Irradiance Prediction Model with Step wise Backward Linear Regression

[1]Rahul Pathak, [2]Mohd. SaquibFaraz, [3]Mayank Jain
[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor
[1]Electrical & Electronics Engg. Deptt.
[1]ADGITM, GGSIPUNew Delhi, INDIA

***Abstract* :.**The objective of this work is to ascertain whether reliable prediction of Global Horizontal Irradiance (GHI) is attainable using predictor variables: wind speed, wind direction, temperature, relative humidity, solar zenith angle, date of year and time. Multivariate backward stepwise linear regressions was utilized to determine significant attributes for GHI estimation. The model was then used to predict value of GHI. The performance of models was investigated using model residuals. The resulting model could explain 67% of variation in data. The model was further tested for multicollinearity and heteroscedasticity. The model gives satisfactory VIF and MAPE values and nominal heterosdasticity. R computation platform was used for modeling and computation.

***IndexTerms*- Multicollinearity, Heteroscdasticity, Linear regression, Variable inflation factor**

I. INTRODUCTION

India receive solar energy equivalent of more than 5000 trillion kWh/year. The Intensity of solarradiation remains at an average value of 200 MW/km2 [1]. 58% of land area receive annual average globalinsolation above 5 kWh/m2/day with no of sunny days250 to 325 over a year. India has tremendous potential for solar energy [2-3].India has achieved record capacity addition of 5525.98 MW of solar power during 2017-18. India has set up ambitious plans of achieving 100 GW of solar energy by year 2022[4].

Prediction models plays important role for estimating solar potential at specific regions. The estimation is generally a two-step process. At first stage, meteorological information is used to evaluate solar irradiance. Second stage involves use of solar irradiance data, PV cell information and temperature in power modelling algorithm to estimate PV cell output [5-6]. Several work exists in literature for irradiance estimation. Methods include use of neural networks, time series forecasting, use of satellite data and numerical weather prediction (NWP) [7-9].Present study deals with predictor model of GHI estimation. The predictor model is built on the basis of meteorological data like wind speed, wind direction, temperature, relative humidity for mehrauli locality in New Delhi, India. Stepwise backward regression was used for the model. All computation work was carried out in R. The article is organized as follows : Section 1 gives background of the study, section 2 enumerates on data set used in the study , section 3 explains linear regression, results are presented in section 4 followed by conclusion in section 5.

II. DATA SET

The data set contains 4745 observations from Mehrauli,New Delhi ,India for year 2012. It has seven variables: UID containing information about date and time,temperature,relative humidity, solar zenith angle, Wind speed, Wind Direction and GHI. Observations were taken on hourly basis between 6 am to 6pm, thus there were 12 observations corresponding to every day.

Data corresponding to month, day of month and time were merged to create single variable UID. Complete data set is available online at National Renewal Energy Laboratory, Department of Energy, and USA [10]. Brief description of data set is presented in table.

Table : 1 DATA SET

Sr.No	Variable Name	Description	Class
1.	UID	Information for Date and Time	Integer
2.	Temperature	Atmospheric Temperature	Integer
3	Relative Humidity	Relative Humidity	Integer
4	SolarZenith.angle	Angle in radians	Integer
5.	Wind direction	Wind Direction	Integer
6.	Wind speed	Wind Speed in meter per second	Integer
7.	GHI	Global Horizontal Irradiance	Integer

III. LINEAR REGRESSION

Linear regression is a ubiquitous statistical technique to formulate a relationship between dependent and independent variables. Independent variables are also known as predictor variable, their values are obtained through experiments. The value of dependent variable or predicted variable is obtained from the predictor variable [11-13]. Mathematically, these two variables are related by equation of straight line:

Y=Bx+c (1)

where,Y is the dependent variable, x is the independent variable.B and c are constants which are called the coefficients. In linear regression, these constant are determined by ordinary least squares algorithm. Ingeneral case of P independent variables, the population regression line for p explanatory variables x1, x2, ... ,xp can be defined as:

Actualdata=Fittedmodel+Error (2)

Where fitted model is represented as:
Y=c+B1x1+B2x2+B3x3+B4x4………..+Bpxp(3)

Error term is also known as residual. The "Residual" is difference between predicted value and actual value. Linear regression assumes normal distribution of residual.

