International Journal of Physics, 2024, Vol. 12, No. 2, 83-88
Available online at http://pubs.sciepub.com/ijp/12/2/4
Published by Science and Education Publishing
DOI:10.12691/ijp-12-2-4

Choice of the Best Regression Model for Estimating the Global Solar Radiation Component in the City of Ouagadougou in Burkina-Faso

Toussaint Tilado Guingane[1,2,*], Mouhamadou Falilou Ndiaye[3], Sosthène Tassembedo[2], Éric Korsaga[2], Dominique Bonkoungou[1,2], Zacharie Koalaga[2], François Zougmore[2]

[1]Laboratoire des Matériaux et Environnement (LA.M.E.), Unité de Formation et de Recherche en Sciences Exactes et Appliquée (UFR/SEA), Université Pr KI-ZERBO, Ouagadougou, Burkina Faso
[2]Laboratoire de Sciences et Technologies (LaST), Unité de Formation et de Recherche en Sciences et Techniques (UFR/ST), Université Thomas SANKARA, Ouagadougou, Burkina Faso
[3]Centre International de Formation et de Recherche en Energie Solaire (C.I.F.R.E.S), ESP-UCAD: BP 5085, Dakar-Fann, Senegal
*Corresponding author: tilado88@yahoo.fr

Received March 16, 2024; Revised April 18, 2024; Accepted April 25, 2024

Abstract The aim of the study was to identify the best-performing regression model for estimating the solar radiation component in Ouagadougou, Burkina Faso, a region characterised by high levels of sunshine. Of the four models evaluated, the Bristow Campbell model emerged as the optimal choice, outperforming the others in terms of accuracy. The importance of this research lies in the need to select a regression model adapted to the specific climatic conditions of the region, thus contributing to more accurate energy planning and sustainable use of solar energy. The evaluation of the models was based on metrics such as RMSE, MBE, t-statistic, R2, and the d metric, reinforced by detailed graphical visualisations and correlations between the models and real data. The results clearly demonstrated the performance of the Bristow Campbell model, underlining its reliability and efficiency in estimating global solar radiation. This conclusion is supported by an in-depth analysis of the metrics and graphs, providing valuable insights for researchers, energy planners and policy makers.

Keywords: *solar radiation, regression model, metrics, performance, Bristow Campbell model*

Cite This Article: Toussaint Tilado Guingane, Mouhamadou Falilou Ndiaye, Sosthène Tassembedo, Éric Korsaga, Dominique Bonkoungou, Zacharie Koalaga, and François Zougmore, “Choice of the Best Regression Model for Estimating the Global Solar Radiation Component in the City of Ouagadougou in Burkina-Faso.” *International Journal of Physics*, vol. 12, no. 2 (2024): 83-88. doi: 10.12691/ijp-12-2-4.

1. Introduction

In the dynamic context of the energy transition, the accurate assessment of solar radiation is of paramount importance in order to make optimal use of solar energy, particularly in regions with high levels of sunshine such as Ouagadougou in Burkina Faso. While various regression models are available to estimate the solar radiation component, choosing the most appropriate one for a specific context remains a crucial challenge. This study focuses on identifying the best regression model among four models, in order to refine solar radiation forecasts in the specific context of Ouagadougou [1,2,3,4].

With the diversity of existing regression models there is a need to determine the one that offers the best accuracy for the specific climatic conditions for Ouagadougou. Each model has advantages and limitations, and the choice of the optimal model is essential to guarantee reliable results. Given this diversity, our main objective is to select the best-performing regression model for estimating the global solar radiation component in this region, thereby contributing to more accurate energy planning and sustainable exploitation of solar energy.

Using a rigorous methodological approach, we will carry out an in-depth comparison of the performance of the four regression models identified. The analysis will be based on meteorological data specific to Ouagadougou, incorporated into each model, in order to assess their ability to capture the subtle variations in solar radiation in this locality. The results of this study will provide crucial information for researchers, energy planners and decision-makers, guiding strategic choices in the exploitation of solar energy in a specific climatic context such as that of Ouagadougou.

2. Materials and Methods

2.1. Study Sites

The Zagtouli solar power plant, located in Ouagadougou, is at latitude 12°19'46" N, longitude

1°37'31" W, and at an altitude of around 312 metres above sea level. Strategically located in this region of high sunshine levels, the Zagtouli solar power plant represents a key infrastructure in Burkina Faso's energy landscape.
With its significant production capacity, this solar power plant benefits from the region's particular climatic conditions, characterised by abundant sunshine throughout the year. Temperature data extracted directly from this site provides a precise perspective on the climatic variations that influence solar radiation in Ouagadougou.

