ASEC 2024 Conference

The 5th International Electronic Conference on Applied Sciences

04–06 December 2024 | Online

MDPI

Bland-Altman Analysis of Open-Access Online Weather Data

Constance Uchechi Iwu, Yongjun Sang, Howard Okezie Njoku, Gang Pei

Sustainable Energy Engineering Research Group, Department of Mechanical Engineering, University of Nigeria, Nsukka 410001, Nigeria.
Department of Thermal Sciences and Energy Engineering, University of Science and Technology of China, Hefei 230027, China.

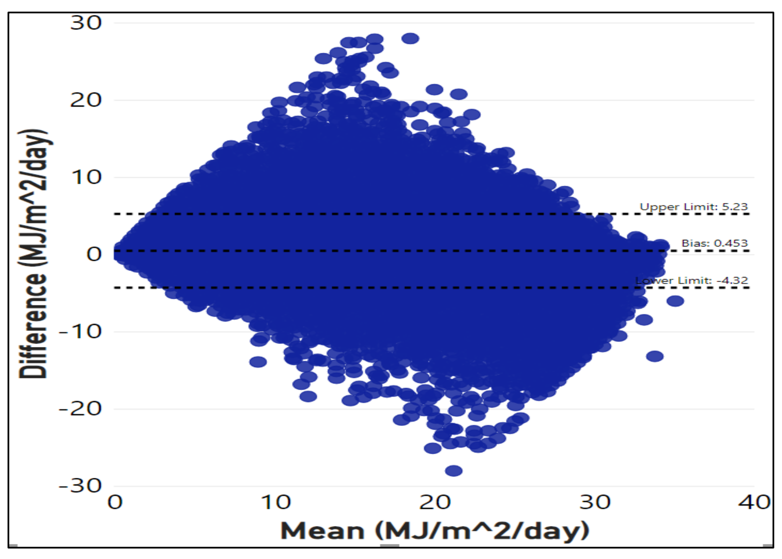
INTRODUCTION & AIM

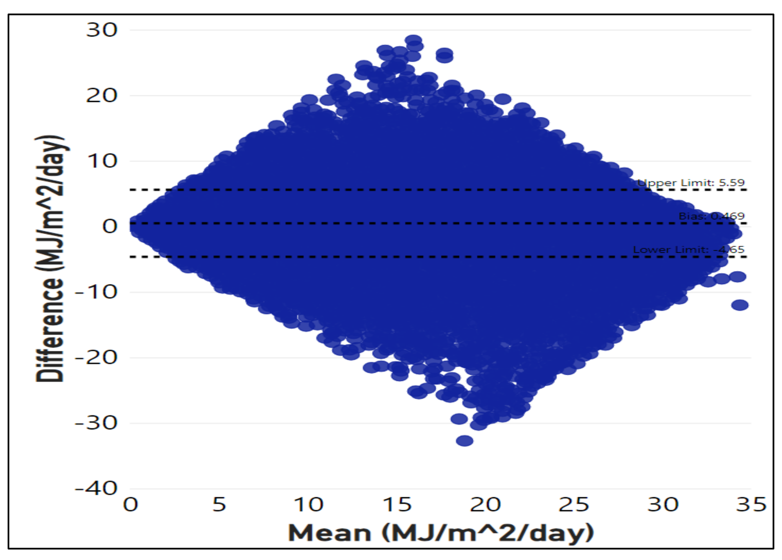
Solar radiation data is crucial for evaluating solar energy potential, designing predictive models, and optimizing solar energy systems. However, the high costs of ground-based weather stations have led to data gaps in many regions. Satellite-derived (SD) data, such as NASA's POWER and the European Commission's PVGIS, offer cost-effective alternatives. This study uses the Bland-Altman (BA) method to compare SD solar radiation data with ground-measured (GM) data from the World Radiation Data Centre (WRDC) to assess their levels of agreement. Unlike standard measures which only indicate overall fit or strength of association, the BA method directly evaluates the agreement by identifying systematic bias and defining the limits within which 95% of the differences lie. Despite its extensive application in medical research, the BA method remains underutilized in the engineering field, highlighting its novelty and potential value for solar energy studies. Additionally, the study investigates how latitude, seasonality, and diurnal variations influence data agreement, offering insights into the reliability of SD data across different contexts.