IV. RESULT

4.1 First Iteration

In this work, our objective is to establish the relationship between; GHI (dependent variable) and predictors given in data set such as, wind speed , day of the year, time,wind direction, relative humidity andtemperature. R studio was usedfor computational platform [14] .During first iteration, all predictor variables were included. Results are tabulated in tables 2-3.

TABLE 2: COEFFICIENTS

Variable	Coeff. value	P- Value
Intercept	1269	<2e-16
UID	0.0001866	0.00159
Temperature	-4.548	<2e-16
Relative.Humidity	-1.858	<2e-16
Solar.Zenith.angle	-1.270	< 2e-16 ***
Wind.direction	0.03.019	< 2e-16 ***
Wind.speed	-5.799e	0.00763

TABLE 3: RESIDUALS

Min	1st quartile	median	3rd quartile	Max.
-773.73	-86.28	17.41	110.57	363.05

Adjusted R-squared: 0.7652

Adjusted R square value for this model is 0.76 which suggests that, model captures 76% variation of data set.

VIF Values of model given in Table 4. It is evident that ,there is no multicollinearity in the model. For performance evaluation of model, MAPE plots are useful tools. MAPE (Mean Absolute Percentage Error)can be evaluated as [15-16]

$$\frac{|\text{Actual value} - \text{Predicted value}|}{|\text{Actual Value}|} \quad *100(4)$$

Figure 1gives MAPE plot of model. One may observe, max MAPE in the model is up to 8%. Histogram of residuals reveal that majority of residuals lie in range of -200 to 200.

TABLE 4:VIF VALUES OF COEFFICIENTS

UID	1.906253
Temperature	1.045675
Relative.Humidity	2.062774
Solar.Zenith.angle	1.355259
Wind.Direction	1.241173
Wind.Speed	1.231764

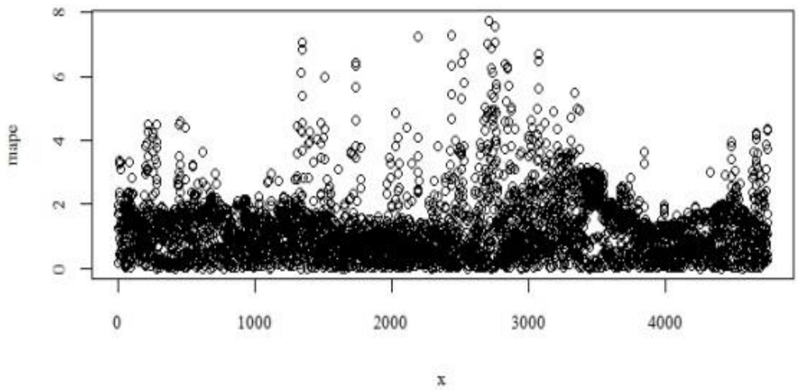

Fig. 1 MAPE plot

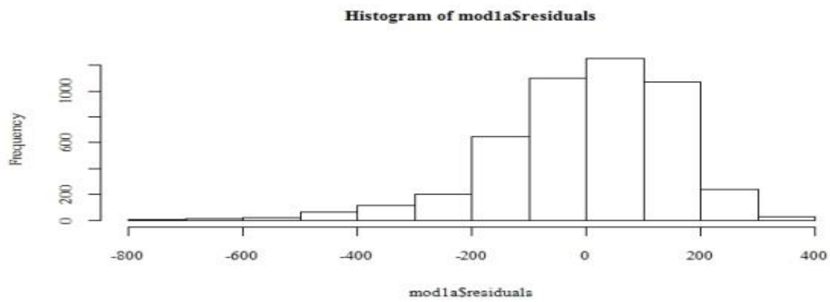

Fig. 2 Histogram of residuals

4.2 Second Iteration

As errors were significant in earlier model , square root transformation was applied . Results are tabulated in Table 5&6. Table 6, reveals that there is improvement in performance of model, as errors are reduced. From table 5, it becomes obvious that variable UID is no more significant.

