Journal of Autonomous Intelligence (2024) Volume 7 Issue 2
doi: 10.32629/jai.v7i2.1032

ORIGINAL RESEARCH ARTICLE

Models for estimation of solar irradiance in Zimbabwe: A statistical and machine learning approach

Kudzanayi Chiteka[1], Rejoice Mwarazi[2], Rajesh Arora[3], Christopher Enweremadu[4,*]

[1] *Faculty of Engineering and the Environment, Gwanda State University, P.O Box 30 Filabusi, Zimbabwe*

[2] *Faculty of Commerce, Innovation and Technology, Catholic University of Zimbabwe, 18443 Harare, Zimbabwe*

[3] *Department of Mechanical Engineering, Amity University, Haryana 122413, India*

[4] *Department of Mechanical Engineering, University of South Africa, Science Campus, Florida 1709, South Africa*

*** Corresponding author:** Christopher Enweremadu, enwercc@unisa.ac.za

ABSTRACT

The present study focused on the statistical development of solar irradiance predictive models for locations with limited solar irradiance measuring equipment. Multiple linear regression models were developed using both measured and satellite corrected meteorological data. The study chose easy to measure and access meteorological data for analysis and modelling. Multicollinearity and correlation analysis were performed to analyse the relationships among the independent and dependent variables. Statistical predictive models were developed, and the prediction accuracy of the developed models was analysed using the coefficient of determination (R^2) and the Mean Absolute Percentage Error (MAPE). The results revealed a higher performance of the developed models compared to generic empirical models. The prediction MAPE for the three models developed were respectively 0.117 kWh/m^2, 0.132 kWh/m^2 and 0.044 kWh/m^2 for H_g, H_b and H_d. The models also had R^2 values of 0.895, 0.972 and 0.993 respectively for global horizontal irradiance (H_g), direct normal irradiance (H_b) and diffuse irradiance (H_d). The developed models outperformed the generic models by a minimum of 5.74%. The study showed that it is more accurate to predict Global Horizontal Irradiance by summing the predicted component of H_b and H_d.

Keywords: solar irradiance; irradiance prediction; predictive modelling; meteorological parameters; regression analysis

ARTICLE INFO

Received: 14 August 2023
Accepted: 15 September 2023
Available online: 29 November 2023

1. Introduction

The economic and industrial growth happening in the world over has resulted in increased energy demand which in turn has increased the use of fossil-based energy systems thus destroying the environment and causing global warming. As a mitigating approach, there is a constant global shift from fossil energy systems in favour of renewable energy systems. Solar energy is one of these important clean energy sources that can alleviate the negative impacts of global warming[1]. New technologies in utilization of renewable energy particularly solar energy, as well as mitigation and adaptation are desirable in reducing pollution levels and mitigating global warming[2,3].

Accurate solar irradiance data is especially important in the design, development, and utilization of solar energy technologies[4]. Therefore there is a need for accurate data on solar radiation although it is not readily available due to the absence of measuring equipment especially in developing countries[5]. This is due to the high cost of maintenance and calibration required for irradiance measuring instruments[6].

Measured solar irradiance data is mostly important but, in its absence, models to obtain such data are useful in the deployment of solar energy systems[7].

Models have been developed to estimate solar irradiance and these include empirical models and artificial intelligence based models[8]. Liu and Jordan[9] are among the pioneers in irradiance predictions. They studied the correlation between the monthly diffuse irradiance and global horizontal irradiance and established a 3rd order correlation based on clearness index. Similarly, Collares-Pereira and Rabl[10] developed a predictive 4th order polynomial using data from five locations in the USA and using the same approach, many other correlations were developed.

For example, Mecibah et al.[11] developed 11 models to estimate the mean monthly global horizontal irradiance using sunshine records and air temperature for six cities in Algeria. Their results revealed that sunshine based models had better performance compared to other models. Pai et al.[12] formulated a statistical model based on the Melo-Escobedo-Oliveira shadowing correction to estimate the anisotropic diffuse fraction from the clearness index. Their study reported an improved model with accuracy improvement from prediction error of 6% to 1.5%.

Feng et al.[13] developed four artificial intelligence-based models to estimate diffuse irradiance at two meteorological stations. The input parameters used were daily global horizontal irradiance, sunshine duration, daily extraterrestrial irradiance and the maximum possible sunshine duration. The performance of these models was found to be better than the empirical model by Iqbal[14]. Roderick[15] developed correlations between the diffuse fraction and clearness index using the daily global horizontal irradiance and the diffuse irradiance. A model was developed to estimate the monthly average daily diffuse radiation and it was found to have reasonable accuracy with an R^2 of 0.86.

Zimbabwe is rich in solar energy and has high potential for the establishment of solar farms which could possibly alleviate the acute energy deficit in the country. Despite having the solar resource, the measured solar irradiance data itself is not readily available and hence there is need for solar irradiance estimation models. For example, Hove et al.[16] expressed the need to estimate solar irradiance in Zimbabwe to make informed decisions in the deployment of solar energy systems. The present study seeks to establish multiple linear regression empirical models to estimate solar irradiance components i.e., the global horizontal irradiance, direct normal irradiance, and diffuse irradiance. This is based on the premise that solar irradiance data is not easily available despite being an important factor in solar system design[17,18]. To improve the accuracy of the prediction models, satellite data is correlated with measured data to establish a relationship between measured and satellite data. This is done to approximate measured data with satellite data using a correction factor and hence improve the accuracy of the developed models. Hove et al.[16] noticed the inaccuracies inherent in satellite measured data and hence the need for more accurate models. In the present study, models were developed specifically for Zimbabwe because generic models are less accurate in their prediction. Furthermore, empirical models are mostly location dependent and hence there is need to develop models specifically meant for each location. It is therefore important to explore novel methods in estimating solar irradiance[19]. Performance evaluation and estimation requires detailed information and data with regards solar radiation and its individual components. In this study, analysis of predictor variables was done, and special attention was made to use parameters which can be easily measured with simple instruments and those which can be obtained analytically.

2. Materials and methods

2.1. Location

Zimbabwe is a southern African country with abundant solar radiation with an annual global horizontal irradiance of 1857 kWh/m^2 to 2257 kWh/m^2 and an average of 6.7 to 8.9 sun hours per day[20].

Data were obtained from 11 different locations in Zimbabwe from each grid point as shown in **Figure 1**. The 11 different locations were meant to provide a full coverage of aggregated data for the whole country. Zimbabwe is a landlocked country with a tropical savanna climate characterized by two major seasons i.e., summer and winter.

