J Ther Eng, Vol. 9, No. 3, pp. 776–785, May, 2023

Journal of Thermal Engineering
Web page info: https://jten.yildiz.edu.tr
DOI: 10.18186/thermal.1300542

Research Article

Statistical analysis of the solar diffuse fraction radiation using regression analysis of longitudinal data in India

Hicham SALHI[1], Abdelmounaim HADJIRA[2], Basharat JAMIL[3,*]

[1]*Laboratory of Applied Research in Hydraulics, University of Mustapha Ben Boulaid-Batna,16010, Algeria*
[2]*Department of Economics, Mohamed Boudiaf University, M'sila, 28000, Algeria*
[3]*Computer Science and Statistics Department, Universidad Rey Juan Carlos Móstoles, Mostoles, 28933, Madrid*

ARTICLE INFO

Article history
Received: 23 November 2021
Accepted: 19 January 2022

Keywords:
Longitudinal Data; Diffuse Fraction; Sunshine Ratio; Clearness Index; India

ABSTRACT

In this study, the validity of the estimation of a single regression equation for the diffuse fraction across 22 stations in India using the two parameters: the clearness index and the sunshine ratio is tested. The homogeneity test based on Fisher's statistics was applied to test the homogeneity of the estimated parameters across all stations. The results showed that the *p*-value at the level of 5% for each model is smaller than 0.05, indicating that all stations were heterogeneous. The Hierarchical Cluster Analysis (HCA) was used to classify the data into homogenous clusters. The results of HCA indicated that the longitudinal data were divided into four main clusters. For each cluster, the regression analysis was applied based on the longitudinal data then, the fixed effects model (FEM) and the random-effects model (REM) were used for the evaluation. Further, the Hausman test was applied to choose between the fixed effects model and the random-effects model. Finally, the results showed that the four best regression models were found for the selected stations in the study area.

Cite this article as: Salhi H, Hadjira A, Jamil B. Statistical analysis of the solar diffuse fraction radiation using regression analysis of longitudinal data in India. J Ther Eng 2023;9(3):776–785.

INTRODUCTION

Solar energy is essentially a renewable energy source, and it plays an important role among other alternative energy sources. For any solar energy research (Solar Electricity, Solar Water Heating, Solar Heating, Solar drying...) [1], solar radiation information in a specific geographic location is very important. India has a sufficient supply of solar energy throughout the year; the average sunshine hours are 2680 hours/year, which is enough to generate 6,081,709 TWh/year of environmentally friendly energy [2]. Many authors have presented empirical equations for estimating diffuse solar radiation using the clearness index, K_t [3-4]. For example, the study [5] estimated the monthly average diffuse radiation by examining the interrelationships using measured data from several sites in Turkey, the study concluded that the average daily spread rate is very related to the number of hours of sunshine.

Or Sunshine ratio (S_t) like the study [6] that proposed a new regression model that can predict daily global solar radiation independent of location. The proposed index quadratic

***Corresponding author.**
*E-mail address: basharat.jamil@urjc.es

This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkılıç

Published by Yıldız Technical University Press, İstanbul, Turkey

model captures the correlation between measured global solar radiation values, sunshine hours, and the air pollution index of Indian cities. This analysis of real data from Indian cities shows that air pollution is a more important factor than location when predicting solar radiation. Finally, the model parameters (regression coefficients) of each model are listed. In addition, generalized model equations for the best performance model are provided [7-9], or combine between clearness index and Sunshine ratio; Salih [10] suggested forty models estimate the diffuse fraction and forty other models to estimate the diffuse coefficient. The monthly average of solar radiation was analyzed at Tamanrasset station and after estimating and comparing the models, the study concluded that the cubic model with sunshine ratio and clearness index is the best accurate model for estimating the diffuse solar radiation in Algeria. There are also many studies combined between S_t and K_t [11–12]. For India, many researchers estimated solar radiation, [13] introduced a new method, i.e., Theory of Experimentation for prediction of monthly average global solar radiation in India. [14] Proposed models of Chennai and Trivandrum in 2015. They linked global radiation to temperature. Based on sunshine duration and air pollution index in India [15] presented a regression model to predict solar radiation without using location as the parameter, on the same basis [16] employed six empirical correlations to assess global solar radiation for Jaipur, India.

Jamil and Akhtar [17] performed a comparison of the models to evaluate the estimates for the monthly average diffuse solar radiation in the humid subtropical climate zone of India. The results of this study are valuable for locations in developing countries and remote areas with similar climatic conditions.

Saud et al. [18] considered twenty-five model forms selected from the literature to correlate the clearness index with the period of sunshine. The coefficients of the model were extracted from the data using k-fold cross-validation thereby improving the performance of the models. The data was divided into k-groups, and each group contains the same amount of data. The (k-1) group is used for model development, and the remaining group is used for model performance testing. This process is repeated k-times, and those coefficients that produce the smallest error are selected. Evaluation and comparison of the models were achieved through the use of statistical errors.