TABLE 5: COEFFICIENTS

Variable	Coeff. value	P- Value
Intercept	4.580e+01	< 2e-16
UID	1.986e-06	0.286
Temperature	-1.314e-01	< 2e-16
Relative.Humidity	-5.731e-02	< 2e-16
Solar.Zenith.angle	-4.153e-01	< 2e-16
Wind.direction	8.076e-03	< 2e-16
Wind.speed	-3.759e-01	4.34e-08

TABLE 6 : RESIDUALS

Mini.	1st quartile	median	3rd quartile	Maxi.
-22.3347	-3.7561	0.6268	3.7732	13.2742

4.3Third Iteration

In third iteration, Variable UID was dropped from the model. Coefficients values are given in table 7. Adjusted R squared value for the model is 0.77. Distribution of residuals is presented in table 8. Table 9 gives VIF values of residuals, clearly there is no multicollinearity in model. Histograms of residuals reveal that, there is improved accuracy in performance of model as range of errors is reduced. Using coefficients in table 7 , prediction model can be written as:

Sqrt(GHI)=45.819-0.130*(temp)+ 0.0568 *(Relative Humidity)- 0.4155725*(Solar.Zenith.Angle + 0.0081915*(wind.direction)-0.37779(Wind.speed) (5)

TABLE 7: COEFFICIENTS

Variable	Coeff. value	P- Value
Intercept	45.8192690	< 2e-16
Temperature	-0.1306465	< 2e-16
Relative.Humidity	-0.0568305	< 2e-16
Solar.Zenith.angle	-0.4155725	< 2e-16
Wind.direction	0.0081915	< 2e-16
Wind.speed	-0.3777960	3.69e-08

TABLE 8 : RESIDUALS

Mini.	1st quartile	Median	3rd quartile	Maxi.
-22.3347	-3.7561	0.6268	3.7732	13.2742

TABLE 9 : VIF VALUES

VARIABLE	VIF VALUE
TEMPARATURE	1.224069
Relative.Humidity	2.053499
Solar.Zenith.angle	1.326476
Wind.Direction	1.897655
Wind.Speed	1.230964

