

The mapping between satellite and ground station data for the estimation of average solar radiation in hilly regions of India

S. Gupta^{1*} and A. K. Singh¹

¹Department of Instrumentation and Control Engineering, Dr. B R Ambedkar National Institute of Technology, Jalandhar, India, 144008

Received Date 07 May 2025; Revised Date 26 October 2025; Accepted Date 30 November 2025

*Corresponding author: shubhamg.ic.19@nitj.ac.in (S.Gupta)

Abstract

This study is performed to increase the bankability and assess the future performance of solar-energy-based applications in hilly areas of India. A solar energy database for these places is necessary to achieve this. There are many obstacles to building measurement stations at these places. Significant problems include expensive measuring instruments, transportation, periodic calibration, and maintenance of the instruments. Utilizing satellite sources can overcome these challenges. This study uses the NASA and ISRO solar radiation databases as satellite sources. The ground station data was obtained from the Center of Energy and Environment (CEE), NIT, Hamirpur. Both overall and seasonal data analysis are performed. The concise mathematical relationship between ground station measured value and satellite radiation has a higher prediction accuracy than the other studies performed for the same study area (site). The lowest mean absolute percentage error (MAPE) obtained in this study was 7.3688%. The present study reveals that satellite solar radiation is a good alternative to developing a long-duration global solar radiation database in the hilly region of India.

Keywords: *surface solar radiation, ground station GSR, satellite data, seasonal analysis, NASA, ISRO.*

1. Introduction

Solar radiation data is essential in the financial planning and performance evaluation of solar energy-based applications. Solar radiation data is also widely applied to calculate buildings' cooling and heating load, lighting requirements in residential areas, the agriculture sector, forestry, weather study, and water resources. The users sometimes procure solar radiation data from the data owner. There is a need to validate these data by the end users before use to improve reliability. A convenient way to validate this data is through gridded data. Due to advancements in remote sensing technology, the upgraded version of gridded solar radiation products is available every few years [1]. The most accurate way to estimate solar radiation data at any location is by pinpointing solar radiation data and its temporal integrated value recording [2]. The high cost of measuring instruments, setup difficulty, periodic calibration, and maintenance difficulty make it inapt for most locations, especially in difficult terrain. The recorded solar radiation data also

suffer from unknown instrument faults and human errors. Solar radiation must be estimated using alternative methods to avoid the measurement setup. Many alternative methods estimate solar radiation at the location of interest. Majorly these are classified as (i) nearby station method [3] (ii) interpolation based on the spatially subjective average of different nearby measurement station data (iii) estimation through measured metrological variables (iv) empirical model based on solar constant (v) Satellite solar radiation data

In the nearby station method, pinpoint ground station measured data are used to estimate solar radiation at a distance by interpolation. Geuder [4] explains that the ground station data are unsuitable for locations far from 25 km from the measurement station. He explained that satellite data are more appropriate and reliable in this case. The model based on the metrological variables again requires the measurement of metrological variables, which are also not available at every location in the world. A similar is true with the model based on solar constant. Satellite data is

the data that is available in most of the locations of the world. Satellite data is derived from the geostationary and polar-orbiting satellites' images. The satellite data is the gridded information. The amount of solar radiation is fixed in the grid area, and the magnitude of radiation changes/varies in the next grid. The change in the magnitude of satellite data depends on the resolution of satellite data sources. So, estimating data at the desired location again requires ground station and satellite data. More details are available in reference [5]. It is difficult to establish the measurement setup in hilly regions of India for all the measurements, i.e., direct solar radiation measurement, measurement of metrological variables, and solar constant.

This study incorporates satellite data to develop a solar radiation database because it is available for hilly regions of India. More precisely this study explains how to find the best-fit model for estimating solar radiation with the help of satellite data. Relevant studies in this field are discussed below.

Assi et al. [1] study shows that direct use of satellite solar radiation and ground station radiation has a higher value of R^2 . This magnitude of R^2 shows further scope for developing different, concise mathematical models with more metrics.

Olomiyesan et al. [7] found that sunshine duration and air temperature are the most relevant input variables to develop highly accurate estimation models. Their study explained that satellite data has reasonable accuracy when compared with empirical statistical models.

Palmer et al. [8] evaluated the reliability of solar radiation data in the UK by comparing satellite estimates with ground station records from 84 UK Met Office stations (MIDAS archive). Satellite datasets such as SARAHE and Solargis, with spatial resolutions ranging from 250 m to 5.6 km and hourly temporal resolution, used inputs including satellite imagery, aerosols, water vapor, ozone, snow, elevation, and terrain shading. Their analysis identified the most reliable data sources depending on the location.

Masoom et al. [9] explain the modern techniques used to obtain global horizontal irradiance (GHI) and direct normal irradiance (DNI) data for specific Indian locations. They found that in cloudy conditions satellite resources are not a good alternative.

The higher solar energy potential is available in the hilly region of India. Mishra et al. [10] explain how it is utilized. They suggest the procedure for how solar PV systems are deployed in the hilly region of India. The challenge in the hilly region was getting the plane surface for installing Solar PV systems. The plane surface is the roof of the houses, and rooftop solar panel implementation is a good alternative to utilize the solar potential of hilly areas of Uttarakhand, India. They estimated the roof area in Uttarakhand with the help of Geographic Information System (GIS) and remote sensing techniques. The fixed map of solar irradiance provided by the Ministry of New & Renewable Energy (MNRE) was used further to calculate the power output of the rooftop SPV system. They explain the SPV system power output variation in cloudy or uncloudy conditions.

Manara et al. [11] explained how the satellite sources were utilized for different geographic locations over the Piedmont region (north-western Italy) to estimate solar radiation. They present seasonal (winter, spring, summer and autumn) and complete year analysis for low and high (orographic complexity index) OCI regions of northwestern Italy. They explain that at high OCI, it is challenging to acquire ground station and satellite data.

At higher altitudes, solar potential is greater due to shorter atmospheric paths, reduced terrestrial losses, and lower dust, making such sites suitable for SPV systems. Mishra et al. [10] highlighted this potential in Uttarakhand, India, while Carpentieri et al. [12] studied the Swiss Alps using SARAHE and Helimont satellite data. They found estimation accuracy improves with longer time scales, being lower for intra-hour and daily data but higher for monthly and annual averages, despite using high-resolution satellite sources.

Kumar Yadav and Chandel [13] developed artificial neural network (ANN) models to estimate solar radiation in 12 different cities of India. The higher-elevation cities Shillong (1598m), Mangalore (897m), and Dehradun (683m) are also considered which are areas of interest for the present study. The explanatory variables of the ANN are the latitude, longitude, sunshine hours, and solar radiation, which are the target variables. Monthly mean solar radiation on a horizontal surface, mean sunshine duration per hour, and height above sea level are the data obtained from the India

Meteorological Department [14]. Mani, A. [15] compiled this data. MAPE values of the ANN model are considered for the 8 cities. They do not cover the higher-elevation city Shillong (1598m) MAPE. However, the use of satellite sources is not preferred by them.