Jamil and Akhtar [19] compared ground-based global solar radiation measurements with available satellite data at the nearest coordinate location. A strong correlation was found between ground measurements and satellite data. In addition, using ground measurement data, models based on single and two input variables (ie, clear sky index and relative sunshine period) were developed to estimate diffuse solar radiation, thus empirically correlating monthly average diffuse solar radiation. 42 new models in 6 different categories were developed. The proposed models were also compared with the well-established models in the literature. Evaluation of the performance of the models was based on the ten most commonly used statistical indicators. It was inferred that the performance of the two-input variable model is much better than that of the single-variable input model. Among the two-variable models, the diffuse fraction model based on the clearness index and relative sunshine period (in order) was the most accurate. There was excellent agreement between the estimated and measured values from the two-variable model. It is also recommended to use univariate models within a reasonable range of accuracy.

The above-mentioned studies focused on proposing a single regression model that includes all stations without doing any statistical test to find out the homogeneity of the data from one station to another. The main objective of this paper is to prove the invalidity of a single regression model for 22 stations in India using Fisher's homogeneity tests, After confirming the heterogeneity of the regression parameters at all stations, a cluster analysis was used to classify the stations into homogeneous subsets. For each subset among these groups, a regression model on the longitudinal data is estimated: the fixed effects model (fem) and the random-effects model (rem); Haussman test is used to compare the two models (fem) and (rem). Finally, we obtain a regression model on the longitudinal data for each subset of stations.

METHODOLOGY

Solar radiation data

India is located between 8°4'-37°6' north latitude and 68°7'-97°25' east longitude. It is the seventh-largest country

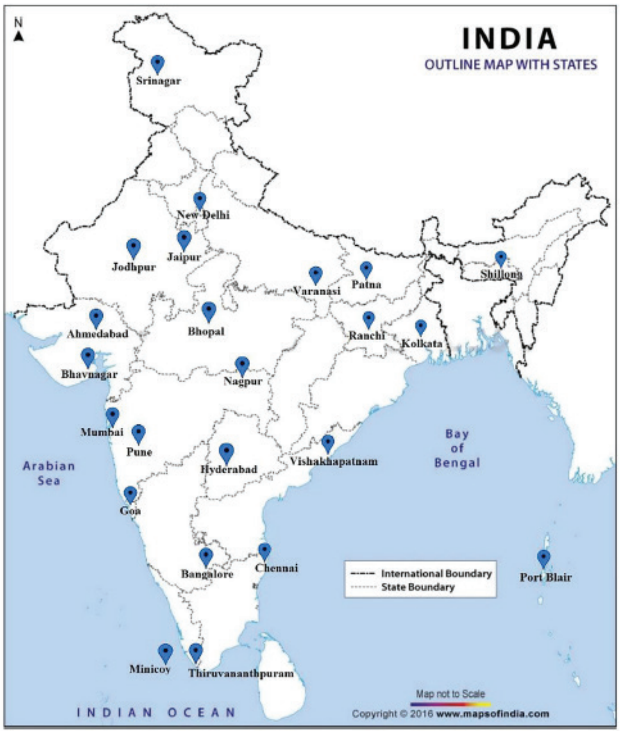

Figure 1. Solar radiation measurement facilities available under IMD (locations are marked on the map of India).

Table 1. Correlation matrix

	MNC	TRV	Blair	BNG	CHN	GOA	HYD	VSK	PNE	MMB	NGP	BHV	KLK	AHM	BHP	RNC	VNS	PTN	JDP	JPR	NDL	SRN
MNC	1																					
TRV	0.85	1																				
Blair	0.92	0.92	1																			
BNG	0.87	0.92	0.96	1																		
CHN	0.58	0.78	0.76	0.85	1																	
GOA	0.93	0.87	0.94	0.95	0.67	1																
HYD	0.89	0.85	0.93	0.96	0.75	0.98	1															
VSK	0.95	0.86	0.95	0.95	0.70	0.98	0.97	1														
PNE	0.91	0.82	0.92	0.95	0.73	0.98	0.99	0.98	1													
MMB	0.88	0.81	0.93	0.96	0.78	0.96	0.98	0.97	0.99	1												
NGP	0.92	0.81	0.92	0.93	0.70	0.98	0.98	0.98	1.00	0.98	1											
BHV	0.89	0.77	0.86	0.89	0.66	0.95	0.95	0.96	0.98	0.96	0.99	1										
KLK	0.97	0.84	0.95	0.92	0.66	0.97	0.95	0.98	0.96	0.96	0.97	0.94	1									
AHM	0.89	0.78	0.88	0.90	0.67	0.96	0.96	0.96	0.99	0.97	0.99	1.00	0.94	1								
BHP	0.91	0.80	0.90	0.90	0.67	0.96	0.96	0.97	0.98	0.96	0.99	0.99	0.95	0.99	1							
RNC	0.87	0.81	0.93	0.95	0.78	0.95	0.97	0.96	0.98	0.99	0.97	0.95	0.94	0.95	0.96	1						
VNS	0.88	0.74	0.84	0.84	0.70	0.86	0.87	0.91	0.91	0.91	0.92	0.93	0.92	0.93	0.91	0.88	1					
PTN	0.92	0.76	0.85	0.84	0.68	0.87	0.87	0.92	0.91	0.91	0.92	0.92	0.93	0.92	0.92	0.89	0.99	1				
JDP	0.91	0.76	0.82	0.80	0.52	0.92	0.89	0.91	0.92	0.87	0.94	0.96	0.91	0.96	0.96	0.85	0.89	0.90	1			
JPR	0.86	0.75	0.81	0.83	0.62	0.90	0.90	0.91	0.94	0.90	0.95	0.98	0.89	0.97	0.96	0.89	0.91	0.90	0.97	1		
NDL	0.92	0.77	0.84	0.84	0.64	0.91	0.90	0.92	0.93	0.91	0.95	0.97	0.93	0.96	0.95	0.89	0.96	0.96	0.97	0.97	1	
SRN	-0.52	-0.65	-0.63	-0.57	-0.24	-0.63	-0.56	-0.54	-0.48	-0.48	-0.48	-0.41	-0.55	-0.42	-0.46	-0.50	-0.26	-0.29	-0.43	-0.34	-0.33	1