As the region's solar energy hub, the Zagtouli power station plays a crucial role in the search for sustainable energy solutions for Burkina Faso. The data collected at this strategic location will serve as an essential basis for our study aimed at choosing the best regression model for estimating the solar radiation component under specific conditions in Ouagadougou. Thus, by exploiting these temperature data from the Zagtouli solar power plant, our research aims to make a significant contribution to understanding and optimising solar energy in this high-potential region.

Figure 1. study site

2.2. Sources of Data

The solar power plant is equipped with four meteorological stations, including instruments for measuring solar radiation, pyranometers, a hygrometer humidity sensor and an anemometer for measuring wind speed. These instruments record the climatic conditions in the area where the PV modules are installed. The information received by each instrument is collected and stored on a server located in the SCADA (Supervisory Control And Data Acquisition) system. The minimum and maximum air temperatures at the location studied have been collected for the year 2019.

3. Working Methodology

Meteorological data, crucial to our analysis, will be extracted from the site, encompassing parameters such as temperature over a significant period.

A data pre-processing phase will follow, involving the elimination of outliers and the management of potential gaps. In parallel, we will select four regression models, such as the Hargreaves and Samani model, the Annandale et al. model, the Bristow and Campbell model and the Goodin et al. model [5], taking into account their relevance to the climatic characteristics of the region.
The models will be validated by applying cross-validation techniques to assess their robustness and by comparing their performance using statistical indicators.

The best model will be chosen by selecting the one that offers the best accuracy, with an in-depth analysis of its sensitivity to climatic variations.

3.1. Hargreaves and Samani model

Hargreaves and Samani [5,6] first suggested in 1982 that global solar radiation (Rs) can be estimated from the difference between maximum and minimum air temperature using a simple equation as follows.

$$R_s = a \times R_a \times \left(T_{\max} - T_{\min}\right)^{0,5} \quad (1)$$

Where Rs(MJ.m^{-2}d^{-1}) is the global solar radiation; T_max and T_min in ℃ are respectively the daily maximum and minimum air temperature; a is an empirical coefficient, the value of which must be 0.16 for inland regions and 0.19 for coastal regions.

Extraterrestrial radiation Ra (MJ. m^{-2}.d^{-1}) can be calculated for any day of the year and latitude using the 1980 equation by Duffie and Beckman [7,8].

$$R_a = \left(\frac{24 \times 60}{\pi}\right) \times S_c \times DF \times \begin{pmatrix} \cos\cos\varphi \times \cos\cos\delta \times \sin\sin\omega_s \\ +\omega_s \times \sin\sin\varphi \times \sin\sin\delta \end{pmatrix} \quad (2)$$

Where Sc is the solar constant (1367 W/m^2 /j or 0.082 MJm^{-2} .mn^{-1} or 118.08MJm^{-2} d^{-1}), DF is the eccentricity correction factor of the Earth orbit, which can be calculated by the expression :

$$DF = 0.1 + 0.033 \times \cos\cos\left(2\pi \times \left(\frac{JulianDay}{360}\right)\right) \quad (3)$$

Where φ is the latitude of the site, can be converted into radians by the expression :

$$\varphi = latitude \times \frac{\pi}{180} \quad (4)$$

And δ (in degrees) is the solar declination, which can be calculated by the expression :

$$\delta = (23.45) \times \sin\sin\left(2\pi \times \left(284 + \left(\frac{JulianDay}{365}\right)\right)\right) \quad (5)$$

ωs (in degrees) is the hourly sunset angle, which can be calculated using the expression

$$\omega_s = \cos^{-1}\cos^{-1}\left(-\tan\tan\varphi \times \tan\tan\delta\right) \quad (6)$$

Table 1 below gives the recommended mean days for the months, the values of the corresponding Julian day and the declination.

Table 2 gives the calendar for the month of December and also the numbers of the days of the week. To obtain the example shown in Table 1, we have considered the first day of December 2019, which starts on a Sunday and is the 7th day of the week. The average day of the month is therefore given by equation 7.