Yadav et al. [16] extended their work by incorporating additional variables such as temperature (T, T_{max}, T_{min}), altitude (H), sunshine hours (SH), clearness index, extraterrestrial radiation, latitude, and longitude. Temperature data were sourced from NASA POWER, while SH and radiation data came from three ground stations (Hamirpur data from CEE, NIT Hamirpur; others from IMD, compiled by Mani [15]). They identified T, T_{max}, T_{min}, H, and SH as the most suitable predictors but noted that SH is often unavailable. An ANN model without SH achieved a MAPE of 9.04%, underscoring the need for models with lower errors.

A study was done by Panthee and Kulchandra [17] to estimate solar radiation in Kathmandu at the Tribhuvan International Airport. The airport's elevation is 1338m, in the Himalayan Mountain range's hilly and plateau region. They obtained desirable magnitude of the relative measure & accuracy in the Himalayan range. They developed five different ANN modes. Yadav and Chandel [18] found that sunshine hours are unavailable for many locations, and it is difficult to establish ground station data at higher altitudes and difficult terrain.

Dogra et al. [19] mapped solar radiation across all districts of Himachal Pradesh using ANN models with inputs such as latitude, longitude, altitude, temperature, humidity, pressure, and wind speed. Input variables were sourced from NASA, while radiation data for 23 cities were taken from the Solar Radiation Handbook, with Hamirpur data obtained from CEE, NIT Hamirpur. The study relied on a trial-

and-error method to select hidden layers and nodes, leading to higher computational cost. Results indicated maximum solar potential in Kangra and minimum in Lahaul Spiti. In contrast, Yadav and Chandel [18] reported the highest potential in Lahaul Spiti, highlighting inconsistencies between the two studies. These findings suggest the need for estimation models that avoid trial-and-error selection, reduce computational cost, and resolve such discrepancies.

Kumar & Kaur [20] performed the study to estimate solar radiation with the most influential explanatory variables for the location of Chamba (H.P.) and verify it for Hamirpur (H.P.) India. The explanatory variables for the modeling were obtained from the NASA power data access viewer. The Chamba solar radiation data were also downloaded from NASA, while the Hamirpur data were taken from the CEE, NIT Hamirpur. They found that air temperature, relative humidity, and atmospheric pressure are more influential variables, while clearness index and precipitation are less influential. They explain that their study's results are better than those performed in the past and results obtained from the empirical models. They obtained the lowest MAPE from the ANN model 5-48-15-1. It is 12.15%.

The related work to the present study is summarized and explained in Table 1. The research studies in Table 1 can further be subdivided into groups. The first group from S. N. 1 to 5 explains the satellite resources utilized worldwide to estimate solar radiation in various forms. The second group of research studies from S. N. 6 to 10 shows the research studies related to the site of the present study. The studies [13, 16, 18] explained that sunshine duration/hours data are unavailable for most Indian locations, especially in the mountain region of India.

Table 1. Summary of the related work to the study area and resources

S.N	Reference	Satellite sources	Ground station sources	Input(I) and target(O) variables	Relative Measures	Remark
1	Assi et al. [6]	SoDA, SEE, SWERA	NCMS	I = [GSR satellite]; O = [GSR ground station]	R^2	further expansion will be remarkable.
2	Olomiyesan et al. [7]	NASA	NIMET	I = [SD, Air temperature (T _{max} and T _{min}); O = [H _m]	MBE, MPE, RMSE, R and R^2	is the most influential variable

3	Masoom et al. [9]	INSAT - 3D, CAMS, SENSE	BSRN	I = [SZA, AOD, TOC, SSA, CWV, AE, COT]; O = [GHI, DNI]	correlation coefficient	Low CC (0.18 to 0.67) shows that Satellite sources are not a good alternative in cloudy conditions.
4	Manara et al. [11]	CMSAF, SSI	PEA	I = [Ground station SSI]; O = [SSI CMSAF SARA-2.1, CLARA-A2]	p-value, S.D., MAD	The study for high elevation areas. At higher OCI, more uncertainty is associated with ground station and satellite SSI.
5	Carpentieri et al. [12]	HelioMont Meteosat product, SARA-2	SMN, SAFaM	I = [Ground measured SSR (SMN, BSRN, SACRaM)]; O = [Satellite SSR (SARA-2, Helio Mont)]	RMSD, MBD, MAD, rMAD	Study shows that the accuracy is much lower in intra-hour and intra-day; apart from this, the resolution of satellite sources is higher.
6	Kumar Yadav and Chandel [13]	-	IMD	I = [lat, long, Elevation, SH, Julian Day, T _{max} , T _{min} , RH, Day length]; O = [GSR]	MAPE, MBE, RMSE	MAPE was not explained for the city with the highest elevation.
7	Yadav et al. [16]	NASA		I = [T _{avg} , T _{max} , T _{min} , altitude, SH, CI, ER, lat and long]; O = [Solar radiation]	MAPE	Avoid sunshine hour as the ANN input variable due to the absence of many locations.
8	Panthee & Kulchandra [17]	-	Meteorological dept. of Kathmandu	I = [Daily T _{avg} , T _{max} , RH, rainfall, SH]; O = [GSR]	RMSE, MPE, MBE, CC	CC among the explanatory and target variables was not covered, so how and when to switch on the ANN model is ambiguous.
9	Kumar & Kaur [20]	NASA	CEE NIT Hamirpur	I = [air temperature, RH, atmospheric pressure, CI and precipitation, lat, long]; O = [GSR]	MAPE, MSE	The lowest MAPE is 12.15%, which shows that further improvement is still necessary.
10	Dogra et al. [19]	NASA	CEE NIT Hamirpur, Solar Radiation Handbook	I = [lat, long, altitude, air temperature, humidity, pressure, WS]; O = [GSR]	MAPE, RMSE, CC, MSE	Pressure, humidity, wind speed, and temperature were taken from NASA, while GSR was not obtained from NASA.

The more relevant studies are discussed above to explain the necessity of the present study. The many other studies are covered but are not discussed in the introduction section due to page constraints. These uncovered studies act as prerequisites in the selection of more relevant studies. The detailed literature survey suggests that much work has been done in this realm. However, a few challenges still exist. These challenges are:

- Due to the absence of a sunshine hour variable for most locations, it is not preferred as an explanatory variable in developing the solar radiation estimation model.
- Soft computing techniques must be avoided if reasonable accuracy is met with hard computing to reduce computational costs. There is a need to avoid higher-order polynomials in mathematical models.
- How can satellite-based sources serve as effective alternatives in hilly regions

of India and difficult terrain for developing highly accurate solar radiation estimation models? Furthermore, how can gridded solar radiation data be utilized to develop pinpoint solar radiation estimation?