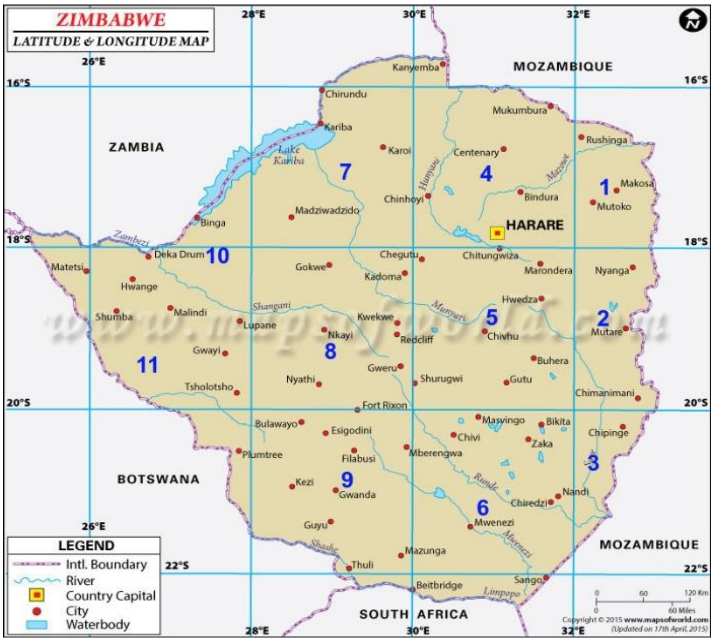


Figure 1. Map of Zimbabwe with some selected locations[21]

The winter season is mostly dry and in some parts of the winter it is often characterized by heavy clouds with sporadic light showers across the country. Generally, the winter season spans from the month of May to the month of September while summer begins in October to April. The average precipitation accumulates to 25 mm during winter season while it can accumulate to 800 mm in the summer season. The average temperatures for the winter and summer season are respectively 17 °C and 24 °C.

2.2. Meteorological data

Meteorological data of 10 different parameters were obtained on NASA Solar Energy and Surface Meteorology (https://power.larc.nasa.gov/data-access-viewer/, accessed on 9 November 2022). Typical Meteorological Year Data were obtained from 2011 to 2021 for the following parameters, Pressure, Temperature, Relative Humidity (RH), Declination (δ), Solar Noon Time, Clearness Index (K_T), Percentage Cloud Cover, Clear Sky Clearness Index, Daily Average Precipitation, Day Length, Sunset Hour Angle (ω), Cumulative Monthly Precipitation, Direct Normal Irradiance (DNI or H_b) and Diffuse Irradiance (DI or H_d). Ground measurements were also collected for a year from 6 locations evenly spread across the country using a Davis Vantage pro 2 plus weather station.

2.3. Data treatment and pre-processing

Python v3.0 was used in the data cleaning process. Incomplete records were removed from the data set and replaced with an average data value for that parameter. The data in which the Global Horizontal Irradiance (GHI or H_g) or direct normal irradiance exceeded the extraterrestrial radiation were also removed and replaced with average values. Data sets where K_T and the diffuse fraction were greater than 1 were also removed. The collected data was divided into training and validation data sets in a ratio of 75:25 respectively for training and validation.

2.4. Dimensionality reduction and parameter correlation analysis

Dimensionality reduction was performed to reduce the number of variables to be used in the modelling process. This approach was used to retain the most important parameters in the predictive modelling process and eliminate variables that are duplicate, misleading or inappropriate[22,23]. The performance of developed models often increases with a reduction in the number of predictor variables[24]. Boruta algorithm was implemented for feature ranking and selection and uses feature importance scoring which is based on Random Forests algorithm. This algorithm has the advantage of selecting features that are statistically significant. Boruta algorithm is different from other feature selection techniques in that rather than ordering, it has a sharp classification of features and it also retains all relevant variables for the decision[25,26]. This algorithm uses shadows which are basically copies of original features with randomly added values intended to retain their distribution while their importance is removed. A complete description and implementation of this algorithm is outlined in the references[27,28]. Boruta algorithm, however, cannot remove multicollinearity as such parameter correlation was performed.

Parameter correlation and curve fitting was performed in this study for all the 10 parameters using Veusz v3.4 software (https://veusz.github.io/)[29]. Relationships were established between each of the solar radiation components and the different meteorological parameters. Curve fitting was done, and performance metrics were recorded to ascertain the relationship between the parameters being investigated and solar irradiance. The models evaluated in the curve fitting process were linear, quadratic, cubic, 4th order, hyperbolic and logarithmic. All these were evaluated for each parameter relation to H_g, H_d and H_b. The model/function that gave the highest value or R^2 was selected as the representative model relating the variable to the solar irradiance component. All the 10 parameters used in this study were evaluated for their contribution in the prediction of solar irradiance. Multicollinearity, a common factor in empirical analysis was also investigated among the parameters to minimise the possibility of increasing the standard error and other complexities emanating from multicollinearity[30–32]. This was intended to identify those parameters which are also correlated in a view of eliminating redundancy and unnecessary complexity in the developed model. All parameters with a correlation coefficient greater than 0.8 had one of them eliminated. The criteria for elimination were based on retaining the parameter whose values are easily obtainable or easily measurable.

2.5. Modelling

A relationship was established between the ground measured data and satellite data to determine a correction factor used to estimate ground measured parameters from satellite data. Ground measured data from the six selected locations across the country were averaged for all the measured parameters and similarly, the satellite data were also averaged for the 11 locations. A correlation was developed between these two data sets to develop correction factors for each parameter. The correction factors were used to correct the satellite data to match ground measured data. This was done to improve the prediction accuracy of the model in the estimation of ground measured data.

Multiple linear regression models were developed to predict Global Horizontal Irradiance, Direct Normal Irradiance, and Diffuse Irradiance. The models were of the form shown by Equation 1 where y is the dependent variable, x_1, x_2, …, x_n are the independent predictor variables while β_0, β_1, β_2, …, β_n are constants and ε is the prediction error. The different measured parameters were used in the modelling process. Information from parameter correlation was collected and used in modelling. As such Vieusz v3.4 software was used in the modelling and data fitting process.

