

Journal of Renewable Energies Revue des Energies Renouvelables

journal home page : https://revue.cder.dz/index.php/rer

Research Paper

Prediction of direct normal irradiation using a new empirical sunshine duration-based model

Abdelatif Takilalte ^a,*, Ali Dali ^a, Mohammed Laissaoui ^a, Amar Bouhallassa ^a

^a Centre de Développement des Energies Renouvelables, CDER, BP 62 Route de l'Observatoire, Bouzaréah, 16340, Algiers, Algeria.

ARTICLE INFO	A B S T R A C T
Article history:	In this work, we are interested in presenting a new approach allowing
Received 15 May 2023	us to express the Direct Normal solar Irradiation (DNI) according to
Accepted 03 July 2023	the sunshine duration essentially. This choice is justified by the fact
Keywords:	that in addition to the sunshine, duration has a strong correlation with
Direct normal irradiation,	solar irradiation, it is measured in many radiometric stations. Some
empirical model,	clear sky models with modifications developed exclusively here are
SVR,	made valid for all types of sky. The proposed model is compared with
sunshine duration,	one of the intelligent models such as the Support Vector Regression
all types of sky.	(SVR) for daily data from Ghardaïa.

* Corresponding author, E-mail address: a.takilalte@cder.dz Tel.: +213(0)23189051

ISSN: 1112-2242 / EISSN: 2716-8247



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. Based on a work at http://revue.cder.dz.

1. Introduction

For most solar energy applications, we need to know precisely the temporal variations of the solar radiation received, in particular, the direct normal irradiation (S_n). Unfortunately, the lack of data is a major constraint in the development of solar systems, namely the concentrating solar power system (CSP) [1].

This deficiency requires the use of models to predict the data that have similar statistical characteristics to measured data. The sunshine duration (n) is measured by the heliograph which is available in many stations, thanks to its lower cost compared to other radiometric measuring devices such as the pyranometer (for the global and diffuse components) and Pyrheliometer (for the direct component) [2].

The literature review shows that many researchers have focused on accurate global horizontal irradiance GHI prediction compared to DNI prediction [3]. However, there are several approaches for modeling and predicting direct solar radiation: Some clear sky models (CSMs) directly output DNI, such as the simplified Solis model [4], the REST2 model [5], the Bird model [6] ...etc. Others are used to separate the GHI into DNI and diffuse components by statistical methods like the DirInt model [7], and the BRL model [8].

On the other side, cloud cover data may be added to modify clear sky irradiance and estimate the solar irradiance in overcast conditions [9], [10], and [11]. As an example, in [12], the authors developed a model for calculating both direct and diffuse solar radiation under cloudy sky based on clear sky models, at Ibadan, Nigeria. The deviations from the data were within 15%.

The prediction could be also achieved for various time series resolutions (monthly, daily, and hourly) and using different input parameters, by including empirical models, stochastic processes (AR, ARMA, ARIMA, and SARIMA), artificial intelligence techniques such as neural networks (ANNs), SVR, fuzzy logic, and genetic algorithms [1] [2].

In earlier studies, empirical models which correlate the monthly average daily beam radiation with relative sunshine duration in linear and quadratic forms, have been developed by Iqbal [13] using measured data from three Canadian stations. Otherwise, [14] emphasized the superiority of a proposed ANN prediction model against some empirical models to predict monthly average daily direct solar radiation for locations in Uganda. The model is based mainly on sunshine hours, monthly average daily values of global solar irradiation, and maximum temperature.

In this paper, we introduce a new simple empirical model, very easy to implement as it uses simplified equations that do not require any previous data to predict the daily direct solar radiation. It is economical since it is based on usually available data, which is the sunshine duration. It is sufficiently precise in all sky conditions to compete with a learning machine like the SVR that uses the same input.

The aim here is to reduce the cost generated by the measurement stations thanks to the development of sufficiently precise estimators of direct solar radiation and this from available parameters such as the duration of insolation. The Ghardaïa site in Algeria is taken as a case study.