To address these challenges, the key contribution of this work is summarized as follows:

In this study, investigations are carried out for the development of a GSR database with the help of two satellite sources at Hamirpur (Himachal Pradesh, India), a hilly city in the Himalayan Mountain range. Estimation models are developed without a sunshine duration variable & try to demonstrate that these can be performed better in the absence of a sunshine duration variable. To avoid the computational cost, this work is performed with the help of mathematical/ hard computing. Comparatively, mathematical/ hard computing techniques were found to be better than the soft computing-based techniques for similar study areas. To

show the strength of the present work, the results of this work were verified for each season as well as the complete year with the help of seven metrics. Pinpoint estimation was performed with the help of gridded data and an attempt was made to show that it is valid for many other locations.

Section 2 describes the study area, datasets (ground and satellite), and evaluation metrics. Section 3 presents seasonal and annual results, followed by discussion in Section 4 and conclusions in Section 5.

2. Materials and methods

2.1 Study area

In the present study, Hamirpur (H.P.), India, is considered as the site. The reason for selecting this hilly region is explained in the Introduction section. The global solar radiation was recorded at the CEE, National Institute of Technology Hamirpur. The geographical coordinates of the location are 31.70720° N, 76.52580° E. Details about the city and its geography are provided in the report [21]. The location is more clearly depicted in Figure 1.

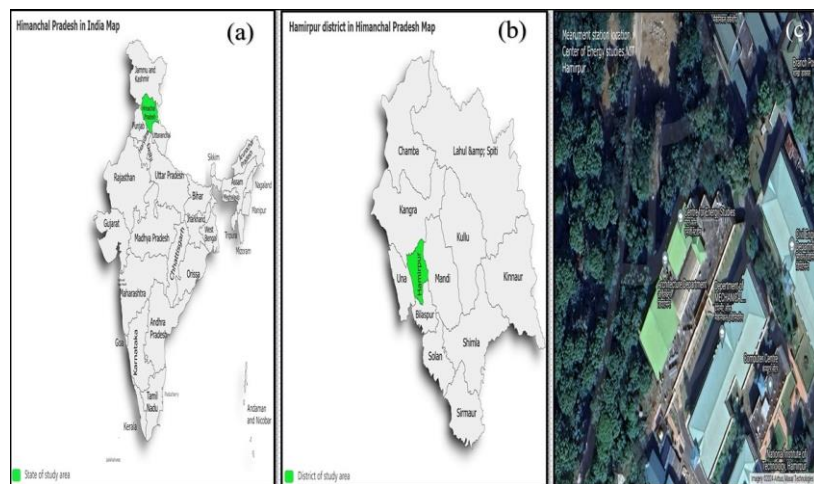


Figure 1 (a) Himachal Pradesh (b) Hamirpur District (c) Center of Energies & Environment, NIT Hamirpur

2.2 Method

The workflow diagram of the method carried out in this study is explained in Figure 2. The first step is the calculation of the Pearson correlation coefficient (PCC) between the ground and satellite data sets. There are a few

days when ground station data are missing; therefore, it is necessary to select data from specific dates to calculate PCC.

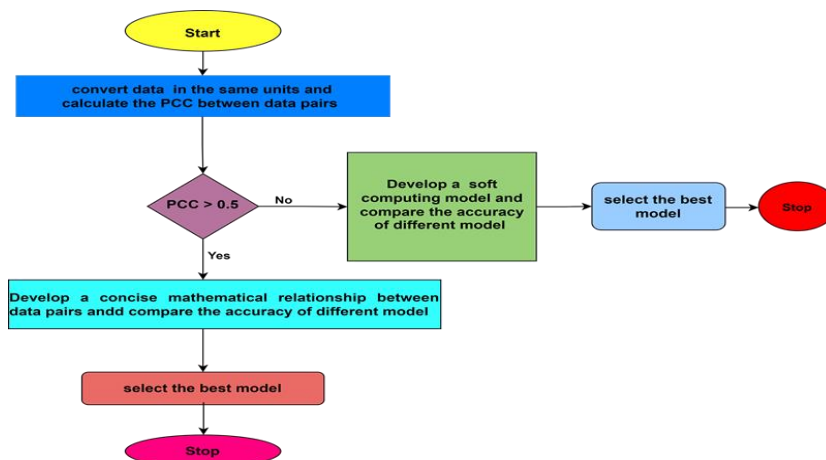


Figure 2 Flow chart of the method used for the mapping between ground and satellite radiation data

The magnitude of PCC is the critical value that decides which selection techniques (soft

computing and concise mathematical) are suitable for establishing a relationship between

independent and dependent variables. In the present study, the absolute magnitude is measured at ground level in megajoules per meter² per day (MJ/m²/day). The corresponding data is downloaded from the NASA (power data access viewer) websites and the ISRO android application (ISRO solar calculator). The downloaded and measured data were both framed in the same unit so that the arithmetic operation could be carried out. The studies suggest that the Pearson correlation coefficient was suitable for the absolute values. The Spearman rank correlation suits qualitative data like beauty, honesty intelligence, etc. The Kendall rank correlation is suitable for ranked data. The two interviewers assign the grading of

ten different candidates. Now, with the help of Kendall's rank correlation, it is possible to find an association between the two interviewers' perceptions. The special case of the Pearson correlation coefficient, "Point biserial correlation coefficients," is not applicable here. The PCC [22], coefficient of determination (R^2), Adjusted R^2 , root mean square error (RMSE), mean bias error (MBE), mean absolute percentage error (MAPE) and standard error of estimate (SEE) were found to be suitable relative measures for strengthening the results. The calculation formula for these relative measures is defined in Table 2.

