

Orthogonal Regression Analyses of solar Irradiation and DC Power Data

Surendra H H, Seshachalam D, Sudhindra K R



Abstract -The rapid deployment of solar photovoltaic cells in recent times is enabling data analysis of renewable energy solar source and its conversion to electrical energy. The need for estimating the capacity of solar irradiation and solar DC power in any geographical location is necessary for short-term and long-term forecasting. In this paper a regression-based data analysis is carried out at different locations in India and analyze the residue for better fit. Here solar DC power is a dependent output variable and solar irradiance is an independent predictor. The orthogonal regression is a standard linear regression technique involves in reducing the sum of squared orthogonal projections. The primary goal of this paper is to analyze orthogonal regression technique with the relationship between response (DC power data) and predictor (solar irradiation) across different diverse locations. The fitted value for the solar irradiance accounts for the uncertainty in the value of the predictor. The study of nonlinear nature of solar irradiation and DC power generated across geographically dispersed sites for a time series data leads to better estimate of solar power.

Keywords- Irradiance, orthogonal regression, least square, error variance ratio, residual

I. INTRODUCTION

The societal impact due to deployment of solar photovoltaics is substantial due to reduced carbon emission. This renewable energy which fluctuates on hourly basis often requires storage due to its inability to generate during no light. The metrological data of country and solar irradiance is estimated and this acts as an input to various power modelling algorithms. The short term and long-term forecasting depend on past and present data. Having more data of the past leads to better forecasting of weather and hence the DC power. Efforts by National renewable energy laboratory has created a large database and has thus created a better prediction and forecasting. The system advisor model (SAM) tool is used to simulate the data over different locations. Most governments have their solar resource maps with generating capability at a resource site. The data accuracy and time over which regression model is analyzed gives the dynamic range of the nation's solar resource [1]. Regression is used for predicting and forecasting. Regression analysis gives the factors that matters and variables having an impact.

Revised Manuscript Received on February 28, 2020.

* Correspondence Author

Mr. Surendra H.H.*, Assistant Professor, Department of Electronics and Communication, BMS College of Engineering, Bangalore, India.

Mr. Seshachalam D, Professor, Department of Electronics and Communication, BMS College of Engineering, Bangalore, India.

Mr. Sudhindra K.R., Associate Professor in the Department of Electronics and Communication, BMS College of Engineering, Bangalore, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

The dependent variable is the main factor for predicting. The solar input variables are the predictors of a statistical box which has a function to predict output variable [2].

Orthogonal regression or deming regression links linear relationship between two continuous variables of predictor and response. It contains measurement errors which are independent. The orthogonal regression fitting method is acceptable if error variance ratio is specified correctly. When assumptions are correct it is a better estimation method [3].

A linear regression represents systematic relations between two variables described by a straight line. In linear regression a straight line that is near to actual data is considered. In least squares method unique line minimizing sum of squared distances between correct data and regression line is considered. Minitab 19 tool is used to simulate and analyze the solar data which is suited for instructional applications and research data. This tool developed in 1972 by Barbara Ryan, Thomas Ryan and Brian Joiner automates calculations and creates graphs useful for any research analysis and interpretation of measured or observed data.

India has abundance solar and wind resources with 850-3400GW for wind and 1300-5200Gw for solar and even if lower estimates are considered it is expected to meet the demand of 2030. Since resources are geographically unevenly spread with more concentrated in western and southern states there is a need for balancing these resources [7]. The government has also revised the target set for installation as it exceeded the target. The solar energy exceeds all of the fossil fuel available in India and also India has lower capital cost per MW generation of solar plant compared to global standards leading to cost effective installation [8].

The average DC power generated can be computed using solar photovoltaic cell model given by the equation shown below [9].

$$DC-P_{avg} = P_s f_s$$

P_s is Photovoltaic module output power,

f_s is irradiance probability density function.

II. METHODOLOGY-LINEAR AND ORTHOGONAL REGRESSION

The typical five techniques for regression analysis are Deming regression, ordinary least squares, orthogonal distance regression, weighted Orthogonal Distance regression and York regression [11]. The regression equation is an algebraic equation that describes relation between response and the terms in the model in the form

$$Y = b + mx \quad (1)$$

Y is the response variable (the output Solar direct current value) b is the constant or intercept,

m is the estimated coefficient for the linear term (also known as the slope of the line), x is the value of the term (the input irradiance).

