direct and diffuse) are only available in a few percentage of them, being necessary to be estimated by means of models in a high number of locations.

In the literature, there can be found many empirical models which estimate the proportion of diffuse radiation, k_d, as a function of different radiometric and meteorological variables such as the sunshine hours [1] or the air temperature and humidity [2]. It has been even studied the relationship of this magnitude with so many as 28 different parameters [2]. These models have been developed for different time intervals (hourly, monthly, etc). However, the most studied relationship of the proportion of diffuse radiation, k_d, is its dependence on the clearness index, k_t. These type of models are interesting because they allow a good estimate of the diffuse component in a simple way being the global solar radiation the only parameter involved.

The main objective of this work is to analyze the validity of five hourly empirical models to estimate k_d using data measured at the radiometric station installed in Badajoz (Spain). These models were proposed by Orgill and Hollands [3], Spencer [4], Erbs et al. [5], Boland et al. [6], and Ruiz-Arias et al. [7]. They have been applied on numerous occasions and in different regions. Another goals of this study are to select the best model for the region of study and to compare the performance of original models with fitted models based on the same functional expressions.

2 DATA

Data analyzed in this study were obtained from the radiometric station located on the terrace of the building of the Physics Department of the University of Extremadura in Badajoz (Spain), with coordinates: 38° 52' 58" N, 7° 0' 38" W and 199 m a.s.l. This location guarantees an open horizon free of obstacles. The city of Badajoz is located in the south-west of Spain. It is characterized by a climate type Csa according to the Köppen classification, i.e., highly influenced by the Atlantic Ocean, with mild winters and very hot summers. Rainfall takes place mainly in winter and spring. This region is characterized by a high number of sunshine hours and high irradiance values reached due to its latitude and to the predominance of cloud-free conditions during the summer.

Measurements were recorded on one-minute basis during the period of study which extends along more

1. *G.Sanchez is with the University of Extremadura, Avda. de Elvas, s/n, 06006 Badajoz. Spain: gsanchezw@alumnos.unex.es*
2. *A.Serrano is with the University of Extremadura, Avda. de Elvas, s/n, 06006, Badajoz. Spain. E-mail: asp@unex.es*
3. *M.L.Cancillo is with the University of Extremadura, Avda. de Elvas, s/n, 06006, Badajoz. Spain. E-mail: mcf@unex.es*
4. *J.A.Garcia is with the University of Extremadura, Avda. de Elvas, s/n, 06006, Badajoz. Spain. E-mail: agustin@unex.es*

than a year, from November 23, 2009 to December 15, 2010. This long period guarantees the existence of a great variety of meteorological situations, allowing to achieve highly representative results.

Global irradiance was measured with a Kipp & Zonen CMP11 pyranometer. This instrument allows to measure global irradiance incident on a horizontal surface in an azimuth angle of 360° with wavelengths between 310 and 2800 nm. Diffuse irradiance was measured by other similar pyranometer installed on a Kipp & Zonen CM121 shadow ring which prevents direct radiation reaching the pyranometer. In order to guarantee the quality of diffuse radiation data, the shadow ring has been daily adjusted to ensure the complete shading of the pyranometer dome. Moreover, it is necessary to correct the shadow ring measurements because the ring obstructs a portion of sky and it prevents that diffuse radiation from this sky region reaches the pyranometer. For this purpose, diffuse measurements have been corrected by the method proposed by Drummond [8] and recommended by the manufacturer of the shadow ring.

Instruments used in this work have been calibrated by intercomparison with a reference pyranometer (Kipp & Zonen CM11 pyranometer, #027771) in the Atmospheric Sounding Station (ESAt, INTA) located at El Arenosillo, Huelva, Spain (37.1 ° N, 7.06 ° W), which had been previously calibrated at the World Radiation Center (WRC) in Davos, Switzerland. The calibration factors were calculated as the average ratios between the voltage signal of our radiometers and the irradiance measured by the reference pyranometer.

Simultaneous data from all sensors were measured every ten seconds and recorded as one-minute average by a CR1000 data-logger (Campbell Sci.). These data were subjected to a quality control to eliminate possible anomalous measurements. To avoid the intrinsic cosine error of the pyranometers, only measurements taken for solar zenith angle below 85° were considered.

After a careful review of data only 1.1% of the original data were discarded. From these one-minute data, hourly average were calculated. Fig. 1 shows hourly k_d values versus hourly k_t values.