2. The proposed model

Indeed, the insolation fraction (sunshine fraction σ) is often used in the estimation of solar radiation, in particular global horizontal irradiation (GHI), G_h [2] and to a lesser extent the DNI (S_n) [15]. This choice is based on the fact that they have almost the same instantaneous variations and consequently the high correlation that exists between these radiometric parameters [2], [16]. Likewise, it is wise to exploit it.

It is worth mentioning that the sunshine fraction is calculated as the ratio between the measured sunshine duration (n) to the daily maximum possible sunshine duration (N), as follows [16]:

$$\sigma = n/N \tag{1}$$

Where *N* is equivalent to the astronomical length of the day and it is calculated by the following formula [2]:

$$N = \frac{2}{15} \cos^{-1}(\tan\varphi\,\tan\delta) \tag{2}$$

 φ and δ are the latitude of the site and the daily solar declination respectively.

We propose in this section, the estimation of the direct normal irradiation (S_n) for all types of sky using an empirical relationship, which is a function of the insolation fraction σ and the calculated value of the direct normal radiation to sky clear $S_{n,c}$

$$S_n = \sigma^2 \times S_{n,c} \tag{3}$$

With

$$S_{n,c} = R_b \times S_h \tag{4}$$

Where

 S_h : is the direct irradiation received on a horizontal surface.

 R_b : is the transposition factor of the direct radiation which is given in [17].



Fig 1. Representative diagram of the procedure proposed for the estimation of S_n for all types of sky

It is easy to see that the proposed model introduces corrections to the expression of S_n in clear skies (see the black curve in Figure 2). The latter is easy to obtain from a simple transposition of the direct irradiation incident on a horizontal surface with a clear sky ($S_{h,c}$) through the factor R_b . Thus, the model proposed to estimate S_n at all types of sky can be written in the form:

$$S_n = \left(\frac{n}{N}\right)^2 \cdot R_b \cdot S_h \tag{5}$$

The S_h component can be calculated by the various clear sky empirical models such as the Perrin-Brichambaut, Liu-Jordan, etc. model. Figure 1 represents the flowchart of the proposed method.

3. The SVR model

The model is developed by « Vapnik » [18] to solve the classification problem, but it has recently been extended to the regression domain [19]. Like all Supervised machine learning techniques, it goes through two distinct stages:

- The learning phase: consists of training the model by giving it examples of input data whose outputs are known beforehand.
- The prediction phase: where new samples for which the outputs are not known are inserted.

The goal is to develop a function y = F(x) which represents the dependence of the output $y_i \in R$ on the input vector x_i of dimension p ($x_i \in R_p$ such that i = 1:m). The form of this function is [20]:

$$y = w^T \phi(x) + b \tag{6}$$

Where w is known as the weight vector and b is the bias. The optimal regression function is given by the minimum of the Lagrangian function [21]:

$$min_{w,b} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*)$$
(7)

Where *C* is a predefined value that controls the trade-off between the overestimation and the generalizability of the algorithm, and ε_i , ε_i^* are variables that represent the upper and lower constraints respectively applied to the outputs of the system [1].

The non-linear solution of the SVR, using the ε -insensitive cost function [22], [23] is given by the following quadratic problem (QP):

$$max_{\alpha,\alpha^*}L(\alpha,\alpha^*) = max_{\alpha,\alpha^*}\left(\sum_{i=1}^n \alpha_i^*(y_i - \varepsilon) - \alpha_i(y_i + \varepsilon) - \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i, x_j)\right)$$
(8)

With constraints:

$$0 \le \alpha_i \le C, \quad i = 1, \dots, n$$

$$0 \le \alpha_i^* \le C, \quad i = 1, \dots, n$$

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$$
(9)

Where ε is the parameter of the ε -insensitive loss function, which represents the error tolerance margin.