Simple regression is a procedure to find specific values for the slope

$$\text{Formula for slope is } m = \frac{\sum(Y - \bar{Y})(X - \bar{X})}{\sum(X - \bar{X})^2} \quad (2)$$

Here Y denotes the dependent variable and X denotes independent variable. In orthogonal regression both predictor and response are unknown [17].

The difference between observed and fitted value is residue which examines how well model fits. The residual is the difference among observed value and the fitted value.

The orthogonal distance regression incorporates error in variables designed to avoid regression attenuation by considering uncertainty in the predictors which is not considered in ordinary linear regression [4]. In orthogonal regression slope becomes a function of error variance ratio. Schematic in fig 1 shows different regression methods with Kendall and Stuart proposing the orthogonal regression taking into account errors which may be assumed in the predictor variables[5]. The schematic shows reduced distance between data points and regression lines compared to standard least square and inverted least square regression but disadvantage is that orthogonal regression needs extra input parameter. The first schematic indicates orthogonal regression and second one represents least square regression which is discussed further in this paper. The third schematic indicates inverted least square regression.

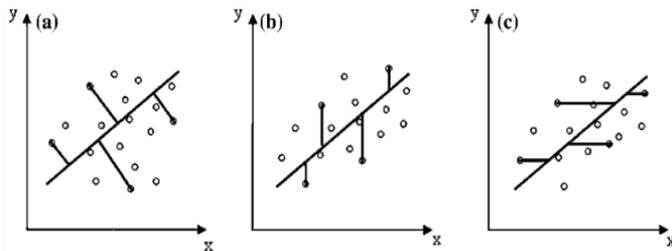


Fig 1 Typical regression methods

The slope-intercept representation is

$$Y = m * X + c \quad (3)$$

where m represents the slope and c indicates the intercept.

If a line is perpendicular to the above equation then with - (1/m) slope the equation will be

$$Y' = -X/m + c' \quad (4)$$

consider data point (X₀, Y₀) where line passes through and its equation is

$$Y' = -X/m + (X_0/m + Y_0) \quad (5)$$

The perpendicular line will intersect the fitted line at a point (X_i, Y_i) where X_i and Y_i are defined by

$$\begin{aligned} X_i &= (X_0 + m * Y_0 - m * c) / (m^2 + 1) \quad (6) \\ Y_i &= m * X_i + c \quad (7) \end{aligned}$$

So, the orthogonal distance between (X₀, Y₀) and (X_i, Y_i) is computed [19]

$$d = \text{sqrt}((X_0 - X_i)^2 + (Y_0 - Y_i)^2) \quad (8)$$

III. DATA COLLECTION

For time series data set of measurements are considered for particular time like hourly, daily, weekly, monthly or yearly data. In this paper time series monthly data is considered for the study. To investigate the regression analysis, ten available measurements from various diverse locations are considered obtained from NREL. The performance model uses PV modules, inverters in an array. The annual results are simulated using system advisor model with the following set of data. The model used for simulation is TP300LBZ solar power panel from Tata power solar systems which uses multi-c-Si material. Each module has an area of 1.97m² and in total 14 strings are used which occupies 27m² floor area. Each module capacity is 300.12 DC watts and total capacity will be 4.2KW DC.

The solar energy being nonlinear varies throughout day and season to season. The input irradiance and output DC power graph is plotted for different locations. Here input irradiance represents the independent variable and output DC power represents the dependent variable. The locations are chosen from north to south and east to west. The north south locations are from Ladakh (34.15 N, 77.45 E) in the north to Kanyakumari (8.05 N, 77.55 E) in the south through Delhi (28.65 N, 77.25 E), Bhopal (23.25 N, 77.35 E) and Bangalore (12.95 N, 77.65 E). The east west locations are from Jaisalmer (26.95 N, 70.95 E) to Siliguri (26.75 N, 88.45 E) through Jaipur (26.95 N, 75.85 E), Lucknow (26.85 N, 80.95 E) and Jorhat (26.75 N, 94.25 E). The system advisor model database contains one year's data divided in terms of hourly time which gives the solar resource at a particular location. The weather file contains data from previous history over a range of years. Typical solar resource is obtained from ground based or satellite data or both [6]. The plane of array irradiance data input to the solar module is useful in predicting the output energy from a solar PV plant. The software considers additional parameters such as location, angles of irradiance, weather parameters, losses such as shading loss and soil losses. Comparison of satellite based GHI and ground-based measurements have been observed for a period of eight months and a relative difference of 1.4% in the total insolation measured is noted. Thus, satellite-based prediction is almost similar to that of ground-based measurement made by pyranometer [10].

