# A Multivariate Regression Model for the Assessment of Solar Radiation in the Senegalese Territories

Ousmane Wane[1(✉)], A. A. Navarro[2], L. Ramírez[2], R. X. Valenzuela[2], José M. Vindel[2], F. Ferrera Cobos[2], Cheikh M. F. Kébé[1], and L. F. Zarzalejo[2]

[1] CIFRES, Ecole Supérieure Polytechnique - UCAD, BP 5085, Dakar-Fann, Senegal
pa.ousmane1@gmail.com
[2] CIEMAT, Energy Department – Renewable Energy Division, Av. Complutense, 40, 28040 Madrid, Spain

**Abstract.** Senegal has a great solar potential, so it could be used to shift from a diesel-based power generation to cheaper renewable energy resources. To exploit this inexhaustible natural resource, the global horizontal irradiation remains one of the key parameters for any solar energy project at a given location. This work establishes a multiple linear regression approach to estimate the solar radiation in the Senegalese territories using the information of the global network of weather geostationary satellites (Meteosat and GOES), satellites database and the ground measurement data available in the website of the World Radiation Data Center (WRDC) as inputs to the model. Jointly a set of multivariate regression models, a statistical analysis between Meteonorm data and outputs of different linear combinations are presented in this work, which also gives the opportunity to appreciate the precision and consistency of each solar radiation model on different locations in the study area.

**Keywords:** Solar radiation · Global horizontal irradiation
Multivariate regression model · Satellites database

## 1 Introduction

Senegal, like sub-Saharan countries, presents a significant energy supply gap characterized by scarcity of petroleum in this part of the sub-region to which is added the constant fluctuations in the price per barrel. Moreover the production equipment aging electric power whose fuel is the dominant one represents a constraint to remedy the inadequate supply to energy demand. For example, in 2000 only 5% of rural households are connected on the national electrical network (Youm et al. 2000). Although nowadays there is a progress in this sector and most of villages in Senegal are unconnected to the national network.

Taking into account what has been stated, the policy orientation toward the strengthening of our production systems becomes a challenge that must be tackled in order to satisfy the distribution of electricity in quality and quantity to the population.


C. M. F. Kebe et al. (Eds.): InterSol 2017/CNRIA 2017, LNICST 204, pp. 3–15, 2018.
https://doi.org/10.1007/978-3-319-72965-7_1

Nevertheless for the realization/execution of solar energy conversion projects in a country or region, it is necessary to collect full information on solar resources. This information on the solar resource facilitates decision-making on different technologies that can be used locally or regionally, as well as to the investments needed for its realization. Therefore, the geographical assessment of the solar resource analysis is without doubt the first step in the deployment of development strategies of solar energy in a particular region.

As much as the study area is deprived of radiometric stations with good spatial coverage, it remains interesting to use an appropriate methodology for estimating solar radiation. Consequently, the solar radiation derived from satellite images is an advanced methodology widely used that offers high reliability and accuracy estimates (Amillo et al. 2014; Rigollier et al. 2004). Nowadays the estimation methods using information from meteorological satellites and/or spatial interpolation are typical when determining the value of solar radiation in a pixel of a rasterized geographic region (Amillo et al. 2014; Perez et al. 1997; Posselt et al. 2014; Rigollier et al. 2004). They found out that is the basis of such web applications PVGIS (Photovoltaic Geographical Information System) and Meteonorm offering solar radiation data and other meteorological parameters (Perpiña Castillo et al. 2016). Šúri et al. (2007) present an analysis and mapping of the potential for solar electricity in Europe producible using a model of solar radiation and climate data available on PVGIS (Šúri et al. 2007). In addition if the estimation model has input data from different sources including in situ data available, these will help consolidate the mathematical approach to estimate solar radiation (Lefèvre et al. 2007; Zarzalejo et al. 2009).

## 2 Study Area

With a Sahelian climate, Senegal is between 12º and 17º north latitude and 11º and 18º west longitude. From a general point of view, the country has two seasons: a rainy season and a dry season. Both transitions to go from one season to another are hardly noticeable. This reflects the consideration of these two seasons in this part of the earth. The winter or rainy season that begins in southern of the country in May, gradually spread over the territory between May and October with a peak in August. However, a great disparity was noted with less precipitation in the northern part, a climate like BWh, compared to the southern which is the tropical climate (Aw) with a dry winter (Kottek et al. 2006). Between November and April the country is watered by continental trade winds, it is the dry season.

The highest temperatures were recorded during the rainy season and in the north east area. They decrease as and as we approach the coastal areas. The lowest values were observed in January and February.