Table 2. The p-value of correlation coefficients

	MNC	TRV	Blair	BNG	CHN	GOA	HYD	VSK	PNE	MMB	NGP	BHV	KLK	AHM	BHP	RNC	VNS	PTN	JDP	JPR	NDL	SRN
MNC																						
TRV	5.02E-04																					
Blair	1.73E-05	2.86E-05																				
BNG	2.41E-04	2.71E-05	1.03E-06																			
CHN	4.68E-02	2.73E-03	3.90E-03	4.71E-04																		
GOA	9.08E-06	2.27E-04	5.05E-06	2.62E-06	1.75E-02																	
HYD	1.18E-04	4.60E-04	9.77E-06	7.91E-07	4.80E-03	1.04E-08																
VSK	1.78E-06	3.78E-04	2.48E-06	1.97E-06	1.17E-02	1.31E-08	2.33E-07															
PNE	5.00E-05	1.16E-03	2.93E-05	3.34E-06	7.16E-03	5.08E-08	3.44E-09	1.11E-08														
MMB	1.42E-04	1.30E-03	1.22E-05	1.25E-06	2.63E-03	9.79E-07	1.37E-08	2.57E-07	1.12E-09													
NGP	2.77E-05	1.28E-03	2.59E-05	1.34E-05	1.14E-02	4.67E-08	1.42E-08	2.34E-08	1.05E-11	1.44E-08												
BHV	1.02E-04	3.54E-03	3.01E-04	1.16E-04	2.02E-02	1.78E-06	1.54E-06	1.15E-06	1.20E-08	8.52E-07	8.66E-10											
KLK	7.42E-08	5.65E-04	2.39E-06	2.88E-05	1.91E-02	3.55E-07	2.24E-06	3.15E-08	7.66E-07	1.28E-06	3.79E-07	7.94E-06										
AHM	8.58E-05	2.92E-03	1.67E-04	7.70E-05	1.66E-02	9.56E-07	4.74E-07	5.39E-07	2.59E-09	2.74E-07	3.58E-11	1.18E-13	4.05E-06									
BHP	4.08E-05	1.78E-03	5.22E-05	5.52E-05	1.81E-02	4.23E-07	7.33E-07	3.20E-07	1.99E-08	6.96E-07	2.27E-10	6.88E-10	3.11E-06	9.49E-11								
RNC	2.42E-04	1.46E-03	1.09E-05	1.41E-06	3.05E-03	2.67E-06	3.20E-07	9.26E-07	3.93E-08	1.38E-10	1.72E-07	2.72E-06	4.92E-06	1.41E-06	1.33E-06							
VNS	1.47E-04	5.50E-03	7.18E-04	7.13E-04	1.05E-02	3.54E-04	2.35E-04	3.97E-05	3.50E-05	4.15E-05	2.23E-05	1.49E-05	2.90E-05	1.18E-05	3.61E-05	1.48E-04						
PTN	2.92E-05	4.32E-03	4.07E-04	5.85E-04	1.51E-02	2.27E-04	2.77E-04	2.43E-05	3.93E-05	4.53E-05	2.43E-05	1.87E-05	9.56E-06	1.70E-05	2.97E-05	1.15E-04	8.10E-10					
JDP	3.77E-05	3.76E-03	1.10E-03	1.68E-03	8.52E-02	1.92E-05	1.02E-04	3.43E-05	2.36E-05	2.25E-04	3.95E-06	5.34E-07	3.74E-05	9.50E-07	7.05E-07	4.15E-04	1.04E-04	7.51E-05				
JPR	3.51E-04	4.94E-03	1.51E-03	9.42E-04	3.08E-02	6.81E-05	7.40E-05	4.68E-05	7.19E-06	6.42E-05	1.65E-06	2.74E-08	1.21E-04	1.11E-07	4.04E-07	1.10E-04	3.27E-05	6.35E-05	1.48E-07			
NDL	2.07E-05	3.23E-03	5.52E-04	6.19E-04	2.50E-02	4.44E-05	6.91E-05	1.99E-05	8.71E-06	3.82E-05	2.10E-06	3.14E-07	9.79E-06	6.11E-07	1.43E-06	1.02E-04	1.19E-06	5.50E-07	2.73E-07	2.32E-07		
SRN	8.03E-02	2.26E-02	2.79E-02	5.30E-02	4.61E-01	2.91E-02	5.75E-02	7.12E-02	1.16E-01	1.17E-01	1.12E-01	1.84E-01	6.53E-02	1.69E-01	1.35E-01	1.01E-01	4.11E-01	3.57E-01	1.62E-01	2.86E-01	2.91E-01	

in the world, with a land area of 2.9 million square kilometers. Horizontal global solar radiation, diffuse solar radiation, and sunshine period data from 23 stations in India [20] taken from the Indian Meteorological Service [21] from 1986 to 2000.