Table 1. Recommended mean days of the month, the values of the corresponding Julian day and the declination

		an average day of the month		
Month	i for ith of month	Average days of the month	Julian Day	δ, Declination
January	i	17	17	
February	i+31	14	45	
March	i+59	16	75	
April	i+90	16	106	
May	i+ 120	17	137	
June	i +151	18	169	
July	i+ 181	16	197	
August	i+ 212	18	230	
September	i+ 243	19	262	
October	i+ 273	17	290	
November	i +304	18	322	
December	i+ 334	20	353	

Table 2. Calendar for December 2019 plus weekday numbers

Mon	Tue	Wed	Thu	Fri	Sat	Sun
25	26	27	28	29	30	1
1	2	3	4	5	6	7
2	3	4	5	6	7	8
8	9	10	11	12	13	14
9	10	11	12	13	14	15
15	16	17	18	19	20	21
16	17	18	19	20	21	22
22	23	24	25	26	27	28
23	24	25	26	27	28	29
28	29	29	30	31	32	33
30	31	1	2	3	4	5
34	35					

Average days of the month = (7+8+9+……..+ 35)/31 =19.65≈ 20 (7)

3.2. Annandale et al. Model

Annandale et al [5][10,11,12] modified the Hargreaves and Samani model by introducing a correction factor as follows.

$$R_s = a.R_a.\left(1+2.7.10^5 Z\right)\left(T_{max} - T_{min}\right)^{0.5} \quad (8)$$

Where Z is the altitude in metres.

3.3. Bristow and Campbell Model

Bristow and Campbell [5] proposed the following relationship for daily values of global solar radiation (Rs) as a function of daily extraterrestrial solar radiation (Ra) and temperature difference (ΔT):

$$R_s / R_a = A * \left[1 - \exp\left(-B.\ \Delta T^C\right)\right] \quad (9)$$

Where ΔT = Tmax - Tmin and A, B and C are the empirical coefficients. The values of A, B and C in the Bristow and Campbell model were taken to be 0.7, 0.004-0.01 and 2.4 respectively.

3.4. Goodin and al. Model

Goodin et al[5] evaluated a form of the Bristow and Campbell model.

$$R_s / R_a = A * \left[1 - \exp(-B.\left(\Delta T^C / R_a\right)\right] \quad (10)$$

Table 2 shows the recommended average days for each month and the value of n (the day of the year) per month according to Duffie and Beckman [5].

4. Statistical Indicators of Model Performance

In the literature, numerous statistical indicators are used by authors to assess model performance.

In this study, the tools used are the root mean square error (RMSE), the mean bias error (MBE), the t-statistic (t), the coefficient of determination (R2), and Willmott's concordance index (d). These statistical tools are defined by the following relationships. [13,14,15,16,17,18,19,20]

4.1. Root Mean Square Error (RMSE)

The RMSE value is always positive, representing zero in the ideal case. The normalised root mean square error provides information on the short-term performance of the correlations by allowing a term-by-term comparison of the actual difference between the predicted and measured values. The smaller the RMSE value, the better the model's performance.

However, this test does not differentiate between underestimation and overestimation.

$$RMSE = \left[\frac{1}{n} \times \sum_{i=1}^{n} \left(R_{s_{meas}}(i) - R_{s_{est}}(i)\right)^2\right]^{\frac{1}{2}} \quad (11)$$

Where Rsmeas (i) and Rsest (i) are, respectively, the ith measured value and the ith estimated value of daily solar radiation (in $kWh.m^{-2}.d^{-1}$), and n is the number of values.

4.2. The Mean Bias Error (MBE)

This test provides information about long-term performance; a low MBE value is desirable.

The lower the absolute value, the better the model's performance. The MBE value represents the systematic error or bias; a negative value indicates an average underestimate of the calculated value, while a positive value indicates an overestimate by the model.

Consequently, one of the disadvantages of these two tests is that the overestimation of an individual observation will trigger the underestimation of another observation.

$$MBE = \frac{1}{n} \times \sum_{i=1}^{n} \left(R_{s_{est}}(i) - R_{s_{meas}}(i)\right) \quad (12)$$

4.3. The t-Statistic Indicator (t)

One of the statistical indicators most commonly used in the literature to assess the performance of a proposed model is the t-statistic. According to Jacovides and Kontoyiannis, using the MBE and RMSE separately can lead to a poor decision in selecting the best model from a series of candidate models. The t-statistic should therefore be used in conjunction with the MBE and RMSE errors to better assess a model's performance. It is defined by the MBE and RMSE errors as follows.

$$t = \left[\frac{(n-1) MBE^2}{RMSE^2 - MBE^2} \right]^{1/2} \tag{13}$$

The smaller the value of t, the better the performance of the model.