Except for the Southern region with some rugged terrain whose altitude does not exceed 581 m at the highest point of the foothills of the Fouta Djallon (Guinea), the topography of the Senegalese territory is more or less flat and does not rise above 130 m.

## 3 Multivariate Regression Model

In general, a regression model consists to study, analyse and interpret the relationship between a dataset ($Y$) called dependent or response variable and an independent or explicative variable ($X$) through a linear function defined by $Y = aX + b$, where $a$ and $b$ are real constants (Núñez et al. 2011). Instead of correlating an independent variable with another dependent variable, the multiple regression model allows to express the dependent variable from a linear combination of two or more independent variables. Thus one of the goals of multivariate regression analysis is to find the correlation coefficients or real constants appropriate to explain relevant aspects between the dependent variable and the set of independent variables (Montgomery and Runger 2003; Núñez et al. 2011). These coefficients weighted independent variables are then used to fit the model. The dependent variable is the parameter to be modelled.

The starting hypothesis model of multivariate linear regression associates the dependent variable to a linear combination of p independent variables weighted by coefficients ($\mathrm{a}_\mathrm{k}^{(\mathrm{i})}$) plus a random perturbation ($\varepsilon^{(\mathrm{i})}$).

$$\mathrm{y}^{(\mathrm{i})} = \mathrm{a}_0^{(\mathrm{i})} + \sum_{\mathrm{k}=1}^{\mathrm{p}} \mathrm{a}_\mathrm{k}^{(\mathrm{i})} \mathrm{x}_\mathrm{k}^{(\mathrm{i})} + \varepsilon^{(\mathrm{i})} \tag{1}$$

Where $y^{(i)}$ is the i-th dependent variable (i = 1,2,…,n) and $x_k^{(i)}$ is the i-th observation of k-th independent variable (k = 1,2,…,p) with $\mathrm{p} \leq \mathrm{n}$.

Each coefficient regression $\mathrm{a}_\mathrm{k}^{(\mathrm{i})}$ reflects the contribution of the variable $x_k^{(i)}$ in the response $\mathrm{y}^{(\mathrm{i})}$. $\varepsilon^{(\mathrm{i})}$ measures the effect of all the variables not included in the model that affect the response variable. For a given observation, this term is achieved by calculating the difference between the observation on the dependent variable and the corresponding estimated value. In addition to the initial hypothesis and an inference in our model the following hypothesis is established (Zarzalejo 2005):

- The number of observations (n) must be greater than the number of explicative variables (p). In other words, if too many independent variables are included in a model it is very likely that the regression coefficients are biased in the direction opposite to the null hypothesis ($\mathrm{H}_0$ below) (Núñez et al. 2011).
- The same as the response variables, the forecast errors are independent between them. The errors are null hope $E\left[\varepsilon^{(i)}\right] = 0$: The random perturbation has on average 0. Otherwise, the expected value of the respond variable is only a function of regression coefficients and explicative variables.
- The distribution of errors follow a normal law of mean equal to zero. It is also assumed which is the case for the distribution of the response variables.
- The error variance is constant and doesn't depend on the independent variables.

From the (Eq. 1), we have the traditional model of simple linear regression:

$$Y = Xa + \varepsilon \tag{2}$$

Where a is the vector of the regression coefficients, ε is the vector of random perturbations and Y is dependent variable. X is the matrix of independent variables.

The regression coefficient matrix is an indicator of the contribution of each explicative variable in the model. However, the regression coefficients are influenced by factors such as variance and linearity of the explicative variables. The variance is a measure of dispersion of the variables in question. Therefore the determination of the optimal regression coefficients leads to study and analyse the perturbations (errors) of the model. Hence, the idea is to minimize the errors. Otherwise one seeks to minimize the distance between the values of the dependent variable and predicted values. This will be done by finding the solution of the equation that minimizes the sum of perturbations (ε) using the method of least squares.

# 4 Data

Some studies like Obrecht in 1990, Ba and Nicholson in 2001 and Diabaté et al. in 2004 estimated/mapped the solar radiation in Africa using satellites data and available pyranometers data (Ba and Nicholson 2001; Diabaté et al. 2004; Obrecht et al. 1990). The problem they faced and which is continuing to recurrent, is the scarcity of radiation stations in parts of the continent to validate their mapping of solar radiation studies using satellite information. The lack of ground measurement solar radiation data with good spatial coverage and over a sufficiently long period on the sub-region, leads us to work with satellite data and reduced ground data available in situ. Even if the data exist, they aren't often over a sufficiently long period or in a nearby present. Besides, it will add the possibility of non-coincidence between the field data and satellite information like in (Diabaté et al. 2004).