Table 1 & 2 shows the correlation coefficient and its significance between the stations. From the results, we can observe that all stations are positively correlated and statistically significant at the 5% level except the Srinagar (SRN) station which is negatively correlated with all stations

Fisher's tests of parameters homogeneity

We assume the following models in which the estimated parameters are constant over time but change from one station to another. Based on the literature, most studies have proven that the best models in India are [17–19]:

$$M_1:\ \left(\frac{H_d}{H}\right)_{it} = \alpha_i^* + \beta_{1i}\left(\frac{S}{S_0}\right)_{it} + \beta_{2i}\left(\frac{H}{H_0}\right)_{it} + \varepsilon_{it}\ \ ; i = 1 \dots\dots 22; t = 1 \dots\dots 12 \quad (1)$$

$$M_2: \left(\frac{H_d}{H}\right)_{it} = \alpha_i^* + \beta_{1i}\left(\frac{S}{S_0}\right)_{it} + \beta_{2i}\left(\frac{S}{S_0}\right)^2 + \beta_{3i}\left(\frac{S}{S_0}\right)^3 + \beta_{4i}\left(\frac{S}{S_0}\right)^4 + \varepsilon_{it}\ \ ; i = 1 \dots 22\ ; t = 1 \dots 12 \quad (2)$$

$$M_3: \left(\frac{H_d}{H}\right)_{it} = \alpha_i^* + \beta_{1i}\left(\frac{H}{H_0}\right)_{it} + \varepsilon_{it}\ \ ; i = 1 \dots\dots 22; t = 1 \dots\dots 12 \quad (3)$$

We use homogeneity tests based on Fisher's statistic [22] to test the homogeneity of the estimated parameters across all stations. The first tested hypothesis is as follows:

$$H_0^1: \alpha_i^* = \alpha^*\ \ ; \beta_{1i} = \beta_1\ ; \beta_{2i} = \beta_2 \quad (4)$$

To test this hypothesis, we use the following Fisher statistic:

$$F_1 = \frac{(RSS_{C1} - RSS)/(N-1)(K+1)}{RSS/(N*T - N(K+1)} \quad (5)$$

Where: RSS_{C1} is the residual sum of squares of the restricted model according to the hypothesis H_0^1 and we get it by estimating the pooled OLS model. The degrees of freedom are equal to the number of total observations ($N * T$) minus a $K + 1$ parameter. RSS is the residual sum of squares for the models estimated using T the number of observations for each station:

$$RSS = \sum_{i}^{N} RSS_i \quad (6)$$

The degrees of freedom are equal to the sum of N degrees of freedom for each station.

$$df = \sum_{i=1}^{N}(T - (K+1) = N*T - (K+1) \quad (7)$$

Degrees of freedom in the denominator is equal to the difference between degrees of freedom of RSS_{C1} and SS:

$$\text{fd} = [(N*T) - (k+1)] - [(N*T) - N(k+1)] = (N-1)(k+1) \quad (8)$$

The F1 statistic is compared with the tabular value of the Fisher distribution at the degrees of freedom of the numerator and denominator, respectively, and at the level 5%. if $F_1 > F^{0.005}_{df1\,;df2}$ We reject the Null hypothesis H_0^1.

The second tested hypothesis is as follows:

$$H_0^2:\ \beta_{1i} = \beta_1\ ;\ \beta_{2i} = \beta_2\ \forall\ i \quad (9)$$

To test this hypothesis, we use the following Fisher statistic:

$$F_2 = \frac{(RSS_{C2} - RSS)/(N-1)K}{RSS/(N*T - N(K+1)} \quad (10)$$

Where: RSS_{C1} is the residual sum of squares of the restricted model according to the hypothesis H_0^2, and we get it by estimating the individual effect model. The degrees of freedom are equal to the total observation $N \times T$ munis $N + K$ parameter (We estimate a K parameter and an N constant). We reject the null hypothesis if $F_2 > F^{0.005}_{df1\,;df2}$.

The third hypothesis tested is:

$$H_0^3: \alpha_i^* = \alpha^*\ \forall\ i \quad (11)$$

To test this hypothesis, we use the following Fisher statistic:

$$F_3 = \frac{(RSS_{C1} - RSS_{C2})/(N-1)}{RSS_{C2}/(N(T-1) - K} \quad (12)$$

if $F_3 > F^{0.005}_{df1\,;df2}$ We reject the Null hypothesis H_0^3.

Hierarchical cluster analysis (HCA)

Through the results of the homogeneity tests, we conclude that the model parameters cannot be equal across all stations, and therefore we applied the Hierarchical cluster analysis to classify the stations into homogeneous groups.