From the hourly data set, 75 % of data have been randomly selected to fit each model and get their new coefficients. The remaining 25% has been used for model validation.

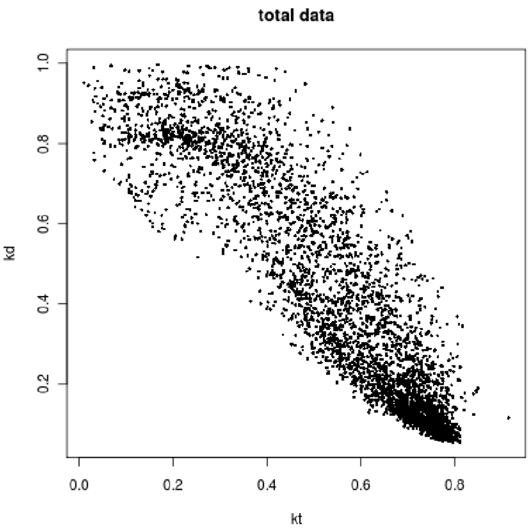

Figure 1. k_d versus kt for the final data set.

3 METHODOLOGY

This study is performed in two ways. First, the functional expressions proposed by each model have been fitted by means of a regression analysis and their new coefficients were obtained. Secondly, the original and new fitted models have been validated and compared with an independent data set.

Statistical parameters such as root mean square error (RMSE), mean bias error (MBE) and the squared coefficient of correlation (R^2) have been used in order to quantify the performance of the models. These statistics are defined as:

$$MBE = \frac{1}{N}\sum_{1}^{N}\left(x_{i,\text{mod}} - x_{i,\text{exp}}\right)$$

$$RMSE = \sqrt{\frac{1}{N}(x_{i,\text{mod}} - x_{i,\text{exp}})^2}$$

$$R^2 = \frac{\left[\sum_{1}^{N}\left(x_{i,\text{mod}} - \overline{x_{\text{mod}}}\right)\left(x_{i,\text{exp}} - \overline{x_{\text{exp}}}\right)\right]^2}{\left[\sum_{1}^{N}\left(x_{i,\text{mod}} - \overline{x_{\text{mod}}}\right)^2\right]\left[\sum_{1}^{N}\left(x_{i,\text{exp}} - \overline{x_{\text{exp}}}\right)^2\right]}$$

where N is the total number of measurements, $x_{i,mod}$ is the *ith-estimate* and $x_{i,exp}$ is the *ith-measurement*. The statistic RMSE is a measure of the differences between the model and the experimental measurements. The relative MBE gives an estimate of the percentage by which a model underestimates or overestimates the experimental results. A positive value of this statistic indicates that the model overestimates the experimental data while a negative value indicates underestimation. Small values of these statistics represent a better fit of the model. On the other hand, the squared coefficient of correlation R^2 between modeled and measured diffuse fraction values represents the proportion of the linear variability explained by the model. It ranges from 0 to, ideally, 1 for a perfect linear relationship.

These statistics provide complementary, but not completely independent information. For this reason, the Taylor Diagram [9], which involves all of them, can be useful to identify the best model. These diagrams provide a way of graphically summarize how closely a model matches the observations. The similarity between a model and the reference observations is quantified in terms of their correlation, centered root-mean-square difference and the amplitude of their variations (represented by their standard deviations)

4 Models

The models analyzed in this work allow the estimation of hourly proportion of diffuse irradiance, k_d, from hourly values of the clearness index, k_t. There are two type of models. The first type proposes different expressions for k_d depending on the value of k_t. The first kt interval ($0 < k_t < k_1$) corresponds to overcast situations where almost radiation reaching the ground is diffuse radiation. The second interval ($k_1 < k_t < k_2$) represents partially covered skies. The last interval ($k_2 < k_t$) corresponds to situations of clear skies, but also includes the possible presence of some clouds that not cover the sun. The models proposed by Orgill and Hollands, Spencer, and Erbs et al. correspond to this type of models. On the other hand, a second type of models propose a single expression valid for all values of kt, not distinguishing different cloud conditions. The models proposed by Boland et al. and by Ruiz-Arias et al. Correspond to this second type of models. Next, the different mathematical expressions proposed by each author are presented.