Table 2. List of Relative Measures Used in this Study

Relative measure	Notation	Formula	Description
Pearson correlation coefficient	R	$R = \frac{\sum_{i=1}^n (Y_{i,g} - \bar{Y}_g)(Y_{i,s} - \bar{Y}_s)}{\sqrt{\sum_{i=1}^n (Y_{i,g} - \bar{Y}_g)^2 \sum_{i=1}^n (Y_{i,s} - \bar{Y}_s)^2}}$	$Y_{i,g}$, $Y_{i,s}$ are the i^{th} ground, satellite observation and $\bar{Y}_g$, $\bar{Y}_s$ are the means of ground and satellite observation of the radiation value. k and n are the number of independent variables and sample size respectively.
Root of mean square error	RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{i,g} - Y_{i,s})^2}$	
Mean bias error	MBE	$MBE = \frac{1}{n} \sum_{i=1}^n (Y_{i,g} - Y_{i,s})$	
Mean Absolute percentage error	MAPE	$MAPE = \left(\frac{1}{n} \sum_{i=1}^n \left \frac{Y_{i,g} - Y_{i,s}}{Y_{i,g}} \right \right) \times 100$	
coefficient of determination	R^2	$R^2 = \frac{\text{Explained variation}}{\text{Total variation}}$ $= \frac{\text{Total variation} - \text{Error variation}}{\text{Total Variation}}$	
Adjusted R^2	Adjusted R^2	$\text{Adjusted } R^2 = R^2 - \frac{k(1 - R^2)}{n - k - 1}$	The sum of square error is the error between the variables while the degree of freedom is the number of independent variables in the model
Standard error of estimate	SEE	$SEE = \sqrt{\frac{\text{sum of square error}}{\text{degree of freedom}}}$	

The mapping function used in this study is explained in Table 3. In Table 3, where X is the independent variable and Y is the dependent variable, b_0 , b_1 , and b_2 are constants, and u represents the upper boundary value for the dependent variable in the logistic model. X corresponds to satellite solar radiation data obtained from ISRO and NASA, while Y is the

recorded ground station GSR at CEE, NIT Hamirpur (H.P.), India.

The present study was conducted on all days as well as four seasons (winter, summer, spring and autumn) radiation data sets. The Box plot [23] analysis is carried out to find the outliers in the data set and compare data sets in different quartile ranges. Similarly, the Taylor diagram [24] is used in this study to show the

different model performances on various relative measures in a single diagram.

Table 3. Mathematical function between dependent variable to independent variables

Model	Mathematical function
Linear	$Y = b_0 + b_1 \times X$
Quadratic	$Y = b_0 + b_1 \times X + b_2 \times X^2$
Compound	$Y = b_0 \times b_1^X$
Growth	$Y = e^{(b_0 + b_1 \times X)}$
Logarithmic	$Y = b_0 + b_1 \times \ln(X)$
Cubic	$Y = b_0 + b_1 \times X + b_2 \times X^2 + b_3 \times X^3$
S	$Y = e^{(b_0 + \frac{b_1}{X})}$
Exponential	$Y = b_0 \times e^{b_1 \times X}$
Inverse	$Y = b_0 + \frac{b_1}{X}$
Power	$Y = b_0 \times X^{b_1}$
Logistic	$Y = \frac{1}{(\frac{1}{u} + (b_0 \times (b_1^X)))}$

2.3 Observations data sets

In this study, data from three different sources were used. The first dataset is the global solar radiation (GSR) measured at the CEE, National Institute of Technology Hamirpur (H.P.), India. A KIPP and ZONEN CM4 high-temperature pyranometer was used to continuously measure GSR data. The average solar radiation for a complete day was selected for analysis. The unit of measured GSR at NIT Hamirpur is MJ/m²/day. The second dataset was downloaded from the NASA website. The geographical coordinates used for this study are 31.70720° N and 76.52580° E. The resolution of the downloaded data is 0.50° for latitude and 0.6250° for longitude, which means the downloaded radiation data represent an area of approximately 55.5 × 69.576 km. The radiation values change beyond this area span. The NASA data used in this study is CERES SYN1deg All Sky Surface Shortwave Downward Irradiance (kWh/m²/day). The third dataset is from the Indian Space Research Organisation (ISRO). ISRO mapped the potential of solar energy to enable beneficiaries to maximize benefits from solar energy-based applications in Indian locations. ISRO data are available in the form of monthly solar insolation. In this study, ground-based solar radiation data were considered for the

period from 31 December 2015 to 31 December 2018, corresponding to the duration for which on-site measurements were available. Although the total period comprises 1,097 days, reliable ground data were recorded for only 729 days, with the remaining 368 days missing. Accordingly, satellite-based data from NASA and ISRO were selected for the same dates to ensure temporal consistency.

The NASA solar radiation data, originally expressed in kWh/m²/day, were converted to MJ/m²/day for the computation of the Pearson correlation coefficient (PCC) with the ground measurements over 729 days. The correspondence between the ground station and satellite datasets was further examined on a monthly timescale, where monthly solar radiation represents the cumulative solar energy received during each month, expressed in MJ/m²/month. Due to missing daily observations, only 12 complete months (January–December 2018) were available for reliable monthly analysis.

Various predictive models were developed using this 12-month dataset to estimate GSR. The same data pre-processing and correlation analysis procedures were consistently applied across different seasonal subsets to ensure comparability. The ground-based and satellite-derived radiation datasets, along with their descriptive statistics and detailed data

characteristics, are presented in Tables A1–A4 of the supplementary material.

3. Results

The introduction section of the paper reveals that the potential of solar energy available in the hilly region of India has yet to be utilized. This study overcame the requirement of establishing a measurement setup and its periodic calibration process for these problematic terrains. The present study explains how the measurement setup requirement can be overcome by using satellite data. The satellite data were used in this study, as described in section 2.3. The flowchart of the methodology explains the selection of suitable techniques based on the Pearson correlation coefficient between ground station data and satellite data. Gupta et al. [25] explain that if the PCC between ground station data and satellite data is above 0.5, then a concise mathematical relationship is suitable for establishing the relationship between the variables. In the

present study, the PCC is much higher than 0.5 for all the data sets explained in Table 4. Therefore, this study discards soft computing techniques and selects mathematical methods to explore the mathematical function for input-output mapping. The box plot of the measured ground station daily GSR and NASA all-sky radiation is shown in Figure 3. It clearly indicates that the magnitude of all-sky radiation is higher than the measured ground station data.

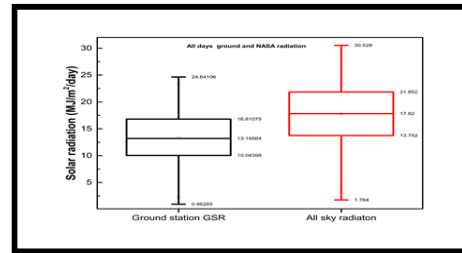


Figure 3 Box plot of the all-day ground station radiation and NASA (all-sky radiation)

Table 4. Correlation coefficient (PCC) and mean bias-error between ground station and satellite (NASA and ISRO) data

Temporal resolution	Period considered	Correlation coefficient (PCC)		Mean bias error	
		Ground - ISRO	Ground - NASA	Ground - ISRO	Ground - NASA
Daily	729 days (31/12/2015 – 31/12/2018)	N. A.	0.950	N.A.	4.4629 MJ/m ² /day
Monthly	01/01/2018 – 31/12/2018(12 months)	0.8639	0.8858	51.1008 (MJ/m ² /month)	156.1266 (MJ/m ² /month)
Seasonal	Winter (D - J - F)	0.9996	0.7284	17.8024 (MJ/m ² /month)	113.9584 (MJ/m ² /month)
	Spring (M – A - M)	0.7512	0.9519	53.2157 (MJ/m ² /month)	175.6637 (MJ/m ² /month)
	Summer (J – J - A)	-0.1845	0.9659	82.3404 (MJ/m ² /month)	146.2837 (MJ/m ² /month)
	Autumn (S – O - N)	0.8638	0.0971	51.0445 (MJ/m ² /month)	188.6005 (MJ/m ² /month)

Similarly, the box plot of the monthly ground station GSR, NASA, and ISRO radiation is shown in Figure 4.