 $K(x_i, x_j)$ is called the kernel function, and its value is equal to the scalar product of two vectors x_i and x_j in the characteristic space $\emptyset(x_i)$ and $\emptyset(x_j)$, it is expressed by:

$$K(x_i, x_j) = \emptyset(x_i) * \emptyset(x_j)$$
⁽¹⁰⁾

The resolution of QP (equations 8 and 9) leads to the determination of the Lagrange multipliers α_i^*, α_i . The regression function is finally given by:

$$y = f(x, \hat{\alpha}_1, \hat{\alpha}_1^*) = \sum_{i=1}^n (\hat{\alpha}_1 - \hat{\alpha}_1^*) K(x, x_i) + b^*$$
(11)

In our case, the core function RBF (Radial basis function) is used. It is defined by:

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2) \quad \gamma > 0$$
⁽¹²⁾

Where γ is the parameter of the RBF kernel function. The use of the RBF function is recommended because it takes into consideration the case of non-linearity between inputs and outputs [24].

4. Methodology

4.1 Site and data description

Ghardaia city is situated in the southern and sunny part of Algeria with a geographical location of latitude: $+32.37^{\circ}$, longitude: $+3.77^{\circ}$, and altitude: 450 m above the mean sea level. This region is characterized by an arid climate, which exhibits mild and dry weather conditions.

The data available for this study are daily measurements of DNI $(S_{n,m})$ and sunshine duration (n), taken between January 1, 2005, and December 31, 2005, at the applied research unit in renewable energies (ARURE) - Ghardaïa.

The quality of the data used is a crucial factor for the accuracy of the models developed [25],[26]. Generally, a data refining procedure aims to improve these qualities by checking them and filtering them from any uncertainty or error, possibly due to a malfunction of the measuring instrument [26]. Each data is checked to extract missing or unreliable values to overcome this problem [27]. Thus, we excluded from the data set the outliers identified as the values whose insolation fraction (σ) is outside the range of 0.015 < σ < 1.

4.2 Statistical performance validation

Several indicators were used to assess the accuracy of the proposed models. The predicted data (\hat{y}_i) is compared to the actual data y_i and the performance measures are calculated as follows, where *m* is the total number of observations and \bar{y} is the average of the measured values [25]. - Normalized Mean Absolute Error (NMAE): This is the average of the deviations in absolute value from the observed values in percentage, which is defined by:

$$NMAE(\%) = \left(\frac{\frac{1}{m}\sum_{i=1}^{m} |y_i - \hat{y}_i|}{\bar{y}}\right) \times 100$$
(13)

-Normalized Root Mean Square Error (NRMSE):

It is frequently used to measure the differences between the values predicted by a model and the values actually observed. Its version normalized (%) by the measured mean is preferable to compare the precision of the models on different databases. And it is calculated by:

$$NRMSE(\%) = \left(\frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}}{\bar{y}}\right) \times 100$$
(14)

According to [2], the precision of the model is considered:

Excellent for NRMSE < 10%;

Good for 10% < NRMSE < 20%;

Fair for 20% < NRMSE < 30%;

Poor for NRMSE > 30%.

-Relative Percentage Error (RPE):

The RPE measures the ratios as a percentage of the absolute errors of the estimation relative to the magnitude of the exact values. It is given by:

$$RPE = \left(\frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i| / |y_i| \right) * 100$$
(15)

-R-Squared or coefficient of determination (R^2)

The R^2 is a parameter that takes possible useful values between 0 and 1. The higher the R^2 , the better it represents the linear relationship between the estimated and the measured values. The R^2 is obtained by:

$$R^{2} = 1 - \frac{\sum_{i=1}^{m} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{m} (y_{i} - \bar{y}_{i})^{2}}$$
(16)

5. Results and discussion

5.1 Performance of the proposed approach in all types of sky

In our approach, we proceed to the prediction of S_n using a single measured parameter which is σ only (Eq. 5). This is feasible based on one of the clear-sky semi-imperial models used to calculate S_h .

Table 1. Performance results of the proposed method applied for all types of sky data, using different S_h estimation models.