4.1. Orgill and Hollands

This model was developed by Orgill and Hollands [3] using data obtained in Toronto (Canada, 43.1°N, 79.31°W) and it proposes the following expression:

$$k_d = \begin{Bmatrix} 1-0.249\cdot k_t & k_t<0.35 \\ 1.577-1.84\cdot k_t & 0.35 \le k_t \le 0.75 \\ 0.177 & k_t<0.75 \end{Bmatrix} \quad (1)$$

4.2 Spencer model

In 1982, Spencer [5] proposed a model to estimate k_d using regression coefficients dependent on the latitude. The model was developed using measurements registered at stations distributed all throughout Australia ranging from 20°S to 45°S latitudes. This model takes into account that the path of radiation in the atmosphere increases with latitude and so does the diffuse component. The model is summarized by the following expressions:

$$k_d = \begin{Bmatrix} a & k_t<0.3 \\ b-c\cdot k_t & 0.3 \le k_t \le 0.75 \\ d & k_t \ge 0.75 \end{Bmatrix} \quad (2)$$

where the coefficients *a, b, c* and *d* depend on latitude in the following form:

$$\begin{aligned} a &= b-0.3\cdot c \\ b &= 0.94+0.0118\cdot|\varphi| \\ c &= 1.185+0.0135\cdot|\varphi| \\ d &= b-0.75\cdot c \end{aligned} \quad (3)$$

where φ is the latitude is given in degrees.
The model proposes a symmetric behavior with respect to Equator so the latitude dependence appears as absolute value.

4.3 Erbs et al. model

This model was developed by Erbs et al. [5] from measurements performed at latitudes between 31°N and 43°N. They propose the following exprressions and intervals:

1. When $k_t \leq 0.3$:

$$k_d = 1 - 0.09 \cdot k_t \qquad (4.a)$$

2. When $0.3 < k_t < 0.75$:

$$k_d = 0.9511 - 0.1604 \cdot k_t + 4.388 \cdot k_t^2 - 16.238 \cdot k_t^3 + 12.336 \cdot k_t^4 \qquad (4.b)$$

3. When $k_t \geq 0.75$:

$$k_d = 0.165 \qquad (4.c)$$

4.4 Boland et al. model

Boland et al. [6] developed this model from data recorded in Australia at radiometric stations located at latitudes between 34ºS and 39ºS. It is given by the expression:

$$k_d = \frac{1}{1 + e^{-5.00 + 8.6 \cdot k_t}} \qquad (5)$$

4.5 Ruiz-Arias et al. model.

This model has been proposed by Ruiz-Arias et al. [7] from measurements performed over Europe and America at latitudes between 30.38º N and 64.82ºN. They proposed the following expression for all k_t values:

$$k_d = 0.952 - 1.041 \exp\left(-\exp\left(2.3 - 4.702 \cdot k_t\right)\right) \qquad (6)$$

5 Results

In this section we will discuss the results obtained applying these five models to the experimental data measured at our station. The analysis consists of two steps: (1) obtaining new values for the regression coefficients of each model expressions and (2) validating and comparing original and new fitted models.

5.1 Fitted models.

~~In this first step, each model has been fitted using the 75% of experimental data and their new regression coefficients were obtained. Models proposed by Boland et al. and Ruiz-Arias et al. present the lowest values for RMSE statistic. The new fitted models neither overestimate nor underestimate the experimental data. Thus, the value obtained for MBE statistic is around zero for all models. The values of squared coefficient of correlation obtained for each fitted are presented in Table 1. This table shows that those models proposing several expression for k_d depending on the value of k_t, present lower square coefficient and performs better.~~In this first step, each model has been fitted using the 75% of experimental data and their new regression coefficients were obtained. The values of squared coefficient of correlation obtained for each fitted model are presented in Table 1. This table shows that all fitted models present similar R^2 values, ranging from 0.830 to 0.843, with the highest ones corresponding to the models that propose several expressions for k_d depending on the value of k_t. Moreover, the value obtained for MBE statistic is around zero for all models, indicating that the new fitted models neither overestimate nor underestimate the experimental data.

Table 1. R^2 for each fitted model.

	R^2
Orgill and Hollands	0.843
Spencer	0.843
Erbs et al.	0.843
Boland et al.	0.830
Ruiz_Arias et al.	0.842
Ruiz-Arias eta al.	0.842

5.2 Validation and comparison of original and new fitted models

The k_d as estimated by each original model has been calculated and represented in Fig.1. It can be observed how the original models reflect the general trend of the data but tend to overestimate the experimental data, especially for low values of k_t. Models proposed by Spencer and by Ruiz-Arias are less biased. This is also confirmed by the values of the statistic MBE in Table 2.