The ISRO monthly radiation is much closer to ground station radiation than NASA radiation, as shown in Figure 4. The resolution

of satellite data (in the form of a grid) is in kilometers. Figures 3 and 4 demonstrate that no outliers exist in the data sets. The analysis was performed in different time slots to enhance the significance of the study. Table 4 presents a month-wise comparison of NASA data with

ground station data. It also illustrates the PCC and mean absolute deviation between satellite data and ground station data. The mapping between ground station data and satellite data is summarized in Tables 5 and 6. The IBM® SPSS® (Statistical Package for Social Science) was used in this study for the mathematical calculations.

The MAPE of the different models explained in Table 5 lies in the range of 7.495255 – 9.272701%, while the RMSE exists in the range of 43.013584 – 45.428498 MJ/m²/month. Similarly, the magnitude of R² ranged from 0.746 – 0.786.

The performance comparison of different mathematical models for ISRO and ground station GSR is presented in Table 5 and illustrated in Figure 5 using a Taylor diagram, while the MAPE of the models is depicted in Figure 6. In the present study, the ascending order of SEE magnitude has been used to assign the name and rank of the model, i.e., the lower SEE model is named R1 and ranked 1. The

same rank and name were assigned to the equal magnitude of SEE models. The predicted values of the similar rank models have an equal magnitude of predicted values. In ISRO mapping, the compound, Growth, Exponential and Logistic have an equal magnitude of the standard error of estimate. Hence, the rank assigned to all the models is 3, and the corresponding model name is R3. The PCC between different models predicated and ground station measured values vary in the range of 0.864 – 0.886, and the difference between the minimum and maximum values is 0.022. So, it has been observed from Figure 5 that the PCC magnitude of different models lies in the span of 0.85 – 0.9. A low span also exists for standard deviation and RMSE, as depicted in Figure 5. Based on the lowest MAPE, the Inverse model was found to be the best-fit model among all the considered models for mapping ISRO and ground station measured data.

Table 5. A different statistical model for the prediction of Ground station data (dependent) from ISRO data in (MJ/m²/month)

Model	R	R ²	Adjusted R ²	RMSE (MJ/m ² /month)	MAPE	Standard Error of Estimate	Rank
S	0.886	0.786	0.764	43.219801	8.072633	0.113	1
Power	0.879	0.772	0.749	43.881087	8.449926	0.116	2
Compound	0.866	0.749	0.724	45.428498	9.272701	0.122	3
Growth	0.866	0.749	0.724	45.428498	9.272701	0.122	3
Exponential	0.866	0.749	0.724	45.428498	9.272701	0.122	3
Logistic	0.866	0.749	0.724	45.428498	9.272701	0.122	3
Inverse	0.871	0.758	0.734	43.013584	7.495255	47.119	4
Logarithmic	0.87	0.757	0.732	43.134991	7.967581	47.252	5
Linear	0.864	0.746	0.721	44.043915	8.615501	48.248	6
Quadratic	0.868	0.753	0.699	43.433426	7.585687	50.153	7
Cubic	0.868	0.753	0.699	43.433426	7.585687	50.153	7

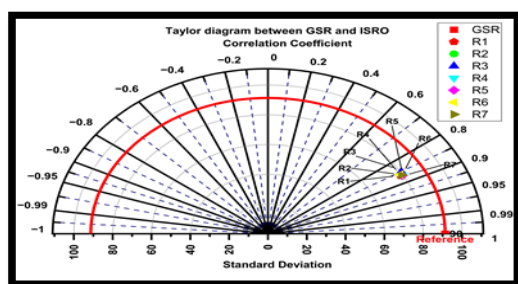


Figure 5 Taylor diagram between measured GSR and different model's predicted values from ISRO data

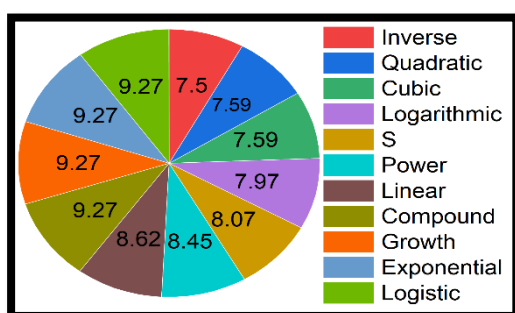


Figure 6 MAPE between measured GSR and predicted values from different models using ISRO data

The mathematical mapping between ground station measured radiation and NASA radiation is explained in Table 6. Similar rank assignment and nomenclature are used here for the different models, as in the ISRO mapping. Figures 7 and 8 compare the performance of the mappings between NASA and ground station GSR. The Taylor diagram in Figure 7 illustrates the linear association between predicted and observed GSR values, while Figure 8 presents the MAPE comparison for different models. The MAPE in the mapping between NASA and ground station radiation ranges from 7.3688% to 8.2153%, and the RMSE ranges from 39.9849 to 44.7973 MJ/m²/month, as shown in Table 6. The coefficient of determination varies from 0.709 to 0.791, and the correlation coefficients range from 0.842 to 0.889. The constant values corresponding to the models in Tables 5 and 6 are provided in Table A5 of the supplementary material. The models' performance comparison is summarized in Table 5 using relevant evaluation metrics. Statistical tests, including ANOVA and t-tests, were employed to determine whether differences in model performance are statistically significant or due to random variation. Results of the ANOVA and t-tests are

presented in Tables A6 and A7 of the supplementary material. For all models in Table 5, the observed F-values exceeded the critical F-values, and p-values were below 0.005 at the 5% significance level, indicating strong statistical significance. Most models also showed t-values higher than the critical t-values, except for the quadratic and cubic models. Although the constants of the S and linear models exhibited lower t-values, this had negligible impact on their overall performance. The inverse model, which achieved the lowest MAPE, was statistically significant in both tests. P-values are reported to three decimal places; values below 0.001 are denoted as 0.000. Similar statistical analysis results for the models in Table 6 are provided in Tables A8 and A9 of the supplementary material, confirming that all models have significant F-values and p-values below 0.005. Only the power and linear models showed lower t-values for their constants, but this did not notably affect performance. The inadequacy of the t-test for higher-order polynomial models (quadratic and cubic), as shown in Tables 5 and 6, suggests that higher-order polynomial models should be avoided when lower-order models can provide sufficient accuracy.