Used models for S_h	NMAE(%)	NRMSE(%)	RPE(%)	R ²
Perrin-Brichambaut	11.05	14.73	19.35	0.87
Liu-Jordan	16.46	20.28	24.07	0.75
Bird-Hulstrom	16.82	23.59	23.86	0.67
Davies-Hay	14.38	19.26	22.03	0.78
Hoyt	14.14	20.70	22.27	0.74

Table 1 summarizes the performance obtained from the approach developed for all types of sky, based on five semi-empirical models tested here for the estimation of S_h . These are: Perrin de Brichambaut model [28], Liu & Jordan model [28], Bird & Hulstrom model [29], Davies & Hay model [30] and Hoyt model [29].

From Table 1, we can note that the estimation of S_h with the "Perrin de Brichambaut" model in the proposed method ($S_n = (\sigma)^2 \cdot R_b \cdot S_{h,Perrin}$), makes it possible to obtain the best results with the least errors by an NRMSE equal to 14.73% which is considered very good results according to the criterion relating to this index (Eq. 14). This confirms that it is the most suitable at least for our case study (Ghardaïa), and therefore it is the one that will be adopted for the rest of our study.

We also notice that the other models have much less performance compared to the model based on "Perrin de Brichambaut" where the difference between error percentages reaches 8% and with a minimum of 3%. This confirms that the choice of the S_h model is a crucial step in determining the quality of this method.

The evolution of the measured irradiations S_n and those estimated by the best model (via the Perrin de Brichambaut model), in the Ghardaïa site during the year 2005, are represented in Figure 2.



Fig 2. Measured values of daily DNI (S_n) and those obtained by the proposed approach based on insolation fraction and the Perrin-Brichambaut model.

We notice that there is a favorable level of agreement between the estimated values and the measured values for the majority of the predictions using the new technique, which is not the

case for the $S_{n,c}$ calculated at the clear sky (Eq.4) that does not follow the fluctuations of solar radiation. This reflects the improvement brought by our model which can be applied to the estimation of S_n for any sky state.

5.2 Comparison with the SVR model

To further appreciate the merit of the proposed empirical technique and draw a more decisive conclusion, a comparison with the model realized in the first part with the SVR model is also carried out.

For the SVR models, the data is divided into two parts: 2/3 of the data is used for training and the rest to test the accuracy of the SVR-RBF models [31]. To obtain the SVR1 model, the S_n is estimated based on the same insolation data σ while in the SVR2 model, two other meteorological and radiometric data are also used as input such as the average daily measured values of the temperature (*T*) and the global irradiation ($G_{h,m}$) respectively. The objective of which is to estimate the DNI regardless of weather conditions. Table 2 presents the results obtained by the various models.

	NMAE(%)	NRMSE(%)	R ²
• S_n as a function of σ			
SVR1 model: $S_n = SVR(\sigma)$	10.71	14.70	0.87
Proposed model: $S_n = \sigma^2 \cdot R_b \cdot S_{h,Perrin.}$	11.05	14.73	0.87
• S_n depending on different parameters			
SVR2 model: $S_n = SVR(\sigma, G_{h,m}, T)$	-	12.62	0.90

Table 2. Comparison between the results obtained by the proposed empirical model and the SVR models constructed for the prediction of S_n .

Although sometimes the results of the two models: the SVR1 and the proposed model are very similar, the proposed empirical model is strongly recommended for the following reasons:

- It is very simple compared to the SVR model, so it is much easier to implement.

- The choice of the empirical model avoids the selection of the parameters of SVR (like C, γ , and ε) which is not an easy task and its precision largely depends on this operation.

- The possibility of estimating the daily *DNI* which corresponds to the relative sunshine duration on any day without the need for a priori knowledge of data, as is the case with SVR models where data sets (training data) must be described beforehand to determine their associated parameters.

However, the fact remains that SVR and artificial intelligence techniques in general have other advantages, the main one being their flexibility to model problems where there are no well-defined mathematical relationships between certain physical parameters. Another important advantage of these artificial intelligence techniques is their ability to include multiple inputs which can enhance the accuracy of the model, as in the case of the SVR2 model.