IV. REGRESSION ANALYSIS FOR SOLAR DATA

In this paper we evaluate orthogonal regression distance in comparison with ordinary least squares. Linear regression is used in analyzing weather science in terms of intercepts of XY datasets and derive the slopes [11]. Ordinary regression is most widely used due to its simple approach where independent variable is error free. Deming (1943) proposed reducing the sum of squares of X and Y residuals to overcome the drawbacks of Ordinary regression. Data analysis of air quality implied a reduced major axis regression is more suitable analysis [12].

The error variance ratio(δ) is the ratio of error variance of the response Y to that of predictor X.

If $\delta > 1$ it implies response is more uncertain than predictor.

If $\delta = 1$ it implies response measurement is equally uncertain as predictor.

If $\delta < 1$ it implies response is more certain than predictor measurements.

The most important thing is to correctly estimate the error variance ratio. Most researchers have accepted error variance ratio as 1 during conversion process. [15][16].

The optimal fit in normal least square regression is less than that of orthogonal regression. The regression by orthogonal concept is more suited if error variances are equal and geometric mean is suited when random errors are present [13]. Regression equations is indicated in terms of form and constant but it is important to include residue analysis also to check maximum and minimum variations from the fitted line. Regression model results in fitted values and residuals. The monthly residual plot represents the difference between the dependent variable solar DC power estimated (Y_i') and real value of dependent variable (Y_i).

$$E_i = Y_i - Y_i'$$

Analytical formulation of regression coefficient is same for general orthogonal regression and chi-square regression with coefficients coinciding[14]. The simulation is carried for ten locations with Error Variance Ratio as one. In orthogonal regression the best fitting line minimizes the weighted distances from plotted points to the line [18]. Figure 2 to figure 12 represents plot of DC power vs Irradiance with fitted line and the residual vs observational months across ten different locations of India.

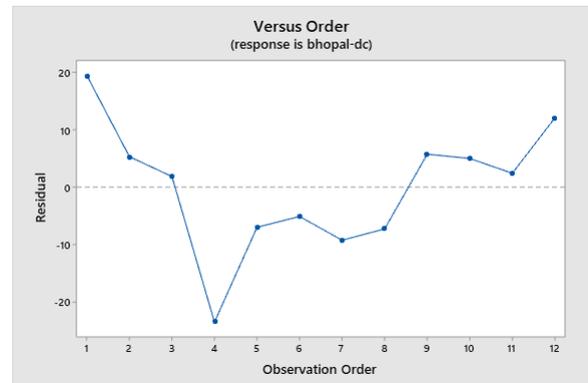
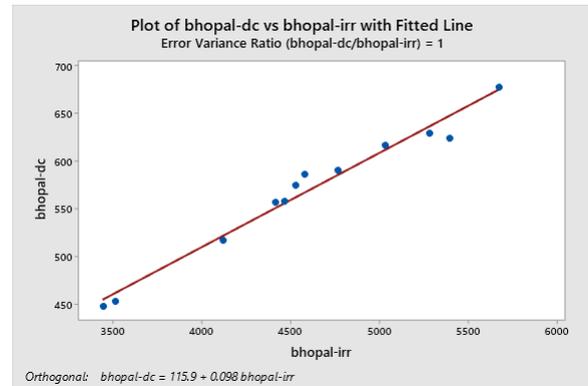


Fig 3 orthogonal regression plot of irradiance vs DC power output and residues for different months of Bhopal

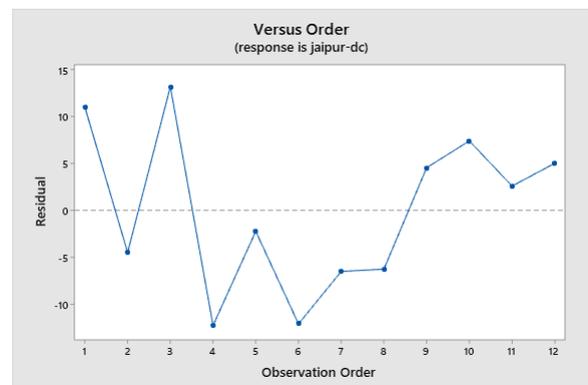
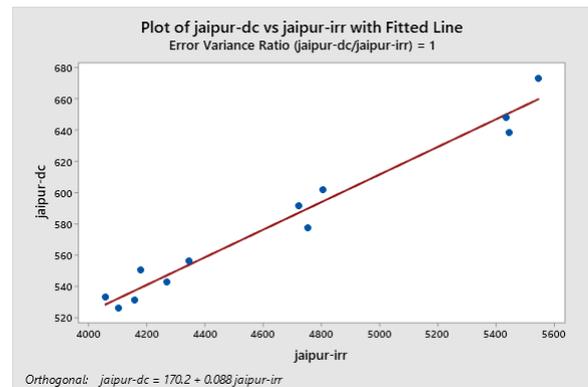