## 4.1 Ground Data

In our study a total of 10 World Radiation Data Center (WRDC) ground stations identified in Senegal are used for modelling the monthly average of global horizontal irradiation (GHI). The WRDC, sponsored by the World Meteorological Organization (WMO), collects and archives radiometric data around the world to ensure the availability of these data for research by the international scientific community. The WRDC is one of the world or national data centers with as much radiometric ground stations in the Senegalese territories. The selected recording period for this work ranges between 1984 and 1991 and is far from completed for these stations identified in Senegal (Table 1). The same locations are used for a set of training points to download the GHI from other and following databases used in this work.

The time series of global horizontal irradiation are measured using thermoelectric pyranometer Kipp & Zonen CM5 model (WRDC 2016). This solar radiation value is

**Table 1.** Details of the 10 WRDC stations using as well as possible to get the maximum solar radiation information on the extent of the Senegal. These geo-locations have been used as the training points for downloading solar radiation data from various databases.

| Station name | Code | Lat. (º) | Long. (º) | Elev. (m) | Years of the measurements |
|---|---|---|---|---|---|
| Bambey | 616411 | 14.42 | −16.28 | 20 | 1984–1988 |
| Dakar/Yoff | 616410 | 14.44 | −17.3 | 27 | 1984; 1986–1991 |
| Kédougou | 616990 | 12.34 | −12.34 | 178 | 1988–1990 |
| Linguere | 616270 | 15.23 | −15.07 | 20 | 1984; 1986–1991 |
| Louga | 616121 | 15.37 | −16.13 | 38 | 1985; 1989–1991 |
| Matam | 616300 | 15.39 | −13.15 | 15 | 1984–1991 |
| Nioro du Rip | 616871 | 13.44 | −15.47 | 18 | 1988–1991 |
| Podor | 616120 | 16.39 | −14.58 | 6 | 1985–1991 |
| Tambacounda | 616870 | 13.46 | −13.41 | 49 | 1984–1987 |
| Ziguinchor | 616950 | 12.33 | −16.16 | 26 | 1986–1991 |

the monthly average of the sum of the energy of solar radiation that reaches one square meter in a horizontal surface in a day (GHI) and is given in J/cm$^2$. For the remainder of this work the data are converted in Wh/m$^2$/day and represent the dependent variable for the multivariate regressive model.

### 4.2 Satellites Data

In each of these ten stations identified in the Fig. 1, we also got the global solar radiation from the NASA-SSE database with monthly average incident data on the surface of the earth time series between 1984 and 2004 (NASA-SSE 2016). The same has been done with the web applications such as PVGIS that provides monthly values. Indeed, PVGIS offers two solar radiation values: a solar radiation calculated from the Helioclim database and the other using the CMSAF data (PVGIS 2016). The monthly average of the sum of GHI estimations from satellite images (GOES and Meteosat) provided by CMSAF, Climate Monitoring Satellite Application Facility between 1998

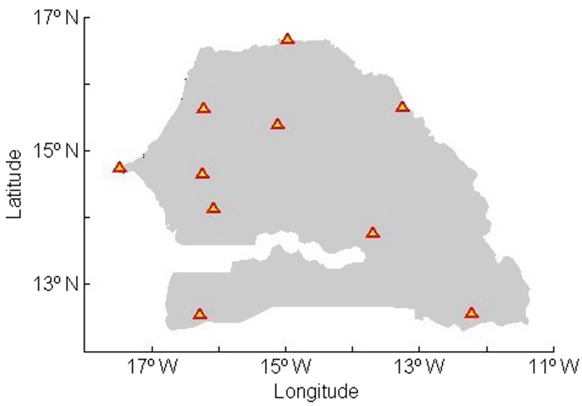


**Fig. 1.** WRDC Station locations identified in Senegal

and 2005 and June 2006–Dec. 2011, Helioclim between 1985 and 2005 (Huld et al. 2012; PVGIS 2016). These datasets represent the sets of independent variables for modelling the solar radiation.

PVGIS: The PVGIS (Photovoltaic Geographical Information System) is a web application that provides climate data and tools needed to assess performance of photovoltaic systems (PV) in Europe, Africa and southwest Asia (PVGIS 2016).