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_n\beta_n + \varepsilon \quad (1)$$

The variables selected through dimensionality reduction and parameter correlation analysis were used in the modelling process. The predicted values of solar irradiance were compared to the measured values to check the performance of the predictive model. The performance of the models were also compared to other empirical models including the model by Page[33] (Equations 2–4) and Liu and Jordan[9] (Equation 5) where $\overline{H_g}$, $\overline{H_d}$, $\overline{H_o}$, $\overline{S}$, $\overline{S}_{max}$ are respectively the monthly average daily values for global horizontal irradiance, diffuse irradiance, extraterrestrial irradiance, sunshine hours and maximum possible sunshine hours. I_{sc} is the solar constant equal to 1367 W/m^2 while a and b are constants normally obtained by regression analysis. ω_s is the sunset hour angle, ϕ is the latitude of the location while δ is the declination angle and is given by Equation 4.

$$\frac{\overline{H_g}}{\overline{H_o}} = a + b\left(\frac{\overline{S}}{\overline{S}_{max}}\right) \tag{2}$$

$$H_o = \frac{12}{\pi} I_{sc}\left(1 + 0.033\cos\frac{360n}{365}\right)\int_{-\omega_s}^{+\omega_s}(\sin\phi\sin\delta + \cos\phi\cos\delta\cos\omega)\,d\omega \tag{3}$$

$$\delta = 23.45\sin\left(\frac{360n}{365}(284 + n)\right) \tag{4}$$

$$\frac{\overline{H_d}}{\overline{H_g}} = 1.390 - 4.027\left[\frac{\overline{H_d}}{\overline{H_g}}\right] + 5.531\left[\frac{\overline{H_d}}{\overline{H_g}}\right]^2 - 3.108\left[\frac{\overline{H_d}}{\overline{H_g}}\right]^3 \tag{5}$$

3. Results

3.1. Dimensionality reduction

Dimensionality reduction performed using Boruta algorithm resulted in the selection of parameters outlined in **Table 1**. The algorithm chose parameters with an apparently low R^2 for inclusion into the model. This is the strength of the algorithm because such parameters would have been left out if simple regression was used to select parameters. Using Boruta algorithm, only relative humidity was selected for the prediction of global horizontal irradiance (H_g). The selected parameters for direct normal irradiance (H_b) were % cloud cover, cumulative precipitation, clearness index, relative humidity, sunset hour angle, declination and temperature. Likewise, the selected parameters for diffuse irradiance (H_d) were % cloud cover, cumulative precipitation, day length, declination, clearness index, relative humidity, pressure, sunset hour angle and temperature. Details of selected parameters are shown in **Table 1** where yes represent selected parameter and no represent non-selected parameter.

Table 1. Selected parameters.

		Dependent variable		
	Parameter	**H_g**	**H_b**	**H_d**
1	% Cloud cover	No	Yes	Yes
2	Cumulative precipitation	No	Yes	Yes
3	Day length	No	No	Yes
4	Declination	No	Yes	Yes
5	Clearness index	No	Yes	Yes
6	Relative humidity	Yes	Yes	No
7	Pressure	No	No	Yes
8	Sunset hour angle	No	Yes	Yes
9	Temperature	No	Yes	Yes
10	Solar noon time	No	No	No

Although the Boruta algorithm was able to select all relevant parameters, it selects all the parameters including parameters correlated to other features. As such, parameter correlation was performed to identify correlations among parameters themselves.

3.2. Parameter correlation

A parameter correlation matrix was developed and analysed for multicollinearity and their correlation with the irradiance components i.e., H_d, H_b and H_g. The results are shown in **Figure 2** and **Table 2** where DNI, GHI, DI, ω_s and K_t are respectively Direct Normal Irradiance, Global Horizontal Irradiance, Diffuse Irradiance, Sunset Hour Angle and Clearness Index. There was a good correlation between H_g and relative humidity with a coefficient of determination (R^2) of 0.78. However, the relation between H_g and all the other parameters was weak and the R^2 values were less than 50%. All the ten parameters analysed had R^2 values above 50% for their relationship with H_d. The notable parameters with a strong relation with H_d were % cloud cover, cumulative precipitation, day length, declination, K_T, pressure, temperature and sunset hour angle. H_b had a strong relationship with % cloud cover, cumulative precipitation, K_T, day length and relative humidity. The strong relationship between H_g, H_b, and H_d with % cloud cover, cumulative precipitation and clearness index is shown in **Figures 2–5**. On the other hand, the relationship between solar noon time and, H_g, H_b, and H_d was poor as shown in **Figure 6**.

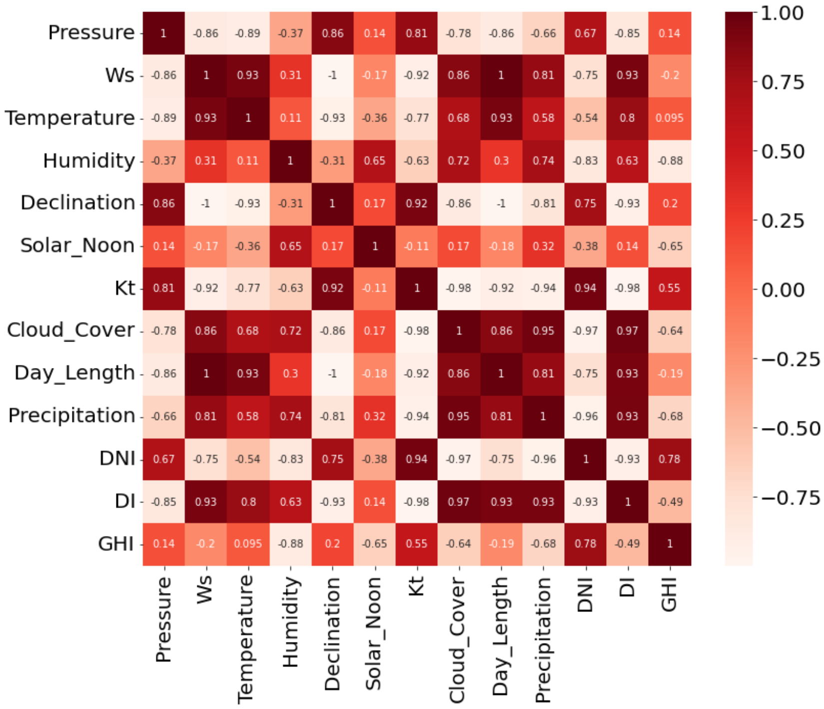


Figure 2. Parameter correlation matrix.

Table 2. Coefficients of determination in parameter correlation.