Fig 4 orthogonal regression plot of irradiance vs DC power output and residues for different months of jaipur

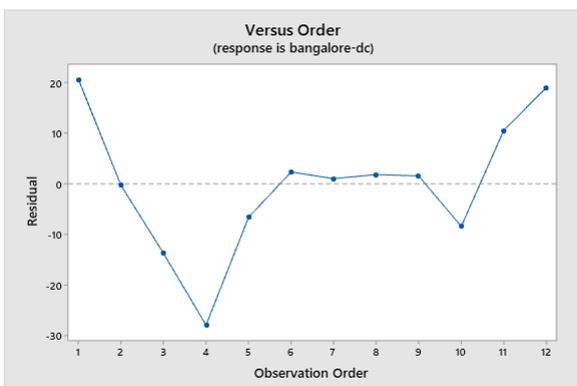
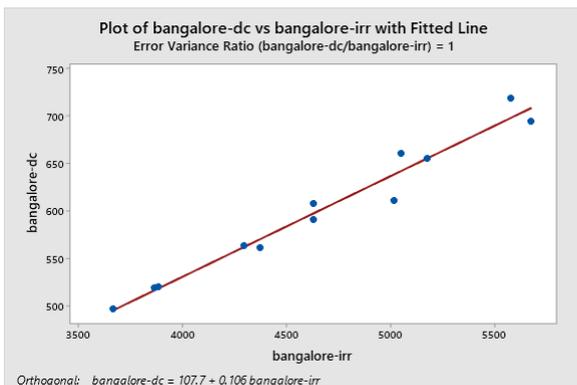


Fig 2 orthogonal regression plot of irradiance vs DC power output and residues for different months of Bangalore

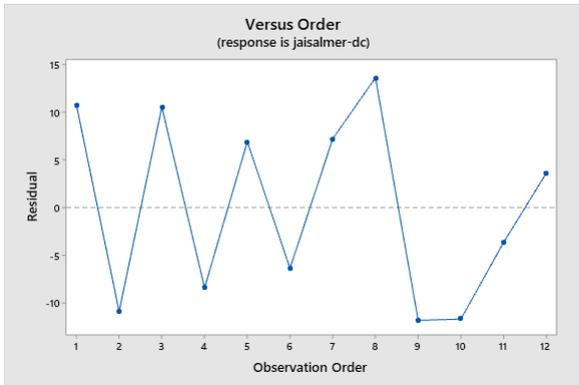
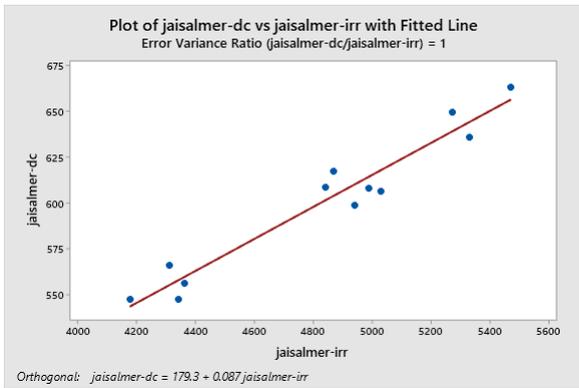


Fig 5 orthogonal regression plot of irradiance vs DC power output and residues for different months of jaisalmer

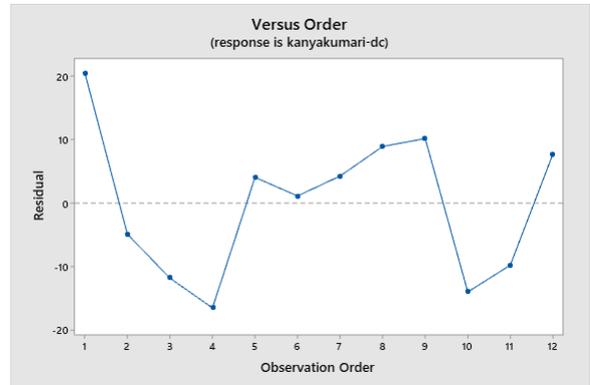
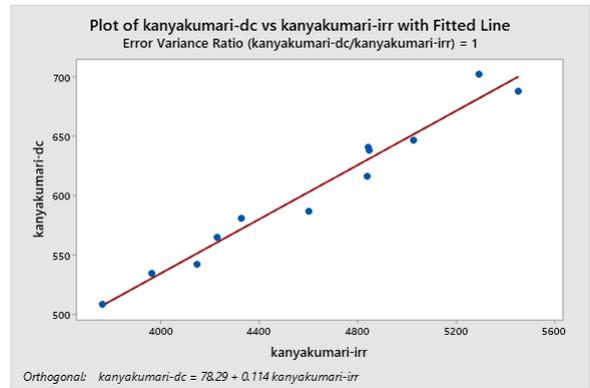