NASA-SSE: National Aeronautics and Spatial Administration - Surface meteorology and Solar Energy is a renewable energy resource website developed under project POWER (Prediction of Worldwide Energy Resource) piloted by NASA (NASA-SSE 2016).

### 4.3 Focus on the Inputs Data

All the data used in this study have been downloaded into a format file that were unusable directly. So, it was only possible to display the data on the computer screen when accessing website and introducing the geographical coordinates of the training point. From there, these data were collected manually and it is likely to induce some error. In addition, one of the obstacles is the non-coincidence of this time series.

All inputs data are measured in Wh/m$^2$/day and referred to as monthly average of the sum of global solar radiation energy that hits one square meter on a horizontal surface in one day.

The rest of databases provides a unique monthly value for each month. So, we are left with an established database from five sources listed above. This allows us to have five different monthly average solar radiation values for each of these ten geo-locations (Fig. 1). As follows, the data of each database in a given location represent a variable with 120 observations: it's the annual period (10 stations, 12 months) for modelling the monthly average of the sum of the global solar radiation. Nevertheless a seasonal separation for the datasets had been made also according to the dry and rainy seasons. In framework of our study, the rainy and dry season correspond respectively to the months from May to October and November to April.

This separation of inputs data are used in the Eq. (1) according to the different linear combinations possible:

$$\text{Model 1}: \; Y_{WRDC} = a_0 + a_1 \cdot X_{CMSAF} + a_2 \cdot X_{Helioclim} + a_3 \cdot X_{NASA} + \varepsilon_1 \tag{3}$$

$$\text{Model 2}: \; Y_{WRDC} = b_0 + b_1 \cdot X_{CMSAF} + b_2 \cdot X_{Helioclim} + \varepsilon_2 \tag{4}$$

$$\text{Model 3}: \; Y_{WRDC} = c_0 + c_1 \cdot X_{CMSAF} + c_2 \cdot X_{NASA} + \varepsilon_3 \tag{5}$$

$$\text{Model 4}: \; Y_{WRDC} = d_0 + d_1 \cdot X_{Helioclim} + d_2 \cdot X_{NASA} + \varepsilon_4 \tag{6}$$

Where $Y_{WRDC}$ represent the dataset of solar radiation obtained from WRDC and $X_i$ is the dataset corresponding to the i database, with $i = \{CMSAF, Helioclim, NASA\}$.

To validate the estimated dataset, the Meteonorm data are used. Meteonorm offers access to accurate meteorological for any place on Earth (METEOTEST 2016).

The Meteonorm database stands out among the various sources used as part of this work. In Meteonorm, several different international databases (ground stations and satellite data) are included, checked for ensuring reliability and forming a single comprehensive database permitting worldwide simulation of solar energy systems, buildings and environmental simulations (Remund et al. 2015). This choice was also motivated by the fact that Meteonorm uses the same weather stations WRDC identified above. The irradiance downloaded from Meteonorm for the same training points in the Fig. 1 is a monthly mean hourly global horizontal solar radiation and is given in W/m$^2$. This data converted in Wh/m$^2$/day.

## 5 Validation Model

One of the steps after the determination of the regression coefficients is to evaluate the contribution of each regression coefficient, following the hypotheses listed below:

- $H0 : a_1 = a_2 = \ldots = a_p = 0$. In other words, there is no contribution of any independent variable in the response. If this hypothesis is true, the model spells: $\mathrm{y}^{(\mathrm{i})} = \mathrm{a}_0^{(\mathrm{i})} + \varepsilon^{(\mathrm{i})}$.
- $H1 : a_p \neq 0$. That is, at least one of the independent variables (explicative) makes a contribution.

The idea is to see whether the addition of a variable as the result of other variables in the regression model makes a significant contribution to the proportion of variance due to regression. For this, we use the theoretical distribution statistics: the test of Fisher for testing the model in its entirety or the student test to see the contribution of each estimator. The Fisher statistic or the Student is an analysis of the variance of the variables by calculating a probability distribution obeying a normal distribution. Under the null hypothesis and the hypothesis of independence, the ratio of means squares regression of the variance and the residual variance, defined as $F_M$, follows a Fisher distribution with $p$ and $[n-(p+1)]$ degrees of freedom:

$$\mathrm{F_M} = \frac{\left(\mathrm{a^t X^t Y} - \mathrm{n}\bar{Y}^2\right)/\mathrm{p}}{(\mathrm{Y^t Y} - \mathrm{a^t X^t Y})/[\mathrm{n}-(\mathrm{p}+1)]} \sim \mathrm{F_{p,n-(p+1)}} \quad (7)$$