		Correlation (R^2)		
	Parameter	H_g	H_b	H_d
1	% Cloud cover	0.41	0.94	0.94
2	Cumulative precipitation	0.46	0.96	0.92
3	Day length	0.25	0.86	0.78
4	Declination	0.25	0.86	0.61
5	Clearness index	0.31	0.97	0.89
6	Relative humidity	0.78	0.59	0.74
7	Pressure	0.14	0.72	0.45
8	Sunset hour angle	0.25	0.86	0.56
9	Temperature	0.28	0.65	0.29
10	Solar noon time	0.41	0.51	0.39

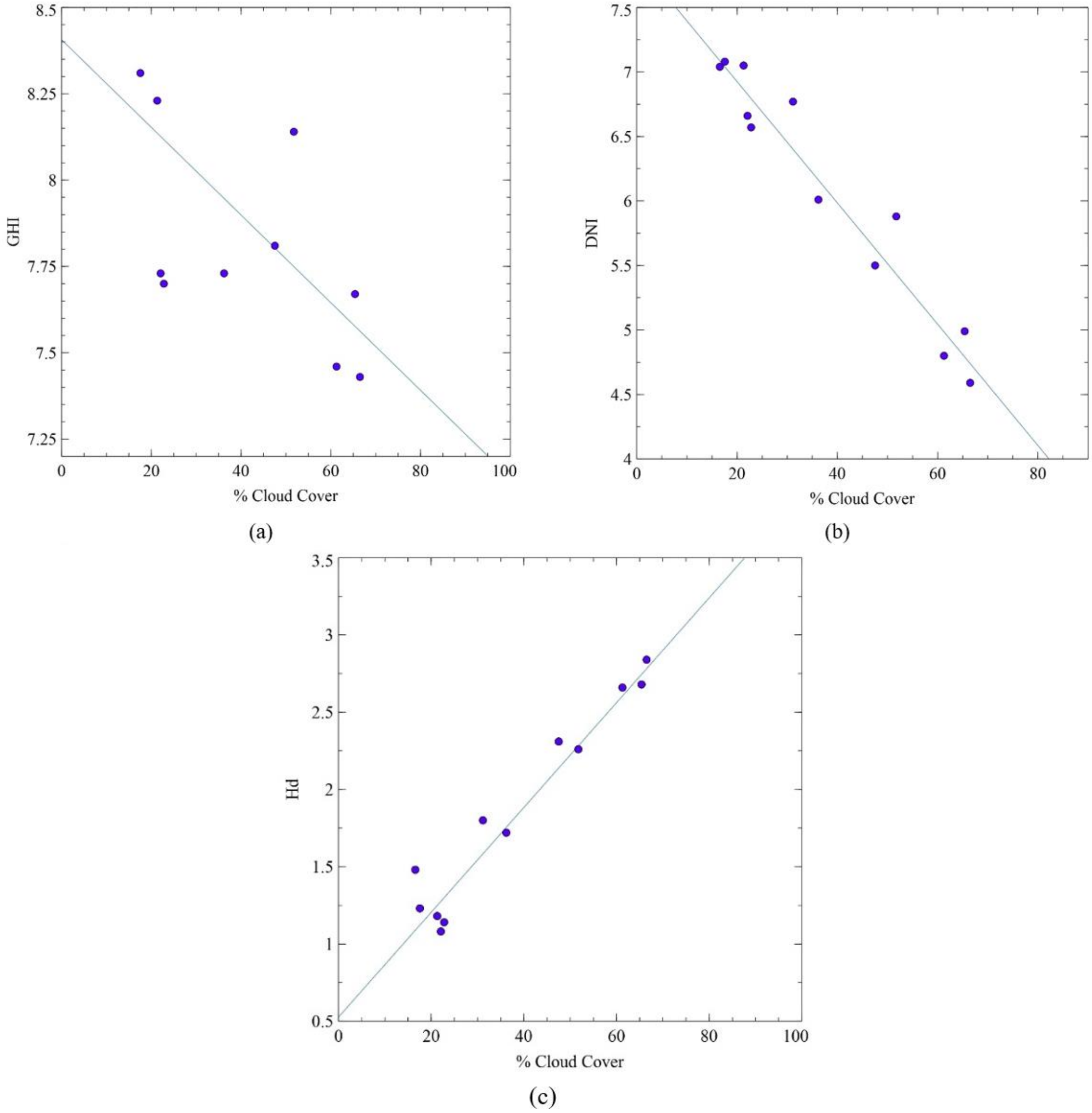


Figure 3. Relationship between **(a)** H_g, **(b)** H_b and **(c)** H_d with % cloud cover.

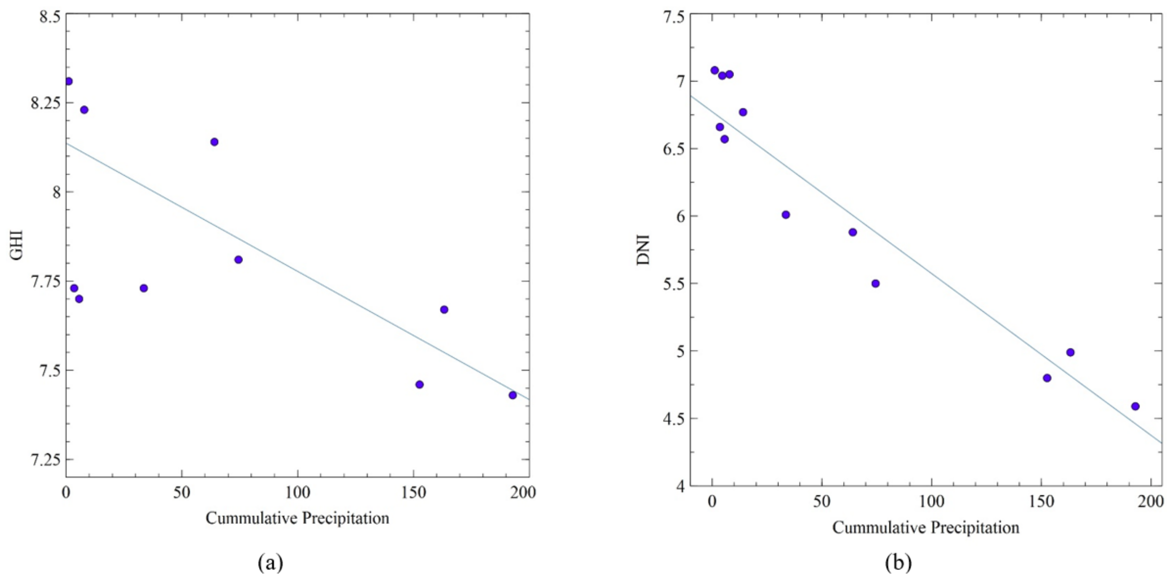
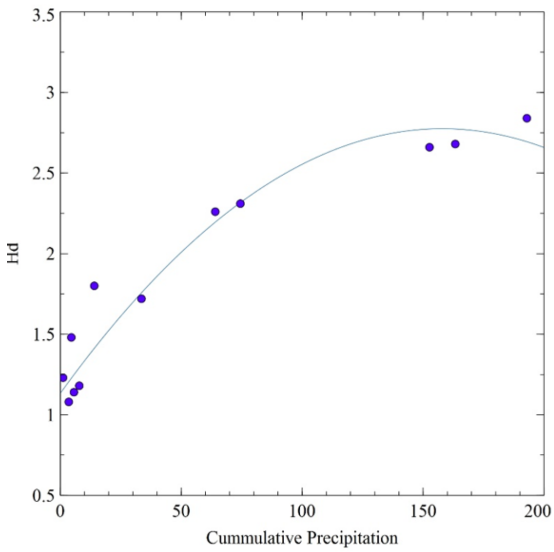
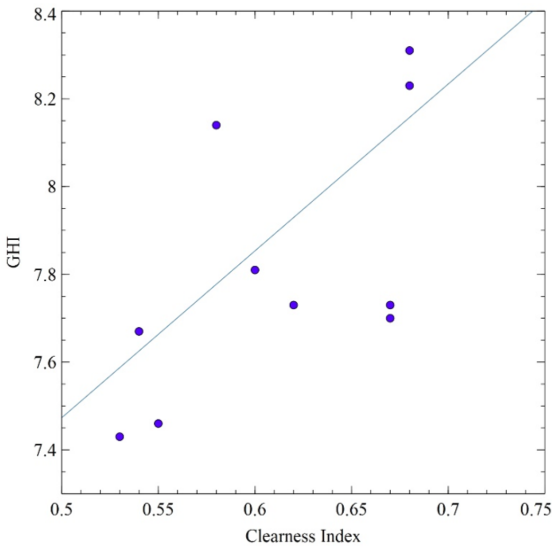