4.4. The Coefficient of Determination R^2

The coefficient of determination R2 is a statistical measure that indicates how well the regression line fits the actual data. An R2 value close to 1 indicates that the regression line fits the data well. This indicator varies between 0 and 1. A value of 1 indicates perfect agreement between the measurement and the model, while a value of 0 indicates total disagreement.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(R_{s_{meas}}(i) - R_{s_{est}}(i) \right)^2}{\sum_{i=1}^{n} \left(R_{s_{meas}}(i) \right)^2} \tag{14}$$

4.5. The Concordance Index ''d'' of Willmott

Willmott's "d" concordance index was calculated using the equation below. For good model accuracy, values of d should be close to 1:

$$d = 1 - \left[\frac{\sum_{i=1}^{n} \left(R_{s_{est}}(i) - R_{s_{meas}}(i) \right)^2}{\sum_{i=1}^{n} \left(\left| R_{s_{est}}(i) - \overline{R_{s_{meas}}(i)} \right| + \left| R_{s_{meas}}(i) - \overline{R_{s_{meas}}(i)} \right| \right)^2} \right] \tag{15}$$

5. Result and Discussion

The performance of the models was tested using statistical metrics. Table 3 shows the values of the metrics for each model.

Table 4 shows the different coefficients found for our site for different models.

The lower the RMSE, the better. In our case, the "Bristow Campbell" model has the lowest RMSE (52.5281), which suggests that it has the best performance among the models we evaluated.

The MBE indicates whether there is a systematic bias in the predictions. An MBE close to zero is generally desirable. In your case, "Bristow Campbell" has a relatively low MBE (-1.0969), which is favourable.

The t-statistic measures the statistical significance of differences between predictions and actual values. A value close to zero suggests better performance. All our models have relatively low t-statistics, indicating good performance except for the Goodin et al model.

R2 close to 1 is desirable, as it means that the model explains a large proportion of the variance in the data. In our case, all the models have high R2, but "Bristow Campbell" has the highest R2 (0.9999), indicating an excellent fit.

d measures the uncertainty in the prediction. A value closer to zero is better. In our case, "Modele Annandale et al" has the lowest value of d (0.9913), which is preferable.

In conclusion, according to these metrics, the "Bristow Campbell" model seems to be the best for our site among those you have evaluated.

Figure 2 presents a graph showing the four models and the measured data.

Table 3. Metric for the different models

Model	RMSE	MBE	t-statistic	R^2	d
Model Hargreaves Samani	90.0683	1.3648	0.0503	0.9999	0.9980
Modele Annandale and al	90.1665	-2.8340	0.1043	0.9996	0.9913
Bristow Campbell model	52.5281	-1.0969	0.0693	0.9999	0.9994
Goodin and al	107.3506	86.0269	4.4432	0.5123	-1.5650

Table 4. Coefficients of the different models

Model	Coefficients
model Hargreaves Samani	a=0.108
Modele Annandale and al	a=0.108
Bristow Campbell model	A = 2.218 B = 0.003 C = 1.569
Goodin and al	A = 2.218 B = 0.003 C = 1.569

Figure 2 gives an idea of the correlation between the models and the measured values. It can be seen that the "Annandale et Al" and "Bristow campbell" models appear to be closer to the regression line.

Figure 3 and Figure 4 show the curves and histograms of monthly measured and predicted values respectively. These two figures confirm that the Bristow Campbell model is the best performing for our site.

Finally, we have chosen the "Bristow campbell" model for our site, given that the various tests, graphs and correlations confirm it.

By replacing the coefficients with their values, the "Bristow campbell" model selected for our site is :

$$R_s / R_a = 2.218 * \left[1 - \exp\left(-0.003 * \Delta T^{1.569} \right) \right] \quad (16)$$