Fig 7 orthogonal regression plot of irradiance vs DC power output and residues for different months of Kanyakumari

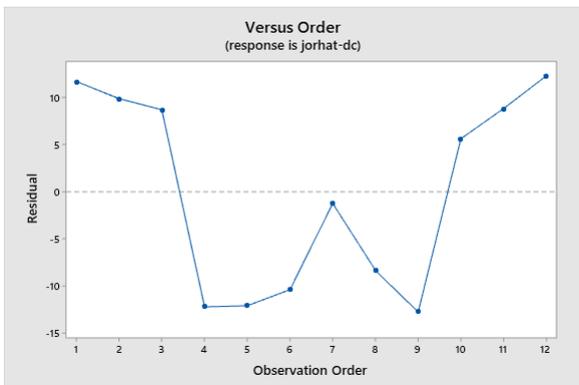
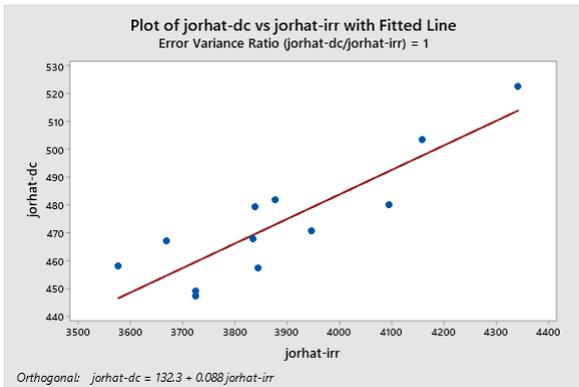


Fig 6 orthogonal regression plot of irradiance vs DC power output and residues for different months of jorhat

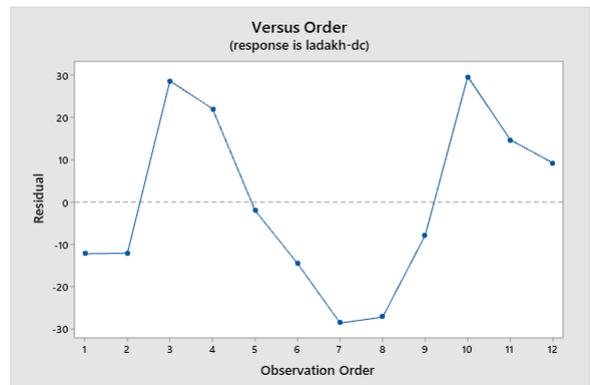
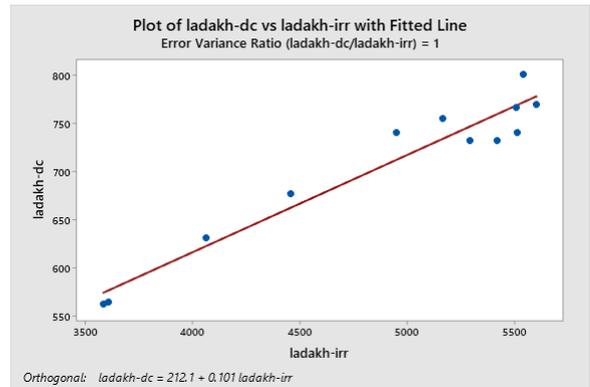


Fig 8 orthogonal regression plot of irradiance vs DC power output and residues for different months of Ladakh