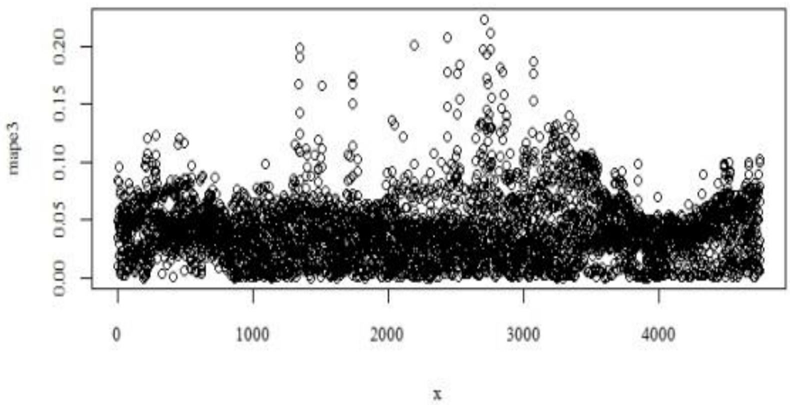

Fig.3 MAPE plot

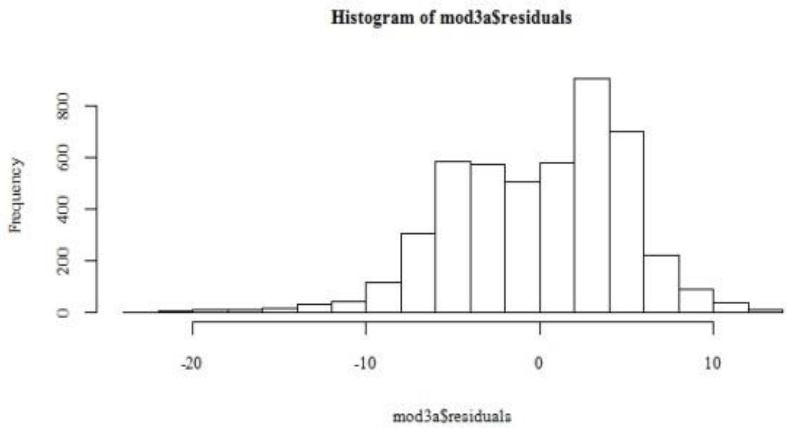

Fig. 4 Histogram of Residulas

CONCLUSION

The objective of present work was to formulate a simple and efficient predictor model for prediction ofGHI, based on 4470 observations.wind speed , day of the year, time,wind direction, relative humidity andtemperature were used as predictor variables to estimate the GHI. Multivariate linear regression model was employed to investigate the influence of independent variables. The p-value corresponding to individual characteristics was utilized to formulate the most significant reduced model for the prediction. The GHI decreased with day of the year, time,wind direction, relative humidity andtemperature while it increased with wind speed. The analysis show that the regression can explain 77% of the variation in GHI. The model is free from multicollinearity.

.

REFERENCES

[1] N. Sharma, P.K Tiwari, Y.R Sood, “Solar energy in India:Strategies, policies, perspectives and future potential” Renewableand Sustainable Energy Reviews 16 933– 941 2012.

[2] R.Aringhoff, G.Brakmann and M.Geyer and S.Teske “Concentrated solar thermal power now.” Greenpeace International, 2005.

[3] http://eai.in/blog/2009/09/national-solar-mission-to-beannounced.Html

[4] Press Information Bureu, Government of India, Ministry of New and Renewable energy, 27 december,2017. http://pib.nic.in/newsite/PrintRelease.aspx?relid=174832

[5] Deshmukh, M. K., and S. S. Deshmukh, 2008: Modeling of hybrid renewable energy systems. Renewable Sustainable Energy Rev., 12, 235–249, doi:10.1016/j.rser.2006.07.011.

[6] Zhou, W., H. Yang, and Z. Fang, 2007: A novel model for photovoltaic array performance prediction. Appl. Energy, 84, 1187–1198, doi:10.1016/j.apenergy.2007.04.006

[7] Geuder, N., F. Trieb, C. Schillings, R. Meyer, and V. Quaschning, 2003: Comparison of different methods for measuring solar irradiation data. Third Int.Conf. onExperienceswithAutomaticWeather Stations, Torremolinos, Spain, Instituto de NacionalMeteorologia.

[8] Myers, D. R., 2005: Solar radiation modeling and measurements for renewable energy applications: Data and model quality.Energy, 30, 1517–1531, doi:10.1016/j.energy.2004.04.034..

[9] "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids", "Renewable and Sustainable Energy Reviews",27,65 – 76, 2013, 1364-0321

[10] National Renewal Energy Laboratory, Department of Energy, USA. https://www.nrel.gov/research/data-tools.html

[11] Courville, T., & Thompson, B. (2001). Use of Structure Coefficients in Published Multiple Regression Articles: β is not Enough. Educational and Psychological Measurement, 61(2), 229–248.

[12] Yi-Chung Hu “Multilayer Perceptron for Robust Nonlinear Interval Regression Analysis Using Genetic Algorithms” ScientificWorldJournal. 2014; 2014: 970931. Published online 2014 Jun 29.

[13] MehranHojati, C.R Bector, Kamal Smimou,“A simple method for computation of fuzzy linear regression” European Journal of Operational Research Volume 166, Issue 1, 1 October 2005, Pages 172-184

[14] R official home page. https://www.rstudio.com/products/rstudio/.

[15] Engineering statistics DC Montgomery, GC Runger, NF Hubele 2011 – Wiley

[16] Capizzi, Giovanna, Masarotto, Guido ,A Least Angle Regression Control Chart for Multidimensional Data,Technometrics 285-296 vol-53 ,Taylor & Francis, doi: 10.1198/TECH.2011.10027