Models for longitudinal data

Based on the Cluster analysis, through which we divided the data into four main groups (the first group, the second, and the third group consisting of 7 stations, and the last group consisting of only one station. We use regression analysis on the longitudinal data where we estimate the fixed effects model (fem) and the random-effects model (rem). For each of the three groups (we will exclude the last group because it consists of one station). Under the fixed effects specification the constants α_i^* are considered as free parameters which are incidental to the analysis, with β being the center of concern. while the random effects model suggests that α_i^* are realizations from a probability distribution function with a finite number of parameters, distributed independently of the regressors [23], and then we will use the Hausman test [24] to compare between the fixed effects model and the random-effects model, the null hypothesis of this test is: H_0: $E(\varepsilon_{it} \backslash x_{it}) = 0$, This hypothesis is considered essential in the random-effects model. Under this hypothesis, the random-effects model estimated by the generalized least squares method is the most appropriate one, and the test statistic is given as follows[24]:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \quad (13)$$

The statistic H is distributed as χ^2 under the null hypothesis with the degree of freedom corresponding to the dimension of β. If $H > \chi^2$ we reject the null hypothesis at level α%.

RESULTS AND DISCUSSION

Homogeneity Tests

Table 3 shows the results of the homogeneity tests for the three models: M_1, M_2 and M_3, where the results of the H_0^1 hypothesis test show that the statistic F_1 greater than the tabulated value of the Fisher distribution in the three models which means that the hypothesis H_0^1 is rejected at the level of 5%. From this, we conclude that the parameters are not homogeneous between all stations. This is confirmed by the p-value which was exactly smaller than 0.05 in the following models.

For the test of H_0^2 hypothesis, the value of F calculated in the three models shows that it is greater than the tabulated F, and from it, we reject the H_0^2 hypothesis at a level of 5% and this is confirmed by a p-value that is smaller than 0.05.

For the H_0^3 the hypothesis that tests the homogeneity of the constants in the panel data model, the results show that the value of F is also completely greater than the tabulated value in the three models, and from this, we reject the H_0^3 hypothesis of homogeneity of the constants at the level of 5% and this is confirmed by a p-value that is smaller than 0.05.

Table 3. Homogeneity tests for the three models

		Calculate Value	Tabulate value	p-value
M_1	F_1	8.8896775	1,36944711	2.721e-11
	F_2	6.4686733	1,43501381	2.450e-07
	F_3	7.3236137	1,58482692	3.502e-06
M_2	F_1	3.7519397	1,33745801	5.439e-14
	F_2	1.6917328	1,36131377	.00247467
	F_3	9.6393957	1,60048735	7.358e-22
M_3	F_1	18.359348	1,44018632	1.446e-51
	F_2	4.6818323	1,60415548	1.332e-09
	F_3	24.255218	1,59992937	7.432e-48

Selection of the best number of clusters

Performed zoning to detect the most similar sites and group them into groups. For this reason, the method and zoning standard of solar radiation zoning must be selected. The different methods to calculate the number of clusters are available in Table 4 and Figure 2. According to the majority rule (40%), the best number of clusters is 4.

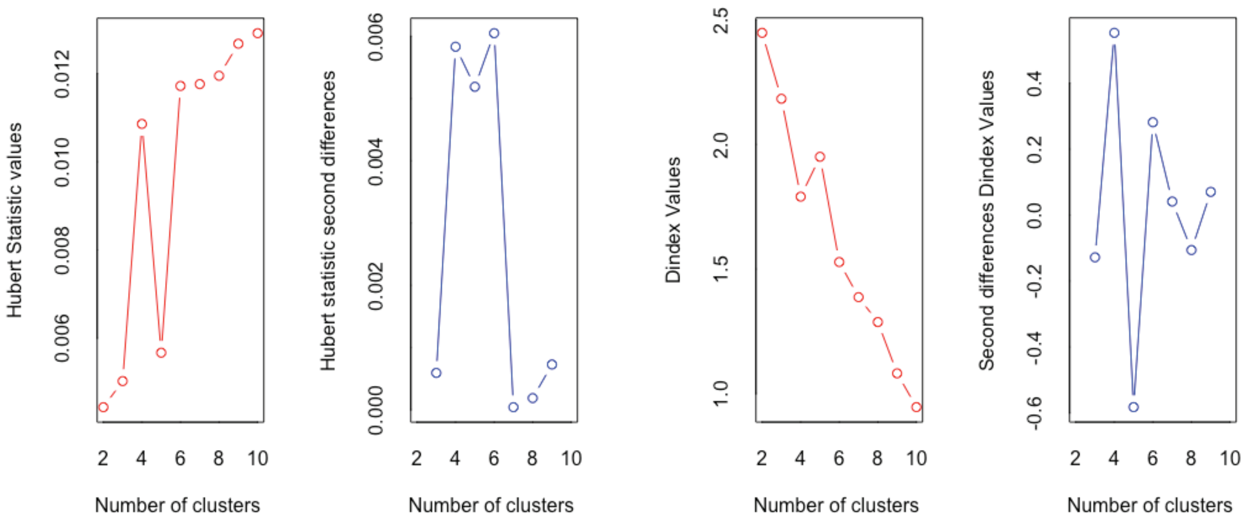

Figure 2. Optimal number of clusters.