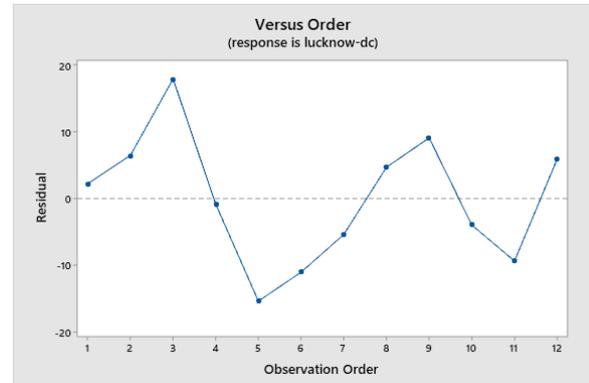
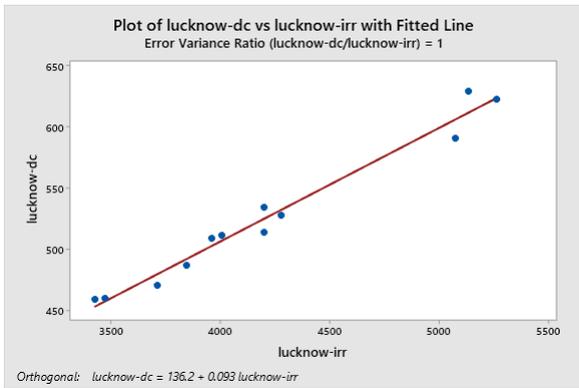


Fig 9 orthogonal regression plot of irradiance vs DC power output and residues for different months of Lucknow

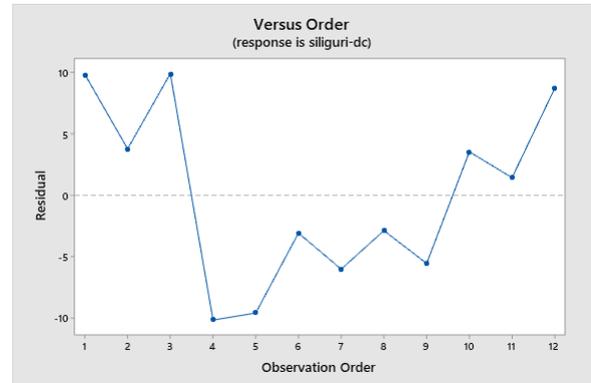
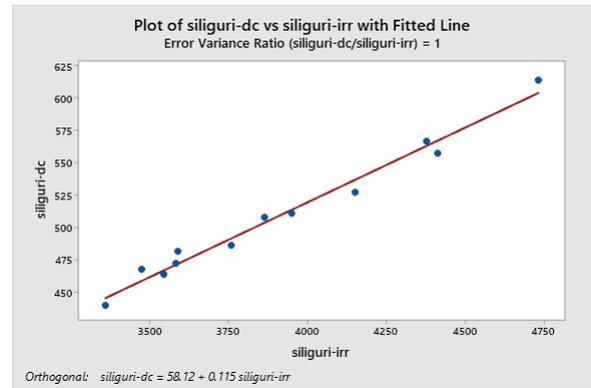


Fig 11 orthogonal regression plot of irradiance vs DC power output and residues for different months of siliguri

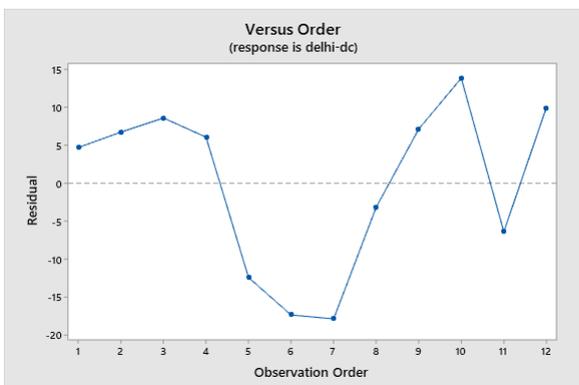
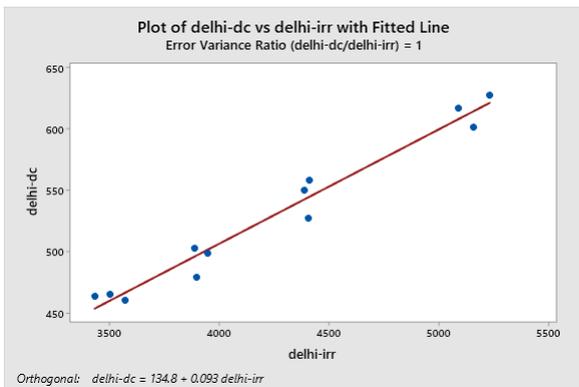


Fig 10 orthogonal regression plot of irradiance vs DC power output and residues for different months of Delhi