Figure 4. (*Continued*).

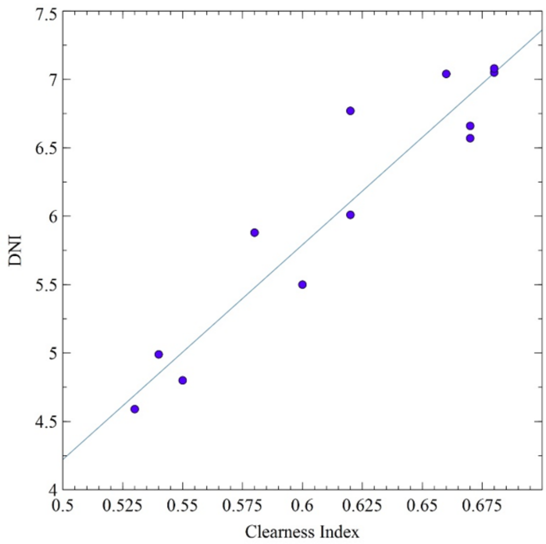


(c)

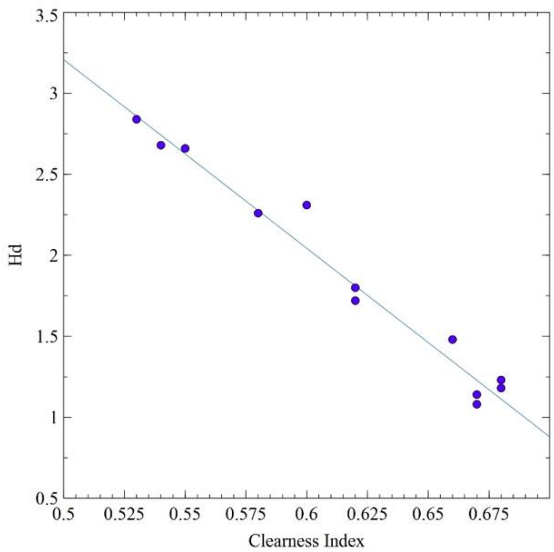
Figure 4. Relationship between **(a)** H_g, **(b)** H_b and **(c)** H_d with cumulative precipitation.



(a)



(b)



(c)

Figure 5. Relationship between **(a)** H_g, **(b)** H_b and **(c)** H_d with K_T.