Where $(a^t\mathrm{X}^tY - n\bar{Y}^2)$ and $(\mathrm{Y}^tY - a^t\mathrm{X}^tY)$ represent matrix form of the expression of regression Sum of squared (SSR) and sum of squared errors (SSE), respectively. $\mathrm{F_{p,n-(p+1)}}$ is the tabulated Fisher distribution function. Thus the critical value (p-value) of $\mathrm{F_M}$ is defined as the probability that $\mathrm{F_{p,n-(p+1)}}$ is superior to $\mathrm{F_M}$:

$$\mathrm{p - value} = \mathrm{P}(\mathrm{F_{p,n-(p+1)}} \geq \mathrm{F_M}) \quad (8)$$

When $\mathrm{F_M}$ > p-value, the null hypothesis (H0) is rejected. At least one of the independent variables makes a significant contribution to the response. This is leading to study the contribution of each independent variable.

With the student distribution, the test focuses on the contribution of each independent variable. Under the normality hypothesis of the distribution of response variables, the sampling distribution of a regression coefficient, $a_1$ for example, is that of a normal distribution whose variance is $s^2(a_1)$ and the mean is equal to the expected value of so-called coefficient. The comparison of this expected value to the estimated value defined by t, follows the student Statistics with $[n-(p+1)]$ degrees of freedom and $(\alpha)$ significance level:

$$t_M = \frac{\hat{a}_k - a_k}{s(a_k)} \sim t_{\alpha/2,(n-p-1)} \tag{9}$$

Where, $\hat{a}_k$ is the k-th estimator of the matrix of regression coefficients, $a_k$ is the expected value of the k-th estimator, $t_{\alpha/2,(n-p-1)}$ is given by the t-student table and $s(a_k)$ is the square root of the k-th diagonal term of the matrix of covariance, $SSE/(n-p-1)\cdot(X^tX)^{-1}$.

The significance level represents the region where the values are not compatible with the null hypothesis. For example, to evaluate the contribution of the k-th estimator, we simply need to apply the null hypothesis (H0) for this estimator, i.e. $a_k = 0$ or the mean value of the distribution is zero. Unlike the fisher statistics, it is possible to obtain the p-value (critical value) and the confidence interval at a desired significance level $(\alpha)$ for each estimator. In the study we choose a threshold of 0.05 for the significance level for the bounds. In other word, the confidence level for the bounds is 95%. The same analysis previously done with the Fisher statistic is used here to calculate the critical value of each estimator. For a given estimator, this critical value represents the probability that $t_{\alpha/2,(n-p-1)}$ is superior to t of the corresponding estimator. The confidence bound defines the area of acceptance or rejection of the null hypothesis. This range is defined by:

$$a_k \pm t_{\alpha/2,(n-p-1)} \tag{10}$$

The null hypothesis is rejected when:

$t_M < -t_{\alpha/2,(n-p-1)}$ or $t_M > t_{\alpha/2,(n-p-1)}$. In other words $P(t_{\alpha/2,(n-p-1)} \geq t_M) < 0{,}05$

In order to evaluate the proposed models in its entirety, we use the classical estimators as Mean Bias Error (MBE) and Root Mean Square Error (RMSE). The combination of the bias and precision is statistical indicator that defines the performance of an estimator. The MBE evaluates the systematic error in the estimation (Zarzalejo 2005; Walther and Moore 2005). The MBE and RMSE are given by the following expressions (Walther and Moore 2005):

$$MBE = \sum_{i=1}^{n} \frac{\hat{y}^{(i)} - y^{(i)}}{n} \tag{11}$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{\left(\hat{y}^{(i)} - y^{(i)}\right)^2}{n}} \tag{12}$$

Where $y^{(i)}$ is the Meteonorm data, $\hat{y}^{(i)}$ is the estimated solar radiation data and n represents the dimension of the dataset.

## 6 Results and Discussion

The regression coefficients and the evaluation parameters of these estimators constituted by the standard deviation, confidence interval, F (Fisher) and t (Student) statistic, and p-value of these estimators are illustrated in this table. Due to the annual and seasonal separation of data, each linear combination or multivariate regression model are tested three times thus allowing to compare the results for a better estimation of solar radiation.