6. Conclusion

This paper was mainly devoted to the modeling of normal direct solar irradiation data based mainly on sunshine duration. The latter is the key parameter that allows the reproduction of solar radiation, based on an appropriate correlation.

Our objective is achieved through the development of innovative and optimal techniques with a reduced number of input parameters, and ensuring a good compromise between computational complexity and accuracy. These techniques will allow us to extrapolate later, and spatially, the synthetic solar radiation data to sites with climatic conditions similar to those of Ghardaïa. The comparative study shows that the proposed methods can compete with the conventional models and even an intelligent model, specifically the SVR model.

7. References

[1] P. Lauret, C. Voyant, T. Soubdhan, M. David, and P. Poggi, "A benchmarking of machine learning techniques for solar radiation forecasting in an insular context," Sol. Energy, vol. 112, pp. 446–457, 2015, doi: 10.1016/j.solener.2014.12.014.

[2] K. Mohammadi, S. Shamshirband, M. Hossein, K. Amjad, and D. Petkovic, "Support vector regression-based prediction of global solar radiation on a horizontal surface," vol. 91, pp. 433–441, 2015, doi: 10.1016/j.enconman.2014.12.015.

[3] E. W. Law, A. A. Prasad, M. Kay, and R. A. Taylor, "Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting - A review," Sol. Energy, vol. 108, pp. 287–307, 2014, doi: 10.1016/j.solener.2014.07.008.

[4] P. Ineichen, "Comparison and validation of three global-to-beam irradiance models against ground measurements," Sol. Energy, vol. 82, no. 6, pp. 501–512, 2008, doi: 10.1016/j.solener.2007.12.006.

[5] C. A. Gueymard, "A Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces," Sol. Energy, vol. 82, no. 3, pp. 272–285, 2008, doi: 10.1016/j.solener.2007.04.008.

[6] R. L. H. Richard E.Bird, simplified clear sky model for direct and diffuse insolation on

horizontal surfaces. Goldon Colorado, USA: SERI/TR-642-761, Solar Energy Research Institute. 1981.

[7] R. R. Perez, P. Ineichen, E. L. Maxwell, R. D. Seals, and A. Zelenka, "Dynamic global-todirect irradiance conversion models," ASHRAE Trans., vol. 98, no. pt 1, pp. 354–369, 1992.

[8] B. Ridley, J. Boland, and P. Lauret, "Modelling of diffuse solar fraction with multiple predictors," Renew. Energy, vol. 35, no. 2, pp. 478–483, 2010, doi: 10.1016/j.renene.2009.07.018.

[9] M. S. Gul, T. Muneer, and H. D. Kambezidis, "Models for obtaining solar radiation from other meteorological data," Sol. Energy, vol. 64, no. 1–3, pp. 99–108, 1998, doi: 10.1016/S0038-092X(98)00048-6.

[10] R. Chen, E. Kang, X. Ji, J. Yang, and J. Wang, "An hourly solar radiation model under actual weather and terrain conditions: A case study in Heihe river basin," Energy, vol. 32, no. 7, pp. 1148–1157, 2007, doi: 10.1016/j.energy.2006.07.006.

[11] D. R. Myers, "General cloud cover modifier for clear sky solar radiation models," Opt.Model. Meas. Sol. Energy Syst., vol. 6652, p. 665208, 2007, doi: 10.1117/12.732472.

[12] F. J. K. Ideriah, "A model for calculating direct and diffuse solar radiation," Sol. Energy, vol. 26, no. 5, pp. 447–452, 1981, doi: 10.1016/0038-092X(81)90224-3.

[13] M. Iqbal, "Correlation of average diffuse and beam radiation with hours of bright sunshine," Sol. Energy, vol. 23, no. 2, pp. 169–173, 1979, doi: 10.1016/0038-092X(79)90118-X.

[14] J. Mubiru, "Using Artificial Neural Networks to Predict Direct Solar Irradiation," Adv.Artif. Neural Syst., vol. 2011, pp. 1–6, 2011, doi: 10.1155/2011/142054.