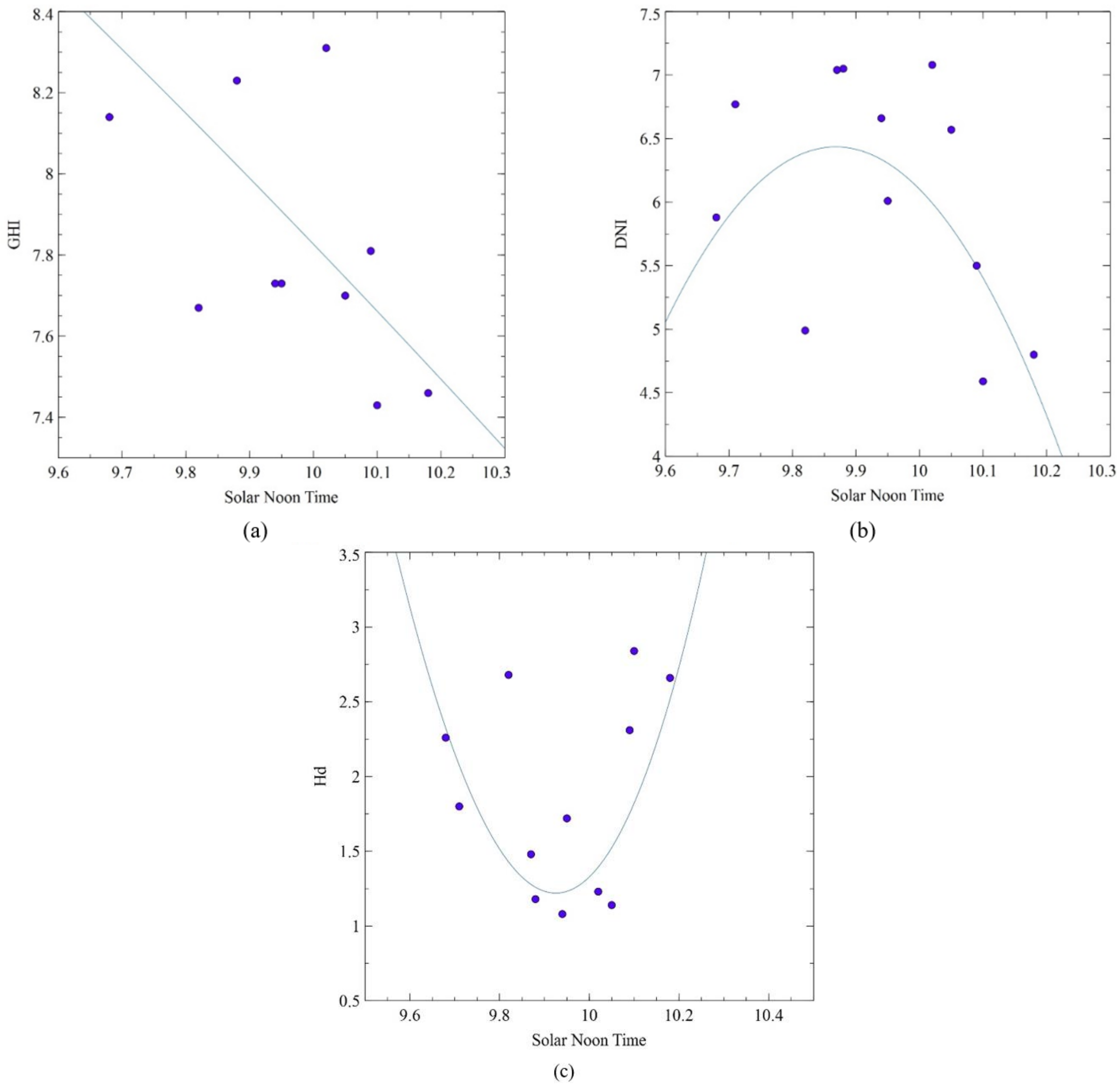


Figure 6. Relationship between **(a)** H_g, **(b)** H_b and **(c)** H_d with solar noon time.

There was also a strong relationship between % cloud cover with H_d and H_b whereas the relationship was relatively weaker with H_g. As shown by **Figure 3**, increasing the % cloud cover had an effect of reducing H_b and increasing H_d which is expected. The study also revealed a strong relationship between precipitation and, H_b and H_d while the relationship was somewhat weaker with H_g. H_b had a linear relationship with precipitation while H_d had a somewhat quadratic relationship. As the precipitation increased, H_b reduced while H_d increased with an overall effect of reducing H_g in a similar way to % cloud cover.

The relationships between day length and, H_d and H_b were relatively strong with respective linear and quadratic relationships. However, the relationship between day length and H_g was weak with a correlation of −0.19. As day length increases, there was an overall increase in H_d while as day length increased beyond 12 hours, there was no significant increase in H_b. On the other hand, the analysis of declination indicated that it had similar relations as those of day length. The relationship between declination and, H_b and H_d had R^2 of respectively 0.61 and 0.86. A similar relationship was observed for the sunset hour angle. Furthermore, there is less H_d in winter due to relatively clear skies. K_T also has a strong relationship with H_b and H_d. The R^2 values for K_T and, H_b and H_d are respectively 0. 97 and 0. 89.

Analysis of pressure and temperature showed a good correlation with H_d while its was weak with H_b and H_g. There is also a notable increase in H_d with increasing temperature. Relative humidity is also an important factor with a strong relationship with H_g and H_b. Solar noon time had no significant relationship with solar irradiance.

3.3. Modelling and analysis

The study analyzed multicollinearity among the parameters and several variables were found to be correlated and these include precipitation and cloud cover, and, day length, declination and sunset hour angle. For example, the relationship between precipitation and cloud cover had an R^2 of 98.24% and day length had a perfect correlation with declination with R^2 of 99%. Further, K_T also had a strong correlation, $R^2 = 0.98$ with precipitation and was also well correlated with cloud cover. Having analyzed all parameters with multicollinearity analysis, the variables in **Table 3** were selected for modelling.

Table 3. Parameters used in modelling.

	Parameters
H_g	Relative humidity
H_b	Relative humidity, declination, temperature and cumulative precipitation
H_d	Pressure, temperature, cumulative precipitation and declination

Meteorological data was measured from 6 different locations across Zimbabwe for one year. Monthly averaged values of the meteorological data were obtained from the measured data. The monthly averaged satellite meteorological data from 11 locations was also obtained. The collected data were relative humidity, pressure, temperature and precipitation. Declination was computed from Equation 4. The two data sets i.e., satellite data and ground measured data were correlated to obtain a correction factor to transform satellite data to approximate ground measured data. The relationships between measured data (y) and satellite data (x) are depicted by Equations 6–9.

$$\text{Pressure: } y = 0.951x + 3.385 \tag{6}$$

$$\text{Temperature: } y = 1.004x - 0.443 \tag{7}$$

$$\text{Relative humidity: } y = 1.011x - 2.539 \tag{8}$$

$$\text{Cumulative precipitation: } y = 0.906x + 5.592 \tag{9}$$

Satellite data were converted into ground measured data and multiple linear regression was employed to correlate the meteorological parameters to H_g, H_b and H_d. The developed models are as shown in Equations 10–12. Let pressure, temperature, relative humidity, declination and precipitation be respectively represented by x_1, x_2, x_3, x_4 and x_5. The study revealed that even those parameters with a lower correlation value when correlation analysis was done are important in the development of a multiple linear regression. This phenomenon was also explained by Smith[34]. The Boruta method used for screening parameters showed that parameters with low correlation with the predicted variable can still have some importance and thus can be incorporated into the multiple linear regression model. For example, correlation analysis showed that H_d has a somewhat weaker relationship with relative humidity, but the parameter selection algorithm still picked it as an important parameter hence even the less correlated variables are also important in predictive modelling.

$$H_g = 9.72 + 0.0354x_3 \tag{10}$$

$$H_b = 6.072 + 0.127x_2 - 0.0462x_3 + 0.0544x_4 + 0.0001x_5 \tag{11}$$

$$H_d = -0.225 + 0.00013x_1 + 0.0787x_2 - 0.0002x_3 + 0.00656x_4 + 0.005x_5 \tag{12}$$

The validation of the models was done using measured data and the graphs for predicted vs actual were plotted as in **Figure 6**. The models were found to have good performance in their prediction with R^2 values of 0.895, 0.972 and 0.983 respectively for H_g, H_b and H_d which was higher compared to results reported by Brahma and Wadhvani[35] and comparable to the results by Singla et al.[36]. The coefficients of multiple regression for the models were respectively 0.946, 0.986 and 0.997 for H_g, H_b and H_d and this was comparable to studies in literature[37–39]. The Mean Absolute Percentage Errors (MAPE) were respectively 0.117 kWh/m^2, 0.132 kWh/m^2 and 0.044 kWh/m^2 for H_g, H_b and H_d. These metrics show a very close agreement between measured and predicted values. The prediction errors are relatively small while the R^2 values are high indicating good performance of the predictive models. Comparison was made with established models in literature proposed by Page[33] for H_g prediction shown by Equation 1 while the Diffuse Irradiance was predicted using a model by Liu and Jordan[9] shown by Equation 5. Comparison with these models revealed better performance of the developed models compared to existing empirical models. For example, the developed models were better by 9.71%, 11.33% and 5.74% for predicting H_b, H_d and H_g respectively. This is attributed to the fact that exact location data were used in the development of these models.