Table 4. The different methods calculate the number of clusters

	Number_clusters	Value_Index
KL	6	6.3478
CH	4	12.9456
Hartigan	5	19.157
CCC	4	-0.8011
Scott	10	128.9526
Marriot	4	572952.9
TrCovW	4	415.9892
TraceW	4	86.901
Friedman	10	26029.01
Rubin	4	-2.1706
Cindex	2	0.2992
DB	10	0.7448
Silhouette	10	0.4134
Duda	2	0.9052
PseudoT2	2	1.2564
Beale	2	0.7694
Ratkowsky	4	0.4123
Ball	3	41.893
PtBiserial	4	0.4599
Frey	1	*NA*
McClain	2	0.8061
Dunn	10	0.4887
Hubert	0	0
SDindex	4	1.1432
Dindex	0	0
SDbw	10	0.1272

In this study, we applied the hierarchical clustering analysis using Ward's algorithm Ward JH [25] to understand the spatial distribution of the diffuse fraction in the study area.

From Figure 3, we can see that the cluster analysis identified four homogeneous groups of the diffuse fraction. Group 1 (G1) was composed of stations (NDL, PTN, CHN, KLK), located on the humid subtropical Zone Figure 4, with tropical wet and dry for stations (MNC, VNS, TRV). Regarding the Group 2 and 3 cohorts, they are distributed over all regions of the country. The last group (G4), consists of one station (SNR) located in mountains.

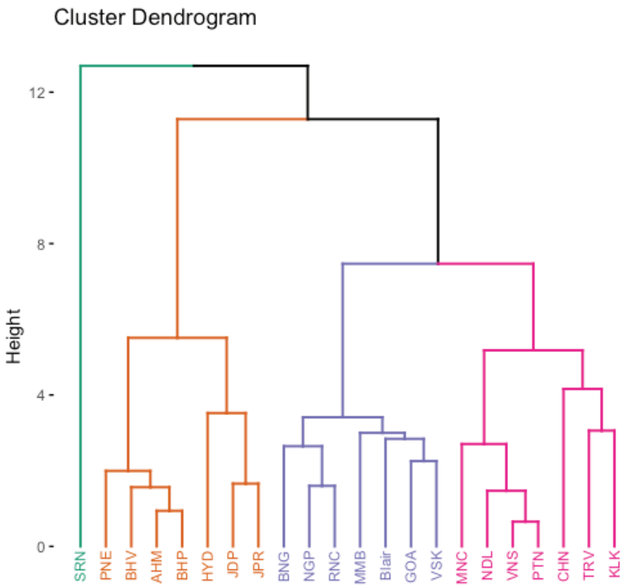

Figure 3. Cluster analysis groups based on the spatial grouping of stations.

Figure 4. Climatic zones in India based on Köppen-Geiger climate classification.

Estimation model

Table 5 shows the results of estimating the fixed effects model and the random-effects model for the proposed model M_1 and for the three groups $\{G_1, G_2, G_3\}$, as there do not seem to be significant differences in the estimation results between the (FEM) and the (REM), to compare between these two models we use the Haussman test whose results appear in Table 6 for the first group, the $p - value = 0.9244 > 0.05$ and from it, we accept the null hypothesis, therefore, the (REM) is the most appropriate model for G_1. For G_2 the $p - value = 0.0004 < 0.05$, we reject the null hypothesis that the random effect model is the appropriate model, so the fixed effect model is the appropriate model. For the G_3 the appropriate model according to the Hausman test is the (REM).

$$M_1:\ \left(\frac{H_d}{H}\right)_{it} = \alpha_i^* + \beta_{1i}\left(\frac{S}{S_0}\right)_{it} + \beta_{2i}\left(\frac{H}{H_0}\right)_{it} + \varepsilon_{it}\ ;\ i = 1 \ldots\ldots 22;\ t = 1 \ldots\ldots 12 \quad (14)$$

Table 5. estimating of fixed effect model and random effect model for M1

Group	parameters	FEM	REM
G_1	α_i^*	1.356985	1.35583
	β_{1i}	-0.3302712	-0.3323464
	β_{2i}	-1.197594	-1.192939
G_2	α_i^*	1.299473	1.291561
	β_{1i}	-0.055939	-0.039764
	β_{2i}	-1.502283	-1.509459
G_3	α_i^*	1.227283	1.223239
	β_{1i}	-0.3276291	-0.319616
	β_{2i}	-1.019594	-1.022661

Regarding the model M_2, the results of estimating the (FEM) and the (REM) are shown in Table 7, and the results of the Hausman test appear in Table 8. Through these results, it is clear that the (REM) is the most appropriate model for group G_1 and the (FEM) is the most appropriate model for the second group G_2 And the third group G_3.

$$M_2:\ \left(\frac{H_d}{H}\right)_{it} = \alpha_i^* + \beta_{1i}\left(\frac{S}{S_0}\right)_{it} + \beta_{2i}\left(\frac{S}{S_0}\right)^2 + \beta_{3i}\left(\frac{S}{S_0}\right)^3 + \beta_{4i}\left(\frac{S}{S_0}\right)^4 + \varepsilon_{it}\ ;\ i = 1 \ldots 22;\ t = 1..12 \quad (15)$$