The statistical analysis between the solar radiation outputs models and Meteonorm data, separated the dataset into annual and seasonal period included in spatially distributed. As the objective is to have a high coefficient of determination with a smallest bias and a good accuracy, it becomes difficult to conform from one model to another or if the model would sealed to choose the right period of analysis. The conclusion on the choice of a model might be the right method to estimate solar radiation in our study can be supported by scientific reasoning on the probability estimators according to the critical value (p-value) of the estimators (correlation coefficients) and the statistical parameter listed above. A p-value less than 0.05 is generally accepted as the threshold where the null hypothesis is rejected and at least one independent variable or the independent variable in question brings significant contribution to the model. From the above, Table 2 summarizing the statistical analysis of the estimators of models we can note that none of the cases studied has a p-value less than 0.05 for all these estimators. This reflects that some variables whose p-value less than 0.05 don't make a significant contribution to the response. The correlation of outputs model with Meteonorm data illustrated in the following table and figures confirm what has been found about the choice of a models with regard to others (Fig. 2).

With the corresponding period from November to April or dry season, the correlation results of the outputs M1 model with Meteonorm data gives a RMSE of the order of 11.49% with 10.74% distortion error and a coefficient of determination of 0.99. Comparing these results with those obtained from other linear combinations for the same period of separation of data (dry season), the model M1 has more inaccuracy and distortion of a higher error than others. Because each of the other models (M2, M3 and M4) have a much lower MBE with a value of approximately −1.94% identical to M2, M3 and M4. The RMSE of each these three is ranging from 7.35% to 7.76% (see Table 3). Thus, for the dry season the M3 model which uses as NASA-SSE data inputs and PVGIS-CMSAF is more suitable than others for the assessment of horizontal global solar radiation on a monthly average.

To evaluate the most ideal model for the rainy season, the same difference noted earlier with the statistical parameters with the dry period is more or less observed for the period corresponding to the month storm. Except the first linear combination,

**Table 2.** Statistical analysis of the models M1, M3 and M4 according to the annual period, dry seasonal and rainy seasonal, respectively.

| Parm. | Coef. | SD | Inf. bound | Sup. bound | $t_M$ | p-value |
|---|---|---|---|---|---|---|
| *Model Nº 1 (M1): Annual period* | | | | | | |
| $a_0$ | 1264,86 | 387,68 | 497.32 | 2032.93 | 3,26 | 0,0015 |
| $a_1$ | 0,57 | 0,13 | 0.32 | 0.82 | 4,56 | 0,0000 |
| $a_2$ | −0,18 | 0,12 | −0.41 | 0.056 | –1,50 | 0,1400 |
| **$F_M$ = 40,43** | | | | **p-value = 0,0000** | | |
| *Model Nº 3 (M3): Dry season* | | | | | | |
| $c_0$ | 804,08 | 485,55 | −168.09 | 1776.26 | 1,66 | 0,1030 |
| $c_1$ | 0,63 | 0,18 | 0.28 | 0.98 | 3,56 | 0,0008 |
| $c_2$ | 0,11 | 0,19 | −0.27 | 0.48 | 0,56 | 0,5767 |
| **$F_M$ = 41.87** | | | | **p-value = 0,0000** | | |
| *Model Nº 4 (M4): Rainy season* | | | | | | |
| $d_0$ | 2468,45 | 583,21 | 1301.52 | 3637.38 | 4,23 | 0,0000 |
| $d_1$ | 0,64 | 0,15 | 0.34 | 0.95 | 4,26 | 0,0000 |
| $d_2$ | −0,15 | 0,11 | −0.37 | 0.08 | −1,28 | 0,2061 |
| **$F_M$ = 14.65** | | | | **p-value = 0,0000** | | |