A great dispersion can be observed in Fig.1. It is mainly due to the great diversity of atmospheric conditions that lead to the same value of k_t. The clearness index is a measure of the attenuation of radiation by the atmosphere. However, such attenuation may be due to different causes such as cloud cover or the action of the aerosols and gases on the radiation. These factors have different influence on the radiation dispersion so the proportion of diffuse can take very different values for the same radiation attenuation.

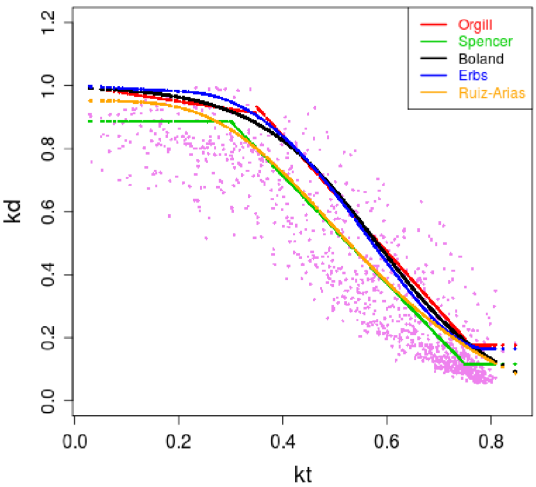

Fig.2. Original models.

In Fig.3 the new fitted models have been represented. If compared to Fig.2, the improvement in the fitted model respect to the original can be addressed. It can be observed that new fitted models better represent the central tendency, being the overestimation lower than for original models. Also, the differences between models seem to have reduced.

A quantitative comparison between each original models and its new fitted approach can be done using the statistics RMSE, MBE and R^2, listed in Tables 2 and 3. It is observed that, in general, statistics are slightly better for the new fitted models than for the original models. Positive MBE values confirm the overestimation of original models, which is lower for the models proposed by Spencer and by Ruiz-Arias et al. (already observed in Fig. 1). For new fitted models (Fig. 2), MBE is almost zero. The lowest value of RMSE statistic for both the original and the fitted models is obtained by the model proposed by Ruiz-Arias et al. followed by the model proposed by Boland et al.. On the other hand, the squared coefficient of correlation presents a very small improvement when models are fitted. This statistic ranges from 0.847 to 0.857 for all models.

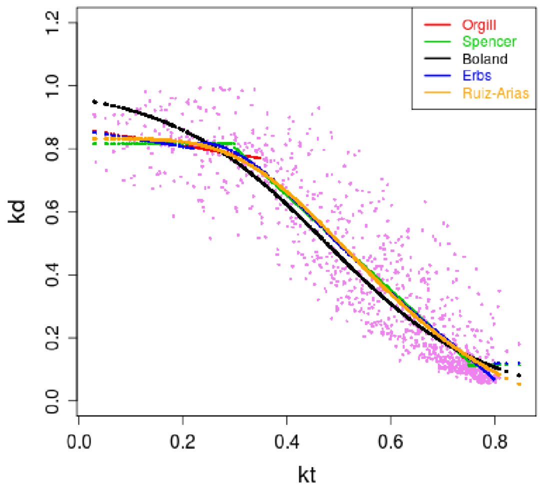

Fig.3. New fitted models

Table 2. RMSE, MBE and R^2 values for original models

	Original models		
	RMSE	MBE	R^2
Orgill and Hollands	0.162	0.116	0.853

Spencer	0.114	0.030	0.855
Erbs et al.	0.160	0.105	0.853
Boland et al.	0.107	0.050	0.852
Ruiz-Arias et al.	0.082	0.025	0.856

~~A quantitative comparison between each original models and its new fitted approach can be done using the statistics~~

Table 3. RMSE, MBE and R^2 values for fitted models

	Fitted models		
	RMSE	MBE	R^2
Orgill and Hollands	0.106	0.001	0.857
Spencer	0.107	0.001	0.855
Erbs et al.	0.107	0.001	0.856
Boland et al.	0.075	-0.003	0.847
Riuiz-Arias et al.	0.072	0.000	0.853
Ruiz-Arias et al.	0.072	0.000	0.857

The distinct statistics provide different information about the performance of the models, being difficult to select the best model and to evaluate if the fitting of models has meant a significant improvement. For this aim, the Taylor's diagram which combines information from different statistics has been represented (Fig. 3).