Table 7. estimating of Fixed effect model and Random effect model for M2

Group	parameters	FEM	REM
G_1	α_i^*	0.976956	0.7718304
	β_{1i}	-2.06129	-1.17832
	β_{2i}	8.857116	7.755474
	β_{3i}	-15.69614	-15.52742
	β_{4i}	8.24894	8.528482
G_2	α_i^*	2.384721	2.363621
	β_{1i}	-14.04328	-14.67045
	β_{2i}	42.44249	45.65797
	β_{3i}	-54.12448	-59.18861
	β_{4i}	23.79596	26.37172
G_3	α_i^*	2.547389	2.356013
	β_{1i}	-13.80677	-12.82972
	β_{2i}	38.48113	36.36594
	β_{3i}	-47.26369	-45.12562
	β_{4i}	20.54023	19.74879

Table 6. Hausman's test for M1

Group	chi-squared	Prob > Kh^2	The best model
G_1	0.16	0.9244	$\left(\frac{H_d}{H}\right)_{it} = 1.35583 + -0.3323464\left(\frac{S}{S_0}\right)_{it} + -1.192939\left(\frac{H}{H_0}\right)_{it}$
G_2	15.61	0.0004	$\left(\frac{H_d}{H}\right)_{it} = 1.299473 - 0.055939\left(\frac{S}{S_0}\right)_{it} - 1.502283\left(\frac{H}{H_0}\right)_{it}$
G_3	2.81	0.2452	$\left(\frac{H_d}{H}\right)_{it} = 1.223239 - 0.3276291\left(\frac{S}{S_0}\right)_{it} - 1.019594\left(\frac{H}{H_0}\right)_{it}$

Table 8. Hausman's test for M2

Groupe	chi-squared	Prob > Kh^2	The best model
G_1	6.10	0.1922	$\left(\frac{H_d}{H}\right)_{it} = 0.7718304 + -1.17832\left(\frac{S}{S_0}\right)_{it} + 7.755474\left(\frac{S}{S_0}\right)^2 - 15.52742\left(\frac{S}{S_0}\right)^3 + 8.528482\left(\frac{S}{S_0}\right)^4$
G_2	24.55	0.0001	$\left(\frac{H_d}{H}\right)_{it} = 2.384721 - 14.04328\left(\frac{S}{S_0}\right)_{it} + 42.44249\left(\frac{S}{S_0}\right)^2 - 54.12448\left(\frac{S}{S_0}\right)^3 + 23.79596\left(\frac{S}{S_0}\right)^4$
G_3	19.29	0.0007	$\left(\frac{H_d}{H}\right)_{it} = 2.547389 - 13.80677\left(\frac{S}{S_0}\right)_{it} + 38.48113\left(\frac{S}{S_0}\right)^2 - 47.26369\left(\frac{S}{S_0}\right)^3 + 20.54023\left(\frac{S}{S_0}\right)^4$

For the model M_3, the results of estimating the (FEM) and the (REM) are shown in Table 9. Also, the results of the Hausman test are shown in Table 10. Through these results, we conclude that the (FEM) is the most appropriate model for all groups $\{G_1, G_2, G_3\}$, as the p-value of Haussmann's test was smaller than 0.05.

$$M_3: \left(\frac{H_d}{H}\right)_{it} = \alpha_i^* + \beta_{1i}\left(\frac{H}{H_0}\right)_{it} + \varepsilon_{it}\ ; i = 1 \dots 22; t = 1 \dots 12 \quad (16)$$

Table 9. estimating of Fixed effect model and Random effect model for M3

Group	parameters	FEM	REM
G_1	α_i^*	1.36585	1.360788
	β_{1i}	-1.645552	-1.637078
G_2	α_i^*	1.298711	1.291739
	β_{1i}	-1.576315	-1.563415
G_3	α_i^*	1.204373	1.196595
	β_{1i}	-1.406264	-1.391667

Table 10. Hausman's test for M3

Groupe	chi-squared	Prob > Kh^2	The best model
G_1	4.33	0.0374	$(\frac{H_d}{H})_{it} = 1.36585 + -1.645552\left(\frac{H}{H_0}\right)_{it}$
G_2	13.07	0.0003	$(\frac{H_d}{H})_{it} = 1.298711 - 1.576315\left(\frac{H}{H_0}\right)_{it}$
G_3	4.26	0.0390	$(\frac{H_d}{H})_{it} = 1.204373 - 1.406264\left(\frac{H}{H_0}\right)_{it}$

CONCLUSION

In the current study, solar radiation data was used to evaluate the diffuse fraction, the sunshine ratio, and the clearness index. Here, we selected 22 stations in India. The homogeneity test, Hierarchical Cluster Analysis (HCA), fixed effects model (FEM), random-effects model (REM), and Hausman test were applied in this study.

The main conclusions can be summarized as follows:

1) Fisher's test of the homogeneity of the regression parameters proved that the choice of the unified model for all stations was wrong. The Fisher values of the three models are equal to 8.88, 6.46, and 7.32, respectively. This is confirmed by the p-value which was exactly smaller than 0.05 in the three models.
2) Four homogeneous groups of the diffuse fraction were identified using Hierarchical Cluster Analysis. Groups 1, 2, and 3 consists of 7 stations, and the last group (G4) consists of a station (SNR) located in a mountainous area.
3) The Haussmann test is used to compare these models, and the results show that each group contains three empirical equations for estimating diffuse solar radiation using the clearness index or Sunshine ratio or combining between clearness index and Sunshine ratio.
4) The main objective of this paper is to prove the invalidity of one regression model for all stations in India. We advise researchers in this field to take this observation in future studies.