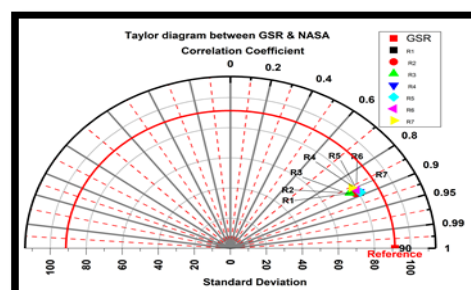


Figure 7 Taylor diagram between measured GSR and different model's predicted values from NASA data

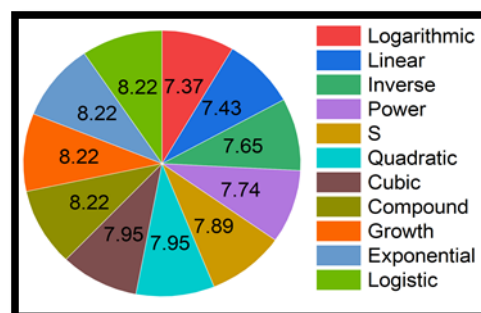


Figure 8 MAPE between measured GSR and different model's predicted values using NASA data

Table 6. A different statistical model for the prediction of Ground station data (dependent) from NASA data in (MJ/m²/month)

Model	R	R ²	Adjusted R ²	RMSE (MJ/m ² /month)	MAPE	Standard Error of Estimate	Rank
Compound	0.858	0.737	0.711	40.0934	8.2153	0.125	1
Growth	0.858	0.737	0.711	40.0934	8.2153	0.125	1
Exponential	0.858	0.737	0.711	40.0934	8.2153	0.125	1
Logistic	0.858	0.737	0.711	40.0934	8.2153	0.125	1
Power	0.853	0.728	0.701	40.7319	7.7406	0.127	2
S	0.842	0.709	0.68	42.6456	7.886	0.131	3
Linear	0.886	0.785	0.763	40.5786	7.4334	44.452	4
Quadratic	0.889	0.791	0.744	39.9849	7.9531	46.171	5
Cubic	0.889	0.791	0.744	39.9849	7.9531	46.171	5
Logarithmic	0.875	0.766	0.743	42.2691	7.3688	46.303	6
Inverse	0.859	0.738	0.711	44.7973	7.6528	49.073	7

4. Discussion

1. Ground-based measurements using the Kipp and Zonen CM4 pyranometer offer high-fidelity, site-specific data (MJ/m²/day) that inherently capture local cloud dynamics and surface albedo, making them ideal for validation and micro-scale solar resource assessment. In contrast, NASA CERES provides spatially extensive daily averages (kWh/m²/day) at approximately 55–70 km resolution through sophisticated radiative transfer modeling, effectively incorporating cloud cover and albedo but potentially smoothing local variability. The ISRO solar insolation data, derived from geostationary satellites and regional modeling, provide broader spatial coverage suited for regional-scale applications but with reduced temporal detail. These differences highlight the complementary nature of the datasets—ground data are best for site-specific validation. The estimation of global solar radiation (GSR) at a

given location, when derived through mapping between ground-based and satellite radiation sources, demonstrates improved accuracy compared to relying solely on satellite-derived radiation data, particularly when ground station measurements are geographically distant from the target site. The comparison among the solar radiation data sources is explained in Table A10 of the supplementary material.

2. From Table 4, in all the cases, the PCC is higher than 0.5. This shows a strong correlation between ground station and satellite radiation. Apart from this, the magnitude in all the cases is more significant than 0.5, which means the mathematical relationship is suitable for the mapping. The performance of different prediction models of both satellite sources is further summarized and explained in Table 7.

Table 7. Comparison of relative measures of different satellite sources mapping

Satellite Sources	MAPE			RMSE (MJ/m ² /month)			R ²			R		
	Max	Min	Diff	Max	Min	Diff	Max	Min	Diff	Max	Min	Diff
ISRO	9.2727	7.4952	1.7775	45.4285	43.0136	2.4149	0.786	0.746	0.04	0.886	0.864	0.022
NASA	8.2153	7.3688	0.8465	44.7973	39.9849	4.8124	0.791	0.709	0.082	0.889	0.842	0.047

The lowest MAPE (7.3688) and RMSE (39.9849) were obtained from the NASA model, as shown in Table 7. Similarly, higher R² (0.791) and R (0.889) were achieved with

the NASA model compared to the ISRO model. This demonstrates that NASA data are better suited for mathematical mapping and estimating ground radiation than ISRO data. If

the researcher aims to avoid establishing a mathematical mapping, selecting ISRO data is a more appropriate option for developing the ground radiation database because the ground station value is much closer to the ISRO value, as shown in Figure 4 with the aid of a box plot. The box plot analysis is further explained in Table 8. Diff1 represents the difference between the box plot values of ISRO radiation and ground station-measured radiation.

Similarly, Diff2 is the difference between the box plot value of NASA and ground station radiation. The min, max, Q1, Q2, and Q3 denote the minimum, maximum, Quartile 1, Quartile 2, and Quartile 3 of the data sets. It can be observed that Diff1 values are smaller than Diff2, indicating that ISRO radiation is closer to ground station radiation than NASA radiation.

Table 8. Comparison of Box plot of three different data sources of radiation

Box plot values	GS	ISRO	NASA	Diff1	Diff2	(Diff2/Diff1)
Min	275.2388	302.4	410.4	27.1612	135.1612	4.97626
Max	537.5242	622.8	746.568	85.2758	209.0438	2.451385
Q1	328.877	342	475.11	13.123	146.233	11.14326
Q2	400.8673	468	555.912	67.1327	155.0447	2.309526
Q3	485.0142	531	650.718	45.9858	165.7038	3.603369

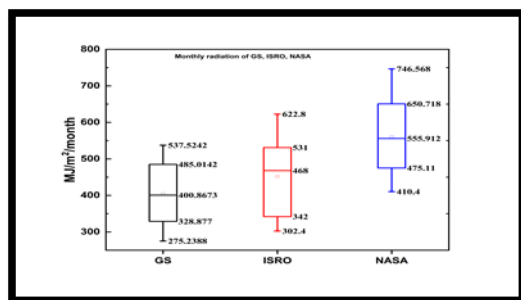


Figure 4 Box plot of the Monthly radiation of the ground station, ISRO and NASA (all-sky radiation)

The ratio of Diff1 and Diff2 ranged between 2.309526 to 11.14326. This shows that the ISRO radiation is much closer to the ground station radiation than NASA radiation.