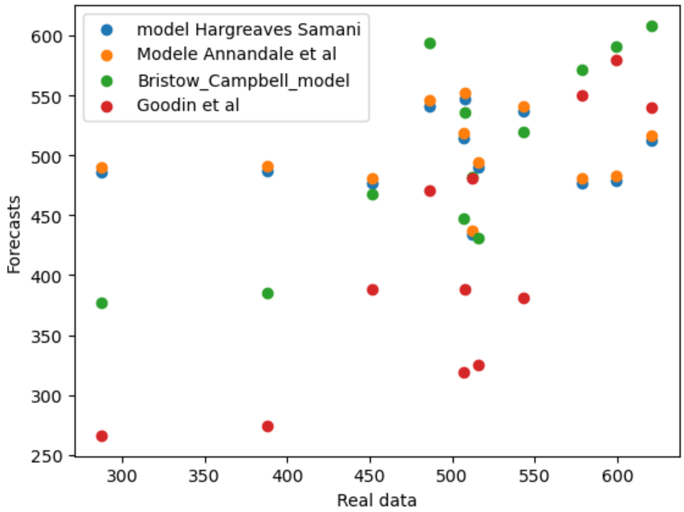

Figure 2. Comparing Models with Real Data

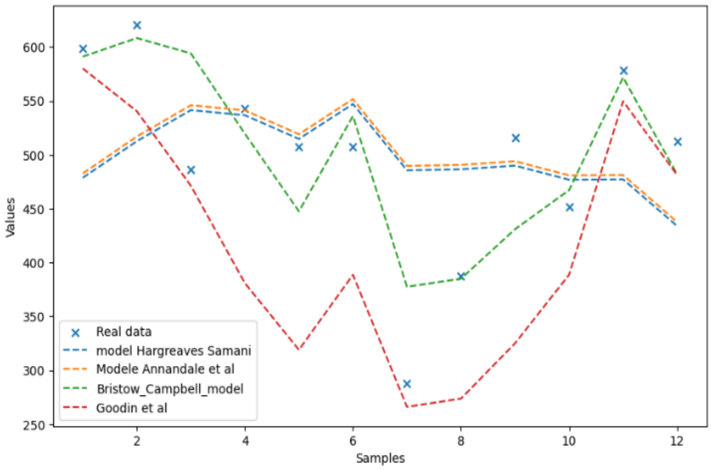

Figure 3. Comparison of Models with Real Data

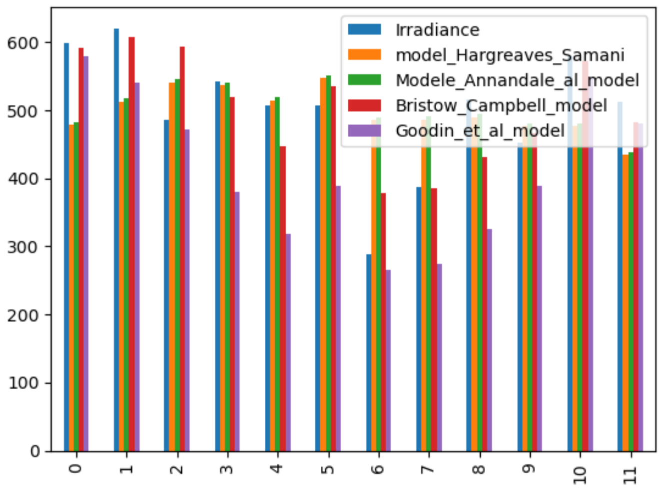

Figure 4. histogram comparing Models with Actual Data

6. Conclusion

At the end of this work, we can conclude that the Bristow Campbell model stands out as the optimal choice.

Our tests involved four well-known models, namely the Hargreaves Samani model, the Annandale et al model, the Bristow Campbell model, and the Goodin et al model. These models were evaluated using various metrics, including the RMSE, the MBE, the t-statistic, the R2, and the d-metric.

The results showed conclusively that the Bristow Campbell model outperformed the others, demonstrating remarkable accuracy in estimating the global solar radiation component.

This conclusion is supported by an in-depth analysis of the metrics and graphs, underlining the reliability and effectiveness of the Bristow Campbell model for the specific conditions at our site in Ouagadougou. In sum, our research provides clear indications as to the choice of the most efficient regression model for the needs of solar radiation estimation in the studied region, thus offering practical perspectives for applications related to solar energy and local meteorology. However, extending our work towards instantaneous models with smaller time steps offers a significant opportunity to improve the temporal resolution of your predictions and enhance the relevance of your models for practical applications.

References

[1] T.T Guingane, “Contribution a l’etude de l’influence d’un systeme pv connecte sur le reseau electrique”, Doctorat Université Ouaga I Professeur Joseph Ki-Zerbo Burkina Faso, Sciences Appliquées, Spécialité : Sciences Physique-semi-conducteurs, 11/05/2018.

[2] P. GROS, «Utilisation du modèle linéaire. Rappels de base – Méthodes de validation», Laboratoires Côtiers de l’Ifremer, (Novembre 2000).