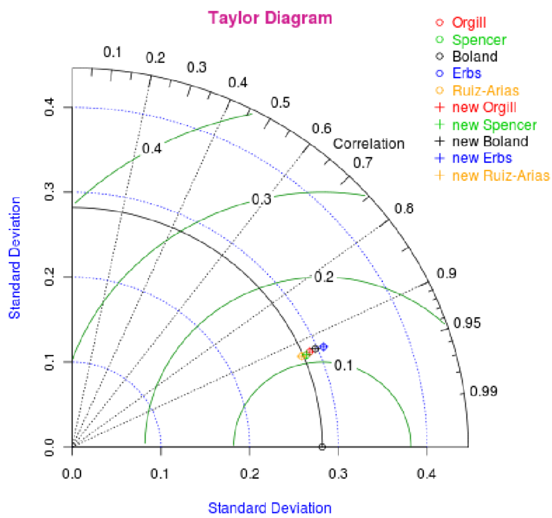

Figure 3. Taylor's diagram.

On Taylor's diagram the correlation coefficient (R), the root-mean-square error (RMSE) and the standard deviation of each model respect to the experimental data are simultaneously considered [9]. Fig.3 confirms the similar values for correlation coefficient and root-mean-square error obtained previously in Table 2 and Table 3. There is only a slight difference in the standard deviation being the model proposed by Ruiz-Arias et al. the one which best performs, followed by the model proposed by Spencer. It can also be seen that no significant differences are found between original and new fitted models.

6 Conclusions

In this study, five empirical models aimed at estimating the portion of hourly diffuse solar irradiance based on clearness index at the radiometric station of Badajoz have been implemented and compared. Using the same functional expressions proposed by the original models, they were fitted to the data measured at the station of Badajoz and new regression coefficients were obtained. Generally, all models reproduced the central tendency of the experimental measurements. Original models overestimate th
e experimental measurements being the models proposed by Spencer and by Ruiz-Arias et al. which better performs among the original models.

The statistics RMSE and MBE, obtained with the new fitted approaches are slightly lower than those found using the original models. According to the results obtained by the Taylor diagram, the fitted models show no significant improvement with respect to original models. The Taylor diagram shows that the estimates provided by all models are very similar, being the expressions proposed by Ruiz-Arias et al. and by Spencer slightly better than the others. To select the best models it is necessary to evaluate the values of MBE and RMSE statistic. According to these values the new fitted models perform better than the original models, resulting the fitted model propsed by Ruiz-Arias the best model, followed by the fitted model proposed by Boland et al. and by Spencer.

ACKNOWLEDGMENT

This work was supported by the research project CGL2008-05939-C03-02/CLI (Ministerio de Ciencia e Innovación).

REFERENCES

[1] Benson R.B., Paris M.V., Sherry J.E., Justus C.G. "Estimation of dayly and monthly direct diffuse and global solar radiation from sun shine duration measurements". *Solar Energy*, vol. 32, pp. 523-535. 1984.

[2] Reindl D.T., Beckman W.A., Duffie J.A. "Diffuse fraction correlation". *Solar Energy,* vol. 45, 1-7. 1990.

[3] Orgill J.F., Hollands K.G.T (1977). Correlation equiation for hourly diffuse radiation on a horizontal surface. *Solar Energy*. **19**, 357-359

[4] Spencer J.W. (1982) A comparison of methods for estimating hourly diffuse solar radiation from global solar radiation. *Solar Energy*. **29**, 19-32

[5] Erbs D.G., Klein S.A., Duffie J.A. (1982). Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. *Solar Energy*. **28**, 293-302.

[6] Boland J., Ridley B., Brown B. (2008) Models of diffuse solar radiation. *Renewable Energy*. **33**, 575-584

[7] Ruiz-Arias J.A., Alsamamra H., Tovar-Pescador J., Pozo-Vázquez D. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. Energy Conversion and Management. 51, 881–893

[8] Drummond, A.J. (1956), On the measurement of sky radiation, *Arch. Meteorol. Geophys. Bioklim*. Ser.B, **7**, 413-436.

[9] Taylor K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research. 106, 7183-7192.