3. The necessity of calculating the standard error of estimate is to demonstrate the precision of the predicted values from different models. A lower standard error of estimate indicates higher precision, and vice versa. Tables 5 and 6 show that models such as S, power, compound, growth, exponential, and logistic exhibit higher precision, while others demonstrate lower precision. The ratio (Rank1 / Rank7) of the magnitude of the standard error of estimate for ISRO is 443.831, whereas for NASA it is 392.584.

4. The inverse model obtained the best-fit model for mapping the ground station and ISRO radiation data among all the considered models. The MAPE and RMSE of the inverse model are 7.495255 % and 43.013584 MJ/m²/month, respectively, as explained in Table 5. A Similarity is not valid for the mapping between ground station and NASA radiation. The logarithmic model has the lowest MAPE, while the quadratic and cubic models pass the lowest RMSE. In this case, MAPE is considered to find the best-fit models while discarding the higher-order polynomial models. So, this study considers the logarithmic model as a best-fit model because it has the lowest MAPE, as explained in Table 6, in the case of NASA radiation.

5. The highest values of R² in the case of NASA and ISRO are 0.791 and 0.781 respectively. This shows that approximately 79% of the variation in ground station radiation is governed by the satellite source radiation, while approximately 21% of the variations are governed by the other variables. Incorporating other terrestrial variables, such as anthropogenic variables, further reduces the MAPE. Suthar et al. [26], Fan et al. [27] and Gupta et al. [25] explained the reduction in the MAPE due to incorporating air pollution variables with other explanatory variables.

6. The performance of this novel approach is compared with the other studies that were accomplished in the exact location or nearby location. Kumar and Kaur[28] developed the ANN model for estimating GSR at Hamirpur (H.P.), India and obtained 16.45% MAPE with the 5-16-1 multi-layer perceptron architecture. However, a study was performed by Kumar and aur[20] for the location of Chamba (a nearby district of the study area of the present study), and they obtained a maximum accuracy of 12.15%. The lowest MAPE obtained in the

present study is 7.3688%. This shows that this novel approach is more valuable than the techniques used in the other studies executed for the same or nearby location. The further performance comparison of various studies is explained in Table 9. In Table 9, the present study is denoted as PS, ANN (5-48-15-1) as ANN 1, and ANN (5-16-1) as ANN 2, while ANN-5, adopted from reference [16], is also included. The same abbreviations are used in Figure 9.

Table 9 Performance comparison of different studies with the present study

Study	location	Model	MAPE	Independent variables
Kumar and Kaur [20]	Chamba	Empirical	21.95	Relative Humidity (RH)
		Empirical	18.39	Temperature average (T_{avg})
		Empirical	21.62	Pressure (P)
		ANN 1 (5-48-15-1)	12.15	T, RH, P, precipitation (PT) and clearness index (CI)
Kumar and Kaur [28]	Hamirpur	ANN 2 (5-16-1)	16.45	T, RH, P, PT and CI
Yadav et al. [16]	All districts of H.P.	ANN (ANN-5)	9.04	T, T_{min} , T_{max} and altitude(H)
Present study (PS)	Hamirpur	Logarithmic, Inverse	7.3688, 7.4952	Satellite solar radiation

However, the maximum accuracy Yadav et al. [16] obtained is 6.89% with the ANN-4 model, i.e., including the sunshine hours as explanatory variables. They did not use the ANN-4 model to depict the solar radiation in all the districts of H.P. because the sunshine hours are unavailable for most Indian locations. That's why the ANN-4 model is not incorporated in Table 9 for the performance comparison from the present study. The performance comparison of present study with the previous related studies are shown in Figure 9.

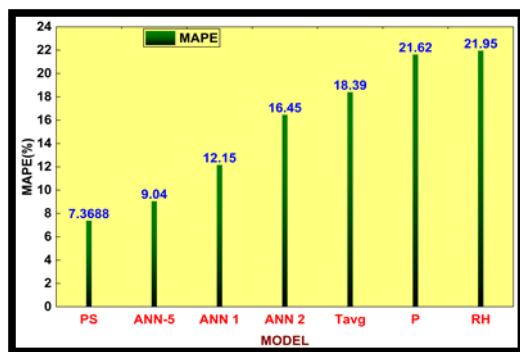


Figure 9 Performance comparison of the present study with previous studies

5. Conclusion

In the present study, the proposed model demonstrates higher accuracy compared to single-variable empirical models based on RH, Tavq, and pressure (P), as well as multivariable ANN models, including ANN 1 (5-48-15-1), ANN 2 (5-16-1) with inputs $I = [T, RH, P, PT, CI]$, and ANN-5 with inputs $I = [T, T_{min}, T_{max}, \text{altitude (H)}]$, developed for estimating GSR in Hamirpur (H.P.), India.

The lowest MAPE, 7.3688%, was obtained in the present study, which is the lowest among all the work conducted for this location. The model developed in this study employed concise mathematical techniques, whereas others utilized mathematical and soft computing methods. If hard computing techniques yield better results than soft computing, the preference for hard computing is justified because soft computing techniques increase computational costs. This study further confirms that mathematical mapping is superior to soft computing when the critical value of PCC exceeds 0.5 between mapped variables. Additionally, the study reveals that ISRO data closely aligns with ground station data, while NASA radiation data is more suitable for mathematical mapping. Lewis [29] explains

that a model with a $MAPE \leq 10\%$ indicates higher prediction accuracy. All models discussed in this study have a MAPE below 10%. The estimated solar radiation remains consistent throughout the grid. Although the proposed model exhibits higher accuracy than both single-variable empirical and multivariable ANN-based models, certain limitations persist. The validation is limited to the Hamirpur (H.P.) region, which may restrict its applicability to areas with different climatic or topographical conditions. Furthermore, the influence of local terrestrial and anthropogenic factors—such as aerosol concentration, land-use variation, and pollution levels—has not been incorporated, which could slightly affect prediction accuracy. Future research should focus on integrating these influential terrestrial variables to further reduce MAPE and enhance model robustness. Developing a hybrid framework that combines the interpretability of mathematical models with the adaptability of soft computing may also improve accuracy across diverse climatic zones. Moreover, expanding the model to a spatially distributed scale using high-resolution satellite datasets (e.g., Solargis, ISRO, and NASA) will facilitate precise, site-specific GSR estimation for various terrains and environmental conditions.