[3] M.S. Okundamiya et al., “Empirical Model for Estimating Global Solar Radiation on Horizontal Surfaces for Selected Cities in the Six Geopolitical Zones in Nigeria”, Research Journal of Applied Sciences, Engineering and Technology 2(8): 805-812, (2010).

[4] Menges H. O., Ertekin C., Sonmete M. H. “Evaluation of global solar radiation models for Konya, Turkey”, Energy Conversion and Management 47: 3149–3173, 2006. Moussa Diop, “Energy Systems: Vulnerability – Adaptation – Resilience (VAR)”, Regional focus: sub-Saharan Africa, Senegal, (2009).

[5] E. Mboumboue, D. Njomo, M. L. Ndiaye, P. A. N'diaye, M. F. Ndiaye, and A. K. Tossa, “ On the applicability of several conventional regression models for the estimation of solar global radiation component in Cameroon and Senegal sub-Saharan tropical regions ”, Journal of Renewable and Sustainable Energy 8, 025906 (2016).

[6] Daniel K. Fisher, H. C. Pringle III, “ Evaluation of alternative methods for estimating reference evapotranspiration”, Agricultural Sciences, Vol.4, No.8A, 51-60 (2013).

[7] Y.M.Irwan, I.Daut, I.Safwati , M.Irwanto , N.Gomesh , M.Fitra, “ An Estimation of Solar Characteristic in Kelantan using Hargreaves Model”, TerraGreen 13 International Conference 2013 - Advancements in Renewable Energy and Clean Environment, Energy Procedia 36 (2013) 473 – 478.

[8] F. M. Abed Al-Dulaimy, and G. Y. M. Al-Shahery,” Estimation of Global Solar Radiation on Horizontal Surfaces over Haditha, Samara, and Beji, Iraq.”, The Pacific Journal of Science and Technology, Volume 11. Number 1. May 2010 (Spring).

[9] J. A. Duffie, W .A. Beckman, “Solar Engineering of Thermal Processes”, Fourth Edition, Published by John Wiley & Sons, Inc., Hoboken, New Jersey,2013.

[10] J. Annandale, N.Z.Jovanovic, N.Benade, R.G.Allen, “Software for missing data error analysis of Penman-Monteith reference evapotranspiration”, Springer-Verlag, 2001.

[11] M. H. Soulouknga, R. Z.Falama, O. O. Ajayi, S. Y. Doka, T. C. Kofane, “ Determination of a Suitable Solar Radiation Model for the Sites of Chad”, Energy and Power Engineering, 2017, 9, 703-722, http://www.scirp.org/journal/epe, ISSN Online: 1947-3818.

[12] D. G. Goodin, J. M. S. Hutchinson, R. L. Vanderlip, and M. C. Knapp, “ Estimating Solar Irradiance for Crop Modeling Using Daily Air Temperature Data”, Agronomy Journal, Vol. 91, September–October 1999, 845-851.

[13] R. Rakotomalala,” Analyse de corrélation Étude des dépendances - Variables quantitatives”, Version 1.1, March 2015, Université Lumière Lyon 2.

[14] P. Dagnelie, “Statistique théorique et appliquée Tome 1 ”, De Boeck Supérieur s.a., 2013, 3e édition, Rue des Minimes 39, B-1000 Bruxelles.

[15] C. J. Willmott, “On the validation of models,” Phys. Geogr. 2, 184–194 (1981).

[16] C. J. Willmott et al., “A refined index of model performance,” Int. J. Climatol. 32, 2088–2094 (2011).

[17] C. P. Jacovides and H. Kontoyiannis, “Statistical procedures for the evaluation of evapotranspiration computing models,” Agric. Water Manage. 27, 365–371 (1995).

[18] R. J. Stone, “Improved statistical procedure for the evaluation of solar radiation estimation models,” Sol. Energy 51(4), 289–291 (1993).

[19] I. S. Nouhoun, “Modélisation du rayonnement solaire pour la simulation des performances thermiques des Eco-bâtiments en Afrique Subsaharienne ”, Master en génie électrique et énergétique, spécialité : énergie renouvelables, 23 Septembre 2020, Institut International d’Ingénierie Rue de la Science Burkina Faso.

[20] J. Liu, T. Pan , D. Chen , X. Zhou , Q. Yu , G. N. Flerchinger , D.L. Liu , X. Zou, H. W. Linderholm , J.Du , D.Wu and Y.Shen,” An Improved Ångström-Type Model for Estimating Solar Radiation over the Tibetan Plateau ”, Energies 2017, 10, 892.

© The Author(s) 2024. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).