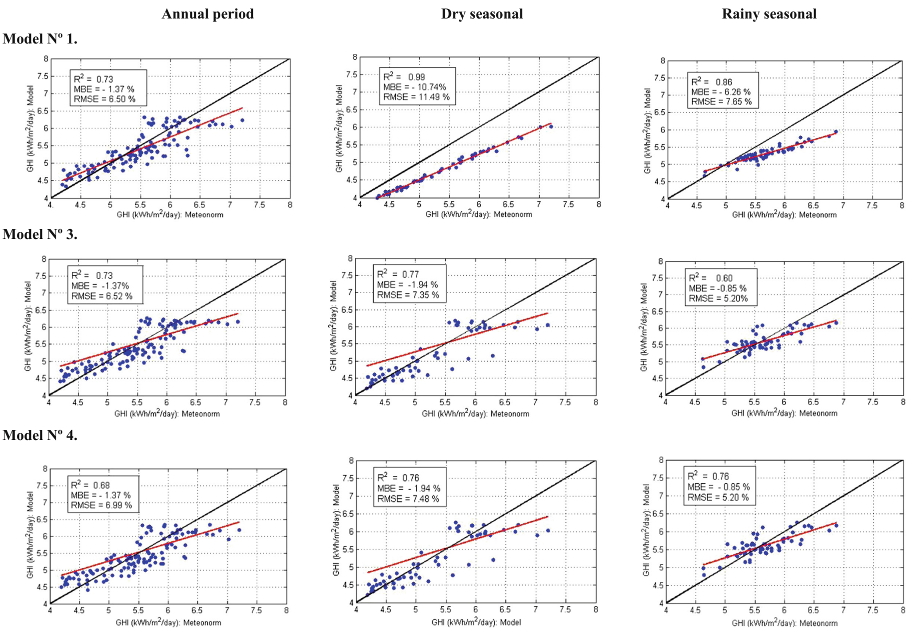


**Fig. 2.** Results of the linear correlation of the models M1, M3 and M4

**Table 3.** Linear correlations coefficient between the estimated values for the different models and the Meteonorm data.

| Model | Linear combinaison | Period | MBE (%) | RMSE (%) |
|---|---|---|---|---|
| M1 | $GHI = f(CMSAF, Helioclim, NASA)$ | Annual | −1.37 | 6.50 |
| | | Rainy | −6.26 | 7.65 |
| | | Dry | −10.74 | 11.49 |
| M2 | $GHI = f(NASA, Helioclim)$ | Annual | −1.37 | 6.93 |
| | | Rainy | −0.85 | 5.76 |
| | | Dry | −1.94 | 7.76 |
| M3 | $GHI = f(NASA, CMSAF)$ | Annual | −1.37 | 6.52 |
| | | Rainy | −0.85 | 5.20 |
| | | Dry | −1.94 | 7.35 |
| M4 | $GHI = f(CMSAF, Helioclim)$ | Annual | −1.37 | 6.99 |
| | | Rainy | −0.85 | 5.20 |
| | | Dry | −1.94 | 7.48 |

whose bias is −6.65% with quadratic error of 7.65%, the MBE is approximately equal to −0.85% M2, M3 and M4 with a RMSE of 5.76% M2 and 5.20% for M3 and M4. In addition, when comparing the coefficients of determination of these different models, we can observed that the M4 remains the most adequate to evaluate the solar radiation during the rainy season.

In an analysis based on annual period, a slight difference was noted compared to the values of the statistical parameters calculated when switching from one model to another. For this period of separation of data, MBE is the same for all models and their value is equal to −1.37% for all considered models. The determination coefficients are between 0.68 and 0.73 and the root mean square errors vary from 6.50% up to 6.99%. In conclusion, the model M1 is the most suitable choice for the estimation of solar radiation without distinction on the dry and rainy seasons.

## 7 Conclusions

The vulnerability of energy supply systems in sub-Saharan Africa is an obstacle to development that must be addressed to ensure energy security in the region. The energy mix is an alternative to offer a qualitative and quantitative supply of electricity to the population. Renewable energy technologies are a way to reach this energy mix. The main advantages of the use of renewable energies are the diversification of energy supply, the use of new production and distribution of energy, which ensures the competitiveness of the Senegalese electrical system, and minimize the impact on environment.

From satellite data and ground data, the monthly mean of the sum of solar radiation energy that reaches a square meter on a horizontal plane in a day has been studied for estimating the solar resource in a geographical point in Senegal necessary for all solar energy project at national level. For example, it is necessary to have a good number of measuring stations or training points with a good spatial distribution in the study area to develop a map of solar energy potential. The constraints related to the radiometric data from available weather stations in the country made the interest of our research project. This linear combination of satellite and ground data is important to elaborate a Typical Meteorological Years (TMYs) in the future solar energy systems studies. The need for a meteorological data base represents an advance in the field of assessing solar resource on the extent of Senegalese territories. Such a study can serve as a scientific contribution or reference for future projects to install photovoltaic or solar thermal power plants.

**Acknowledgement.** An acknowledgement to the Radiation Solar Group of CIEMAT for supervising this work, not forgetting the Spanish Cooperation Agency (AECID) for financing of my stay at this research center.