References

- [1] D. Yang and J. M. Bright, "Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years," *Sol. Energy*, vol. 210, no. November 2019, pp. 3–19, 2020, doi: 10.1016/j.solener.2020.04.016.
- [2] M. Sengupta et al., "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications," 2015. [Online]. Available: <https://www.nrel.gov/docs/fy15osti/63112.pdf>.
- [3] R. Perez, R. Seals, and A. Zelenka, "Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data," *Sol. Energy*, vol. 60, no. 2, pp. 89–96, 1997, doi: 10.1016/S0038-092X(96)00162-4.
- [4] N. Geuder, "Solar Resource Measurements and Satellite Data," 2013.
- [5] I. Karakoti, A. Datta, and S. Singh, "Validation of satellite based solar radiation data with ground measurements," in *Proceedings of the 14th annual international conference and exhibition on geospatial information technology and application*, 2012, pp. 1–10.
- [6] A. H. Assi, M. H. Al-shamisi, and H. A. N. Hejase, "solar radiation in UAE – a comparison between ground station," 2012.
- [7] B. M. Olomiyesan and O. D. Oyedum, "Comparative Study of Ground Measured, Satellite-Derived, and Estimated Global Solar Radiation Data in Nigeria," *J. Sol. Energy*, vol. 2016, pp. 1–7, 2016, doi: 10.1155/2016/8197389.
- [8] D. Palmer, E. Koubli, I. Cole, T. Betts, and R. Gottschalg, "Satellite or ground-based measurements for production of site specific hourly irradiance data: Which is most accurate and where?," *Sol. Energy*, vol. 165, no. February, pp. 240–255, 2018, doi: 10.1016/j.solener.2018.03.029.
- [9] A. Masoom, P. Kosmopoulos, A. Bansal, and S. Kazadzis, "Solar energy estimations in india using remote sensing technologies and validation with sun photometers in urban areas," *Remote Sens.*, vol. 12, no. 2, pp. 1–25, 2020, doi: 10.3390/rs12020254.
- [10] T. Mishra, A. Rabha, U. Kumar, K. Arunachalam, and V. Sridhar, "Assessment of solar power potential in a hill state of India using remote sensing and Geographic Information System," *Remote Sens. Appl. Soc. Environ.*, vol. 19, no. March 2020, p. 100370, 2020, doi: 10.1016/j.rsase.2020.100370.
- [11] V. Manara et al., "Comparison of surface solar irradiance from ground observations and satellite data (1990–2016) over a complex orography region (piedmont—Northwest Italy)," *Remote Sens.*, vol. 12, no. 23, pp. 1–26, 2020, doi: 10.3390/rs12233882.
- [12] A. Carpentieri, D. Folini, M. Wild, L. Vuilleumier, and A. Meyer, "Satellite-derived solar radiation for intra-hour and intra-day applications: Biases and uncertainties by season and altitude," *Sol. Energy*, vol. 255, no. March, pp. 274–284, 2023, doi: 10.1016/j.solener.2023.03.027.
- [13] A. KumarYadav and S. S. Chandel, "Artificial Neural Network based Prediction of Solar Radiation for Indian Stations," *Int. J. Comput. Appl.*, vol. 50, no. 9, pp. 1–4, 2012, doi: 10.5120/7796-0907.
- [14] "Solar Insolation Data," [Online]. Available: <http://www.indiaenvironmentportal.org.in/files/srd-sec.pdf>.
- [15] A. Mani, *Handbook of solar radiation data for India*, 1980. 1981.
- [16] A. K. Yadav, H. Malik, and S. S. Chandel, "Application of rapid miner in ANN based prediction of solar radiation for assessment of solar

energy resource potential of 76 sites in Northwestern India,” *Renew. Sustain. Energy Rev.*, vol. 52, pp. 1093–1106, 2015, doi: 10.1016/j.rser.2015.07.156.

[17] J. Panthee and A. K. Kulchandra, “Estimation of global solar radiation using Artificial Neural Network in Kathmandu, Nepal,” *Int. J. Eng. Res. Sci.*, vol. 2, no. 1, pp. 62–68, 2016.

[18] A. K. Yadav and S. S. Chandel, “Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model,” *Renew. Energy*, vol. 75, pp. 675–693, 2015, doi: 10.1016/j.renene.2014.10.046.

[19] R. Dogra, S. Kumar, and N. Gupta, “Application of Artificial Neural Network to Solar Potential Estimation in Hilly Region of India,” *J. Renew. Energy Environ.*, vol. 9, no. 3, pp. 10–16, 2022, doi: 10.30501/jree.2022.307064.1267.

[20] S. Kumar and T. Kaur, “Efficient solar radiation estimation using cohesive artificial neural network technique with optimal synaptic weights,” *Proc. Inst. Mech. Eng. Part A J. Power Energy*, vol. 234, no. 6, pp. 862–873, 2020, doi: 10.1177/0957650919878318.

[21] “District environment plan district hamirpur(HP),” Hamirpur (HP), India, 2021. [Online]. Available: <https://hphamirpur.nic.in/document/district-environment-plan-hamirpur/>.

[22] N. Malhotra and S. Dash, *Marketing Research An applied orientation*, Seventh. Chennai 600113, Tamil Nadu, India: Pearson India Education Services Pvt. Ltd., 2016.

[23] D. F. Williamson, R. A. Parker, and J. S. Kendrick, “The box plot: A simple visual method to interpret data,” *Ann. Intern. Med.*, vol. 110, no. 11, pp. 916–921, 1989, doi: 10.7326/0003-4819-110-11-916.

[24] K. E. Taylor, “Summarizing multiple aspects of model performance in a single diagram,” *J. Geophys. Res.*, vol. 106, no. D7, pp. 7183–7192, 2001.

[25] S. Gupta et al., “Estimation of Solar Radiation with Consideration of Terrestrial Losses at a Selected Location—A Review,” *Sustain.*, vol. 15, no. 13, pp. 1–29, 2023, doi: 10.3390/su15139962.

[26] M. Suthar, G. K. Singh, and R. P. Saini, “Effects of air pollution for estimating global solar radiation in India,” *Int. J. Sustain. Energy*, vol. 36, no. 1, pp. 20–27, 2017, doi: 10.1080/14786451.2014.979348.

[27] J. Fan et al., “Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature,” *Renew. Sustain. Energy Rev.*, vol. 94, no. June, pp. 732–747, 2018, doi: 10.1016/j.rser.2018.06.029.

[28] S. Kumar and T. Kaur, “Development of ANN Based Model for Solar Potential Assessment Using Various Meteorological Parameters,” *Energy Procedia*, vol. 90, no. December 2015, pp. 587–592, 2016, doi: 10.1016/j.egypro.2016.11.227.

[29] C. D. Lewis, *Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting*. London: Butterworth Scientific, 1982.

Appendix A. Supplementary material