NOMENCLATURE

H_d diffuse solar radiation on a horizontal surface (MJ/m^2-day)
H global solar radiation on a horizontal surface (MJ/m^2-day)
H_0 extraterrestrial radiation
k_d diffuse fraction (or cloudiness index)
K_t clearness index
S_t sunshine ratio
S_o maximum possible sunshine duration (hours)
S sunshine duration (hours)

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Natarajan K, Thokchom SS, Verma TN, Nashine P. Convective solar drying of Vitis vinifera & Momordica charantia using thermal storage materials. Renew Energy 2017;113:1193−1200. [CrossRef]

[2] Kar SK, Sharma A, Roy B. Solar energy market developments in India. Renew Sustain Energy Rev 2016;62:121−133. [CrossRef]

[3] Li H, Ma W, Wang X, Lian Y. Estimating monthly average daily diffuse solar radiation with multiple predictors: a case study. Renew Energy 2011;36:1944−1948. [CrossRef]

[4] Khorasanizadeh H, Mohammadi K. Diffuse solar radiation on a horizontal surface: Reviewing and categorizing the empirical models. Renew Sustain Energy Rev 2016;53:338−362. [CrossRef]

[5] Taşdemiroglu E, Sever R. Estimation of monthly average, daily, horizontal diffuse radiation in Turkey. Energy 1991;16:787−790. [CrossRef]

[6] Suthar M, Singh GK, Saini RP. Effects of air pollution for estimating global solar radiation in India. Int J Sustain Energy 2017;36:20−27.

[7] Raja Sekhar Y, Sharma KV, Rao AS. New correlations for estimation of monthly average daily solar radiation on a horizontal surface using meteorological data. Int J Ambient Energy 2013;34:160−714. [CrossRef]

[8] Aras H, Balli O, Hepbasli A. Global solar radiation potential, part 1: model development. Energy Sources 2006;1:303−315. [CrossRef]

[9] Ouali K, Alkama R. A new model of global solar radiation based on meteorological data in Bejaia City (Algeria). Energy Procedia 2014;50:670−676. [CrossRef]

[10] Salhi H, Belkhiri L, Tiri A. Evaluation of diffuse fraction and diffusion coefficient using statistical analysis. Appl Water Sci 2020;10:1−12. [CrossRef]

[11] Bailek N, Bouchouicha K, Al-Mostafa Z, El-Shimy M, Aoun N, Slimani A, et al. A new empirical model for forecasting the diffuse solar radiation over Sahara in the Algerian Big South. Renew Energy 2018;117:530−537. [CrossRef]

[12] Kambezidis HD, Psiloglou BE, Karagiannis D, Dumka UC, Kaskaoutis DG. Meteorological Radiation Model (MRM 274 v6. 1): Improvements in diffuse radiation estimates and a new approach for implementation of cloud products. Renew Sustain Energy Rev 2017;74:616−637. [CrossRef]

[13] Makade RG, Chakrabarti S, Jamil B, Sakhale CN. Estimation of global solar radiation for the tropical wet climatic region of India: A theory of experimentation approach. Renew Energy 2020;146:2044−2059. [CrossRef]

[14] Ravichandran S, Rathnaraj J. Modeling of global solar radiation by using ambient air temperature at coastal cities in India. Int J Appl Eng Res 2015;10:16843−16852.

[15] Suthar M, Singh GK, Saini RP. Effects of air pollution for estimating global solar radiation in India. Int J Sustain Energy 2017;36:20−27. [CrossRef]

[16] Marwal VK, Punia RC, Sengar N, Mahawar S, Dashora P. A comparative study of correlation functions for estimation of monthly mean daily global solar radiation for Jaipur, Rajasthan (India). Indian J Sci Technol 2012;5:2729−2732. [CrossRef]

[17] Jamil B, Akhtar N. Comparison of empirical models to estimate monthly mean diffuse solar radiation from measured data: Case study for humid-subtropical climatic region of India. Renew Sustain Energy Rev 2017;77:1326−1342. [CrossRef]

[18] Saud S, Jamil B, Upadhyay Y, Irshad K. Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach. Sustain Energy Technol Assess 2020;40:100768. [CrossRef]

[19] Jamil B, Akhtar N. Comparative analysis of diffuse solar radiation models based on sky clearness index and sunshine period for humid-subtropical climatic region of India: A case study. Renew Sustain Energy Rev 2017;78:329−355. [CrossRef]

[20] Maps of India. Information about India Maps. Available at: https://www.mapsofindia.com/maps/india/ Accessed on October 21, 2017.

[21] Tyagi AP. Solar Radiant Energy Over India. India: Indian Meteorol Dep Minist Earth Sci 2009:4179.

[22] Hsiao C. Analysis of panel data. 3rd ed. Cambridge: Cambridge University Press; 2014.

[23] Pesaran MH. Time series and panel data econometrics. 1st ed. Oxford: Oxford University Press; 2015. [CrossRef]

[24] Hausman JA. Specification tests in econometrics. Econometrica 1978;46:1251−1271. [CrossRef]

[25] Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58:236−244. [CrossRef]