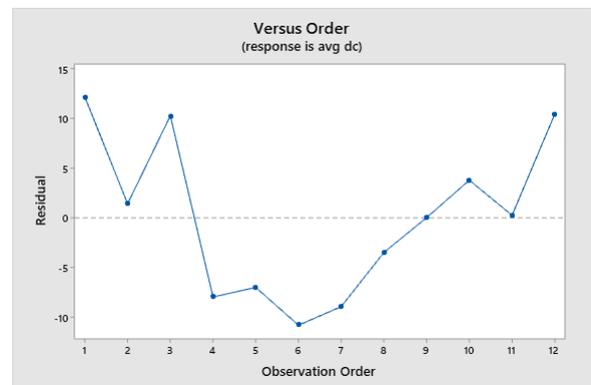
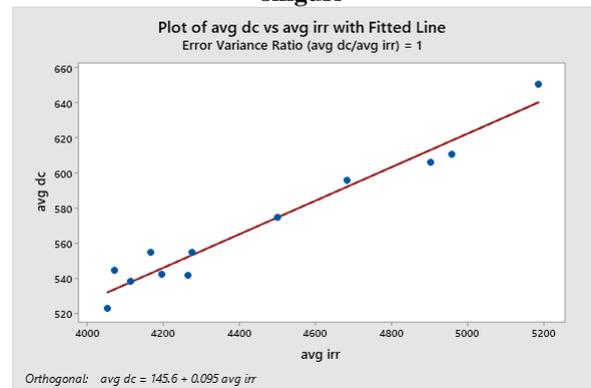


Fig 12 orthogonal regression plot of irradiance vs DC power output and residues for different months considering average of ten locations

Table 1 Regression Equations of different stations.

Stations	Orthogonal	Simple Linear
Bangalore	$Y = 107.7 + 0.106 X$	$Y' = 107.9 + 0.1057 X'$
Bhopal	$Y = 115.9 + 0.098 X$	$Y' = 116.0 + 0.09844 X'$
Jaipur	$Y = 170.2 + 0.088 X$	$Y' = 170.3 + 0.08824 X'$
Jaisalmer	$Y = 179.3 + 0.087 X$	$Y' = 179.5 + 0.08713 X'$
Jorhat	$Y = 132.3 + 0.088 X$	$Y' = 133.1 + 0.08764 X'$
Kanyakumari	$Y = 78.29 + 0.114 X$	$Y' = 78.53 + 0.1140 X'$
Ladakh	$Y = 212.1 + 0.101 X$	$Y' = 212.4 + 0.1010 X'$
Lucknow	$Y = 136.2 + 0.093 X$	$Y' = 136.3 + 0.09248 X'$
Delhi	$Y = 134.8 + 0.093 X$	$Y' = 134.9 + 0.09286 X'$
Siliguri	$Y = 58.12 + 0.115 X$	$Y' = 58.24 + 0.1153 X'$
Average	$Y = 145.6 + 0.095 X$	$Y' = 145.7 + 0.0953 X'$

As such there is choice of regression is to estimate the relation between two variables and which method is more suitable for the analysis for the given datasets. The difference between regression equation fit and actual data gives the residue

V. CONCLUSION

With dig data analytics regression is used to predict outcomes for new data. If any change in irradiance leads to change in solar DC power output the regression model fits the data well. Here the residuals appear random spread across different months and the plot of residual vs observation shows no trend which implies the errors are independent. The positive coefficient in regression equation implies as the value of irradiance increases the mean of DC power also tends to increase. The negative coefficient implies as the independent variable irradiance increases, the dependent variable DC power of solar panel tends to decrease. The residue analysis of DC power shows maximum residue for Ladakh followed by Bangalore and Bhopal. The minimum residue is in Jorhat followed by Jaisalmer and Jaipur which infers that the regression equation is better suited in these locations. Reduction in residue results in better fit of actual data leading to better estimate of the monthly irradiance and DC power data for a given location. Thus, regression models can be used to estimate the DC power generated in a geographical location.

REFERENCES

- Christopher t. M. Clack January 2017 “modeling solar irradiance and solar pv power output to create a resource assessment using linear multiple multivariate regression” journal of applied meteorology and climatology volume 56doi: 10.1175/jamc-d-16-0175.1
- Wilks, D.S., “Statistical methods in the atmospheric sciences, in International Geophysics Series”, Dep. of Earth and Atmospheric Sciences. 2006, Academic Press: Cornell University. p. 179-548.
- Carroll, David Ruppert “The use and misuse of orthogonal regression in linear errors in variables models” the American statistician, feb 1996 vol. 50 no. 1
- Boggs, P. T., and J. E. Rogers (1990), “Orthogonal distance regression, in Statistical Analysis of Measurement Error Models and Applications”:Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference Held June 10–16, 1989, Contemp. Math., vol. 112, edited by P. J. Brownand W. A. Fuller, pp. 186–194, American Mathematical Society, Providence, doi:10.1090/conm/112.
- Kendall MG, Stuart A (1979) “The advanced theory of statistics”, vol 2, 4th edn. Charles Griffin Co. Ltd., London <https://sam.nrel.gov/weather-data.html>
- R. Deshmukh et al. “Geospatial and techno-economic analysis of wind and solar resources” Renewable Energy 134 (2019) 947e960
- P. Vanitha, P. Swamy Sharan, K. David “Solar Energy In India: An Overview” IJIRMPMS | Volume 7, Issue 5, 2019
- Bogdan S. Borowy, Ziyad M. Salameh, “Methodology for optimally sizing the combination of a battery bank and PV array in a Wind/PV