[15] T. Khatib, A. Mohamed, and K. Sopian, A review of solar energy modeling techniques, vol. 16, no. 5. Elsevier Ltd, p. 2864–2869.

[16] H. Suehrcke, R. S. Bowden, and K. G. T. Hollands, "Relationship between sunshine duration and solar radiation," Sol. Energy, vol. 92, pp. 160–171, 2013, doi: 10.1016/j.solener.2013.02.026.

[17] M.R. Yaïche et S.M.A. Bekkouche, "Conception et Validation d'un Programme Sous Excel pour l'Estimation du Rayonnement Solaire Direct en Algérie. Cas d'un Ciel Clair," Rev. Int. d'Héliotechnique, vol. 39, pp. 50 – 55, 2009.

[18] V. Vapnik, "The Nature of Statistical Learning Theory." Springer-Verlag, New York, 1995.

[19] V. Vapnik, S. E. Golowich, and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing," Adv. Neural Inf. Process. Syst.

9, pp. 281--287, 1997, doi: 10.1007/978-3-642-33311-8_5.

[20] Z. Ramedani, M. Omid, A. Keyhani, B. Khoshnevisan, and H. Saboohi, "A comparative study between fuzzy linear regression and support vector regression for global solar radiation prediction in Iran," Sol. Energy, vol. 109, no. 1, pp. 135–143, 2014, doi: 10.1016/j.solener.2014.08.023.

[21] F. Girosi, "An equivalence between sparse approximation and support vector machines,"A.I. Memo No. 1606, MIT, vol. 1480, pp. 1455–1480, 1997.

[22] S. Belaid and A. Mellit, "Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate," ENERGY Convers. Manag., vol. 118, pp. 105–118, 2016, doi: 10.1016/j.enconman.2016.03.082.

[23] R. Urraca, J. Antonanzas, M. Alia-Martinez, F. J. Martinez-De-Pison, and F. Antonanzas-Torres, "Smart baseline models for solar irradiation forecasting," Energy Convers. Manag., vol. 108, pp. 539–548, 2016, doi: 10.1016/j.enconman.2015.11.033.

[24] M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke, "ScienceDirect A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production," Sol. ENERGY, 2014, doi: 10.1016/j.solener.2013.12.006.

[25] K. Mohammadi, S. Shamshirband, M. H. Anisi, K. Amjad Alam, and D. Petković, "Support vector regression based prediction of global solar radiation on a horizontal surface," Energy Convers. Manag., vol. 91, pp. 433–441, 2015, doi: 10.1016/j.enconman.2014.12.015.

[26] L. Martín, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, "Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning," Sol. Energy, vol. 84, no. 10, pp. 1772–1781, 2010, doi: 10.1016/j.solener.2010.07.002.

[27] K. Dahmani, R. Dizene, G. Notton, C. Paoli, C. Voyant, and M. L. Nivet, "Estimation of 5-min time-step data of tilted solar global irradiation using ANN (Artificial Neural Network) model," Energy, vol. 70, pp. 374–381, 2014, doi: 10.1016/j.energy.2014.04.011.

[28] A. Takilalte, S. Harrouni, M. R. Yaiche, and L. Mora-López, "New approach to estimate 5-min global solar irradiation data on tilted planes from horizontal measurement," Renew. Energy, vol. 145, pp. 2477–2488, 2020, doi: 10.1016/j.renene.2019.07.165.

[29] R. L. H. Richard E.Bird, simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Goldon Colorado, USA: SERI/TR-642-761, Solar Energy Research Institute. 1981.

[30] Muhammad Iqbal, An Introduction to Solar Radiation. New York - London: Academic

Press, 1983.

[31] A. Takilalte, S. Harrouni, and J. Mora, "Forecasting global solar irradiance for various resolutions using time series models - case study: Algeria," Energy Sources, Part A Recover. Util. Environ. Eff., vol. 44, no. 1, pp. 1–20, 2022, doi: 10.1080/15567036.2019.1649756.