4. Discussions

There was also a strong relationship between % cloud cover with H_d and H_b whereas the relationship is relatively weaker with H_g. Cloud cover has an overall effect of reducing H_g mainly because most of the incoming irradiance will be reflected into space thus reducing solar irradiance reaching the surface. Precipitation was also found to affect H_d and H_b in that it reduces the incident beam irradiance (H_b) while amplifying the presents of diffuse irradiance (H_b).

As day length increases, there was an overall increase in H_d while as day length increased beyond 12 hours, there was no significant increase in H_b. This is because day length increases by an early sunrise and a delayed sunset. As expected, this increase will only impact the H_d component because there is a low H_b component in the morning and sunset. The relationship between declination and, H_b and H_d had R^2 of respectively 0.61 and 0.86. This is because declination is negative in summer and positive in winter for Zimbabwean locations. A similar relationship was observed for the sunset hour angle. Despite having shorter days in winter, the clear skies have an effect of increasing the H_b component while the H_d component reduces unlike in summer where there is more rain and cloud cover. This is also confirmed by low values of H_b in the wettest periods in summer which are December and January. Furthermore, there is less H_d in winter due to relatively clear skies. K_T also has a strong relationship with H_b and H_d. There was a positive relation between H_b and K_T because as the skies become clearer, there is more Direct Normal Irradiance reaching the earth while Diffuse Irradiance is reduced. The R^2 values for K_T and, H_b and H_d are respectively 0. 97 and 0. 89. This could be the reason why many empirical models developed in literature mainly considered K_T as a key parameter.

Pressure and temperature had a good correlation with H_d while its was weak with H_b and H_g and a notable increase in H_d with increasing temperature was noted and this is attributed to the fact that cloud cover has a greenhouse effect which increase temperature and reduce H_b while increasing H_d. Relative humidity is also an important factor with a strong relationship with H_g and H_b. Increasing relative humidity reduces both H_g and H_b due to the increased particulate matter in the atmosphere caused by the presents of water droplets which reflect and scatter incoming irradiance.

5. Conclusions and future studies

The study seeks to develop irradiance prediction models which are critical in forecasting solar irradiance in design and maintenance of solar plants. Machine learning and statistical approaches were adopted in model development to increase the prediction accuracy. The location dependent models developed predict solar irradiance in the absence of solar irradiance data by making use meteorological data.

The study analyzed the influence of each of the selected meteorological parameters on solar irradiance. Boruta algorithm was used to select all the relevant parameters while multicollinearity and correlation analysis were also performed to analyse the impact of each of the parameters on one another. Models were developed to predict solar irradiance and the following were noted:

- The individual correlation of H_g with the parameters analysed only showed a strong correlation with relative humidity.
- Increasing the day length does not significantly increase H_b but has an overall effect of increasing H_g.
- Lower correlation between solar irradiance and a selected parameter does not necessarily mean the parameter is insignificant. The parameter selection algorithm used indicated that even the less correlated parameters could still be important in predictive modelling.
- The study noted that direct prediction of Global Horizontal Irradiance using meteorological variables is less accurate compared to using the sum of predicted Diffuse Irradiance and Direct Normal Irradiance.

The developed models were however dependent on satellite measured historical data which may not take into consideration the constant effect of global warming. It is therefore recommended that models be developed that take real-time data as input.

Author contributions

Conceptualization, KC and RM; methodology, KC; software, RM; validation, RA and CE; formal analysis, KC; investigation, KC; resources, CE and RA; data curation, RM; writing—original draft preparation, KC; writing—review and editing, CE; visualization, RM; supervision, CE; project administration, CE; funding acquisition, CE. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Eitan A. Promoting renewable energy to cope with climate change—Policy discourse in Israel. *Sustainability* 2021; 13(6): 3170. doi: 10.3390/su13063170
2. Jamil B, Bellos E. Development of empirical models for estimation of global solar radiation exergy in India. *Journal of Cleaner Production* 2019; 207: 1–16. doi: 10.1016/j.jclepro.2018.09.246
3. Li D, Heimeriks G, Alkemade F. The emergence of renewable energy technologies at country level: Relatedness, international knowledge spillovers and domestic energy markets. *Industry and Innovation* 2020; 27(9): 991–1013. doi: 10.1080/13662716.2020.1713734
4. Müller R, Pfeifroth U. Remote Sensing of solar surface radiation—A reflection of concepts, applications and input data based on experience with the effective cloud albedo. *Atmospheric Measurement Techniques* 2022; 15(5): 1537–1561. doi: 10.5194/amt-15-1537-2022
5. Liu P, Tong X, Zhang J, et al. Estimation of half-hourly diffuse solar radiation over a mixed plantation in north China. *Renewable Energy* 2020; 149: 1360–1369. doi: 10.1016/j.renene.2019.10.136
6. Nunez Munoz M, Ballantyne EEF, Stone DA. Development and evaluation of empirical models for the estimation of hourly horizontal diffuse solar irradiance in the United Kingdom. *Energy* 2022; 241: 122820. doi: 10.1016/j.energy.2021.122820
7. El Mghouchi Y, El Bouardi A, Choulli Z, Ajzoul T. Models for obtaining the daily direct, diffuse and global solar radiations. *Renewable and Sustainable Energy Reviews* 2016; 56: 87–99. doi: 10.1016/j.rser.2015.11.044
8. Ye H, Yang B, Han Y, Chen N. State-of-the-art solar energy forecasting approaches: Critical potentials and challenges. *Frontiers in Energy Research* 2022; 10: 875790. doi: 10.3389/fenrg.2022.875790
9. Liu BYH, Jordan RC. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy* 1960; 4(3): 1–19. doi: 10.1016/0038-092X(60)90062-1
10. Collares-Pereira M, Rabl A. The average distribution of solar radiation-correlations between diffuse and hemispherical and between daily and hourly insolation values. *Solar Energy* 1979; 22(2): 155–164. doi: 10.1016/0038-092X(79)90100-2
11. Mecibah MS, Boukelia TE, Tahtah R, Gairaa K. Introducing the best model for estimation the monthly mean daily global solar radiation on a horizontal surface (Case study: Algeria). *Renewable and Sustainable Energy Reviews* 2014; 36: 194–202. doi: 10.1016/j.rser.2014.04.054
12. Dal Pai A, Escobedo JF, Dal Pai E, dos Santos CM. Estimation of hourly, daily and monthly mean diffuse radiation based on *MEO* shadowring correction. *Energy Procedia* 2014; 57: 1150–1159. doi: 10.1016/j.egypro.2014.10.102