## References

Amillo, A.G., Huld, T., Müller, R.: A new database of global and direct solar radiation using the eastern meteosat satellite, models and validation. Remote Sens. **6**, 8165–8189 (2014). https://doi.org/10.3390/rs6098165

Ba, M.B., Nicholson, S.E.: Satellite-derived surface radiation budget over the african continent. Part II: climatologies of the various components. J. Clim. 14, 60–76 (2001). https://doi.org/10.1175/1520-0442(2001)014<0060:SDSRBO>2.0.CO;2

Diabaté, L., Blanc, P., Wald, L.: Solar radiation climate in Africa. Sol. Energy **76**, 733–744 (2004). https://doi.org/10.1016/j.solener.2004.01.002

Huld, T., Müller, R., Gambardella, A.: A new solar radiation database for estimating PV performance in Europe and Africa. Sol. Energy **86**, 1803–1815 (2012). https://doi.org/10.1016/j.solener.2012.03.006

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F.: World map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift **15**, 259–263 (2006). https://doi.org/10.1127/0941-2948/2006/0130

Lefèvre, M., Wald, L., Diabaté, L.: Using reduced data sets ISCCP-B2 from the Meteosat satellites to assess surface solar irradiance. Sol. Energy **81**, 240–253 (2007). https://doi.org/10.1016/j.solener.2006.03.008

Zarzalejo, L.F.: Estimación de la radiación global horaria a partir de imagenes de satelite. Desarrollo de modelo empíricos. Universidad Complutense de Madrid (2005)

METEOTEST: Meteonorm [WWW Document] (2016). http://meteonorm.com. Accessed 9 Feb 2016

Montgomery, D.C., Runger, G.C.: Applied Statistics and Probability for Engineers. Wiley, Hoboken (2003)

NASA-SSE: Langley Research Center Atmospheric Science Data Center Surface meteorological and Solar Energy (SSE) web portal supported by the NASA LaRC POWER Project [WWW Document] (2016). https://eosweb.larc.nasa.gov/sse/. Accessed 8 Feb 2016

Núñez, E., Steyerberg, E.W., Núñez, J.: Estrategias para la elaboración de modelos estadísticos de regresión. Rev. Esp. Cardiol. **64**, 501–507 (2011). https://doi.org/10.1016/j.recesp.2011.01.019

Obrecht, D.: Météorologie solaire et images satellitaires : cartographie du rayonnement solaire, détermination de l'albédo des sols et évaluation de l'ennuagement (1990)

Perez, R., Perez, R., Seals, R., Zelenka, A.: Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data. Sol. Energy **60**, 89–96 (1997). https://doi.org/10.1016/S0038-092X(96)00162-4

Perpiña Castillo, C., Batista e Silva, F., Lavalle, C.: An assessment of the regional potential for solar power generation in EU-28. Energy Policy **88**, 86–99 (2016). https://doi.org/10.1016/j.enpol.2015.10.004

Posselt, R., Mueller, R., Trentmann, J., Stockli, R., Liniger, M.A.: A surface radiation climatology across two Meteosat satellite generations. Remote Sens. Environ. **142**, 103–110 (2014). https://doi.org/10.1016/j.rse.2013.11.007

PVGIS: Photovoltaic Geographical Information System - Interactive Maps [WWW Document] (2016). http://re.jrc.ec.europa.eu/pvgis/apps4/pvest.php?map=africa. Accessed 2 Jan 2016

Remund, J., Müller, S., Kunz, S., Huguenin-Landl, B., Studer, C., Klauser, D., Schilter, C., Lehnherr, R.: Meteonorm: Global Meteorological Databases. Handbook Part I : Software v7 (2015)

Rigollier, C., Lefèvre, M., Wald, L.: The method Heliosat-2 for deriving shortwave solar radiation from satellite images. Sol. Energy **77**, 159–169 (2004). https://doi.org/10.1016/j.solener.2004.04.017

Šúri, M., Huld, T.A., Dunlop, E.D., Ossenbrink, H.A.: Potential of solar electricity generation in the European Union member states and candidate countries. Sol. Energy **81**, 1295–1305 (2007). https://doi.org/10.1016/j.solener.2006.12.007

Walther, B.A., Moore, J.L.: The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography **28**(6), 815–829 (2005)

WRDC: World Radiation Data Centre Online Archive (2016) [WWW Document]. http://wrdc-mgo.nrel.gov/. Accessed 30 Jan 2016

Youm, I., Sarr, J., Sall, M., Kane, M.M.: Renewable energy activities in Senegal: a review. Renew. Sustain. Energy Rev. **4**, 75–89 (2000). https://doi.org/10.1016/S1364-0321(99)00009-X

Zarzalejo, L.F., Polo, J., Martín, L., Ramírez, L., Espinar, B.: A new statistical approach for deriving global solar radiation from satellite images. Sol. Energy **83**, 480–484 (2009). https://doi.org/10.1016/j.solener.2008.09.006