- hybrid system”, IEEE Transactions on Energy Conversion, vol. 11, No. 2, pp. 367-375, June 1996. (Article)
- Hlaluku W Mkasi, Manjunath Basappa Ayanna, Lawrence Pratt, Kittessa Roro “GLOBAL IRRADIANCE ON PHOTOVOLTAIC ARRAY” CSIR Energy Centre, Council for Scientific and Industrial Research
- Cheng Wu, Jian Zhen Yu “Evaluation of linear regression techniques for atmospheric applications: the importance of appropriate weighting” <https://doi.org/10.5194/amt-11-1233-2018> Atmos. Meas. Tech., 11, 1233–1250, 2018
- Ayers, G. P.: “Comment on regression analysis of air quality data”, Atmos. Environ., 35, 2423–2425, [https://doi.org/10.1016/S1352-2310\(00\)00527-6](https://doi.org/10.1016/S1352-2310(00)00527-6), 2001
- Ling Leng, Tianyi Zhang, Lawrence Kleinman, Wei Zhu “Ordinary Least Square Regression, Orthogonal Regression, Geometric Mean Regression and their Applications in Aerosol Science” Brookhaven National Laboratory
- Stromeyer, D., Gr’unthal, G. & Wahlstrom, R., 2004. “Chi-square regression for seismic strength parameter relations, and their uncertainties, with applications to an Mw based earthquake catalogue for central, northern and northwestern Europe”, J. Seism, 8(1), 143–153.
- A. Deniz M. S. Yucemen “Magnitude conversion problem for the Turkish earthquake data” Nat Hazards (2010) 55:333–352 DOI 10.1007/s11069-010-9531-8
- Castellaro S, Bormann P (2007) “Performance of different regression procedures on the magnitude conversion problem”. Bull Seismol Soc Am 97:1167–1175
- Andrew Bruce, Peter Bruce “Practical statistics for data scientists”<https://www.oreilly.com/library/view/practical-statistics-for/9781491952955/ch04.html>
- <https://support.minitab.com/en-us/minitab/19/help-and-how-to/modeling-statistics/regression>
- <http://www.curve-fitting.com/orthogonal.htm>

AUTHORS PROFILE



Mr. Surendra H.H. completed his Bachelor of Engineering from Bangalore University in 2000 and Master of Technology from Visvesvaraya Technological university in year 2004. He is currently pursuing Ph.D. and working as Assistant Professor in the Department of Electronics and

Communication, BMS College of Engineering since 2008. He has over 15 years of teaching experience and has published two conference papers and two international journal papers. His main research work focuses on design of renewable resource largely solar energy analysis and the statistical data of solar energy resources provides the platform to analyze the system using various tools. He has guided various under graduate and post-graduate projects.



Mr Seshachalam D has completed his Phd from MNNIT, Allahabad and currently working as Professor in the Department of Electronics and Communication, BMS College of Engineering since 2009. His area of interest is Control Systems,

Embedded System and Power electronics. He has over 25 years of teaching experience and 4 years in research experience. He is IEEE and ISTE member and is currently guiding five research scholars. He was head of the department for around 8 years. He has published papers in 3 international journals and 7 conferences. He is currently guiding 5 research scholars and guided 20 projects at master’s level.



Mr Sudhindra K.R. has completed his Phd from VTU, Belgaum and is currently working as Associate Professor in the Department of Electronics and Communication, BMS College of Engineering since 2014. His area of interest is in wireless communication and is a member of LMISTE and

AMIETE. He has 10 years of teaching experience and 6 years of industrial experience. He has published 6 national and 9 international journals. He is also guiding 4 research scholars. He is a subject expert for Nihon communications, Bangalore. He has two grants fetched under TEQIP II. He has also delivered technical talks in ZTE and Engineering colleges.