13. Feng Y, Cui N, Zhang Q, et al. Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain. *International Journal of Hydrogen Energy* 2017; 42(21): 14418–14428. doi: 10.1016/j.ijhydene.2017.04.084
14. Iqbal M. *An Introduction to Solar Radiation*. Academic Press; 1983.
15. Roderick ML. Estimating the diffuse component from daily and monthly measurements of global radiation. *Agricultural and Forest Meteorology* 1999; 95(3): 169–185. doi: 10.1016/S0168-1923(99)00028-3
16. Hove T, Manyumbu E, Rukweza G. Developing an improved global solar radiation map for Zimbabwe through correlating long-term ground- and satellite-based monthly clearness index values. *Renewable Energy* 2014; 63: 687–697. doi: 10.1016/j.renene.2013.10.032
17. Boriratrit S, Fuangfoo P, Srithapon C, Chatthaworn R. Adaptive meta-learning extreme learning machine with golden eagle optimization and logistic map for forecasting the incomplete data of solar irradiance. *Energy and AI* 2023; 13: 100243. doi: 10.1016/j.egyai.2023.100243
18. Tahir ZUR, Hafeez S, Asim M, et al. Estimation of daily diffuse solar radiation from clearness index, sunshine duration and meteorological parameters for different climatic conditions. *Sustainable Energy Technologies and Assessments* 2021; 47: 101544. doi: 10.1016/j.seta.2021.101544
19. Chen JL, He L, Chen Q, et al. Study of monthly mean daily diffuse and direct beam radiation estimation with MODIS atmospheric product. *Renewable Energy* 2019; 132: 221–232. doi: 10.1016/j.renene.2018.07.151
20. van Kuijk K. *Solar PV Potential in Rural Zimbabwe: Analysing Natural and Economical Potential* [Master's thesis]. Vrije Universiteit Amsterdam; 2012.
21. Maps of World. Available online: https://www.mapsofworld.com/lat_long/zimbabwe-lat-long.html/ (accessed on 9 November 2022).
22. Khan W, Walker S, Zeiler W. A bottom-up framework for analysing city-scale energy data using high dimension reduction techniques. *Sustainable Cities and Society* 2023; 89: 104323. doi: 10.1016/j.scs.2022.104323
23. Velliangiri S, Alagumuthukrishnan S, Thankumar joseph SI. A review of dimensionality reduction techniques for efficient computation. *Procedia Computer Science* 2019; 165: 104–111. doi: 10.1016/j.procs.2020.01.079
24. Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. *Journal of Big Data* 2020; 7(1): 52. doi: 10.1186/s40537-020-00327-4
25. Farhana N, Firdaus A, Darmawan MF, Ab Razak MF. Evaluation of Boruta algorithm in DDoS detection. *Egyptian Informatics Journal* 2023; 24(1): 27–42. doi: 10.1016/j.eij.2022.10.005
26. Kursa MB. Boruta for those in a hurry. Available online: https://cran.r-project.org/web/packages/Boruta/vignettes/inahurry.pdf (accessed on 10 December 2022).
27. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *Journal of Statistical Software* 2010; 36(11): 1–13. doi: 10.18637/jss.v036.i11
28. Kursa MB, Jankowski A, Rudnicki WR. Boruta—A system for feature selection. *Fundamenta Informaticae* 2010; 101(4): 271–285. doi: 10.3233/FI-2010-288
29. Veusz—A scientific plotting package. Available online: https://veusz.github.io/ (accessed on 12 March 2023).
30. Allen MP. The problem of multicollinearity. In: Understanding Regression Analysis. Springer; 1997. pp. 176–180.
31. Asadi Shamsabadi E, Salehpour M, Zandifaez P, Dias-da-Costa D. Data-driven multicollinearity-aware multi-objective optimisation of green concrete mixes. *Journal of Cleaner Production* 2023; 390: 136103. doi: 10.1016/j.jclepro.2023.136103
32. Voss DS. Multicollinearity. In: Kempf-Leonard K (editor). *Encyclopedia of Social Measurement*. Elsevier; 2005. pp. 759–770
33. Page JK. The estimation of monthly mean values of daily total short wave radiation on vertical and inclined surfaces from sunshine records for latitudes 40°N–40°S. Available online: https://digitallibrary.un.org/record/3827996 (accessed on 1 December 2022).
34. Smith G. Step away from stepwise. *Journal of Big Data* 2018; 5(1): 32. doi: 10.1186/s40537-018-0143-6
35. Brahma B, Wadhvani R. Solar irradiance forecasting based on deep learning methodologies and multi-site data. *Symmetry* 2020; 12(11): 1830. doi: 10.3390/sym12111830
36. Singla P, Duhan M, Saroha S. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. *Earth Science Informatics* 2022; 15(1): 291–306. doi: 10.1007/s12145-021-00723-1
37. Bamigbola OM, Atolagbe SE. Empirical models for predicting global solar radiation on the African continent based on factors of location and season. *Open Journal of Modelling and Simulation* 2021; 9(1): 59–73. doi: 10.4236/ojmsi.2021.91004
38. Bamisile O, Cai D, Oluwasanmi A, et al. Comprehensive assessment, review, and comparison of AI models for solar irradiance prediction based on different time/estimation intervals. *Scientific Reports* 2022; 12(1):9644. doi: 10.1038/s41598-022-13652-w
39. Benatiallah D, Bouchouicha K, Benatiallah A, et al. Forecasting of solar radiation using an empirical model. *Algerian Journal of Renewable Energy and Sustainable Development* 2019; 1(2): 212–219. doi: 10.46657/ajresd.2019.1.2.11