METHOD

The study utilized solar radiation data from 171 WRDC stations between 2005 and 2020, along with corresponding SD data from POWER and PVGIS. Data cleaning involved removing missing values and standardizing units. Bland-Altman analysis was applied to assess agreement, focusing on calculating the limits of agreement (LOA) to determine the range within which 95% of differences between SD and GM data are expected to lie. The concept of the "range of LOA" was introduced to quantify the extent of variability between the datasets, providing a direct measure of agreement, where narrower ranges indicate stronger agreement. Other statistical methods such as root mean square error (RMSE), correlation coefficient (R-value), and the paired T-value were employed to validate the BA findings. Additional analyses examined the effects of latitude, seasonal variations, and diurnal periods on data agreement.

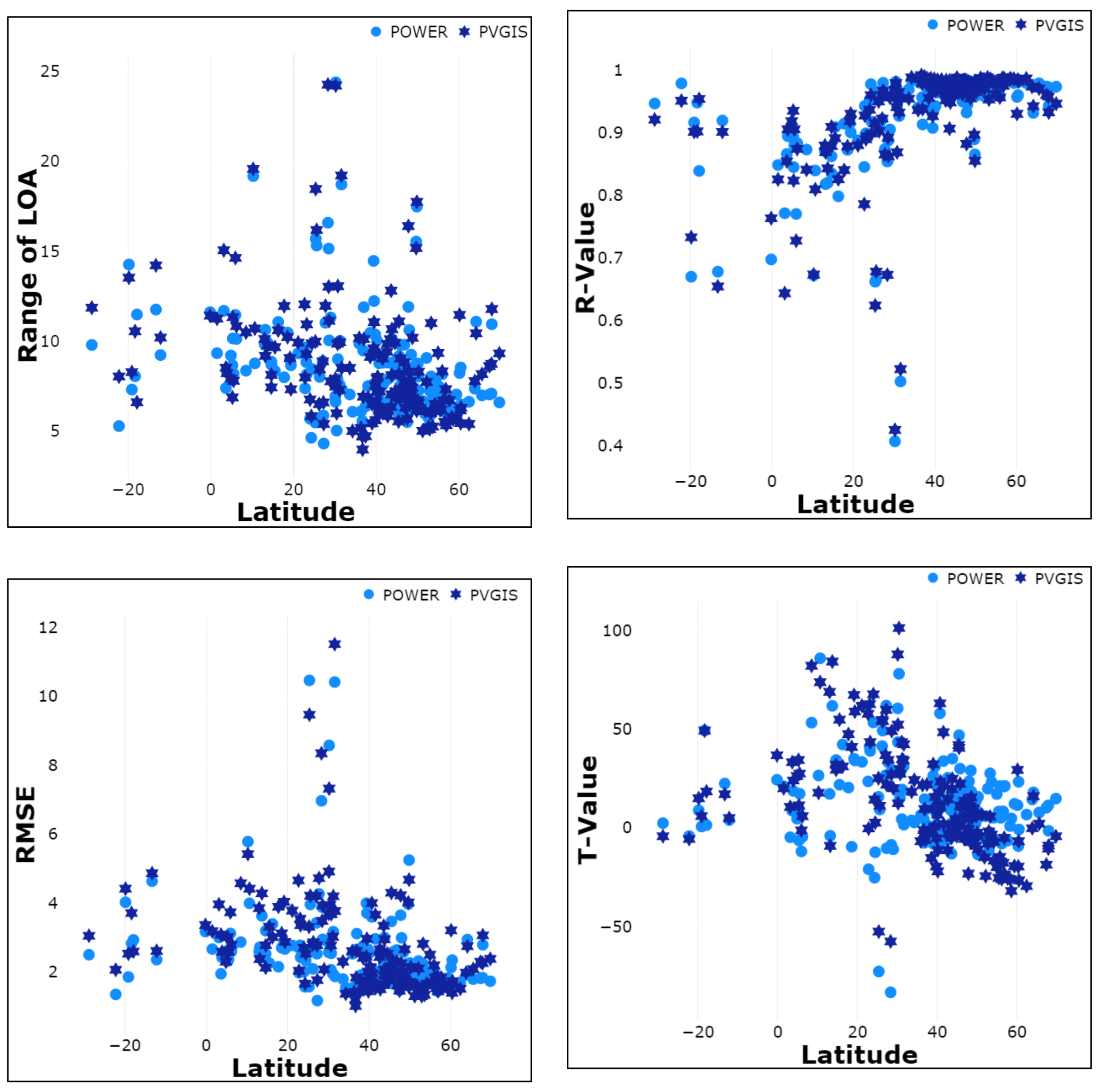
RESULTS & DISCUSSION



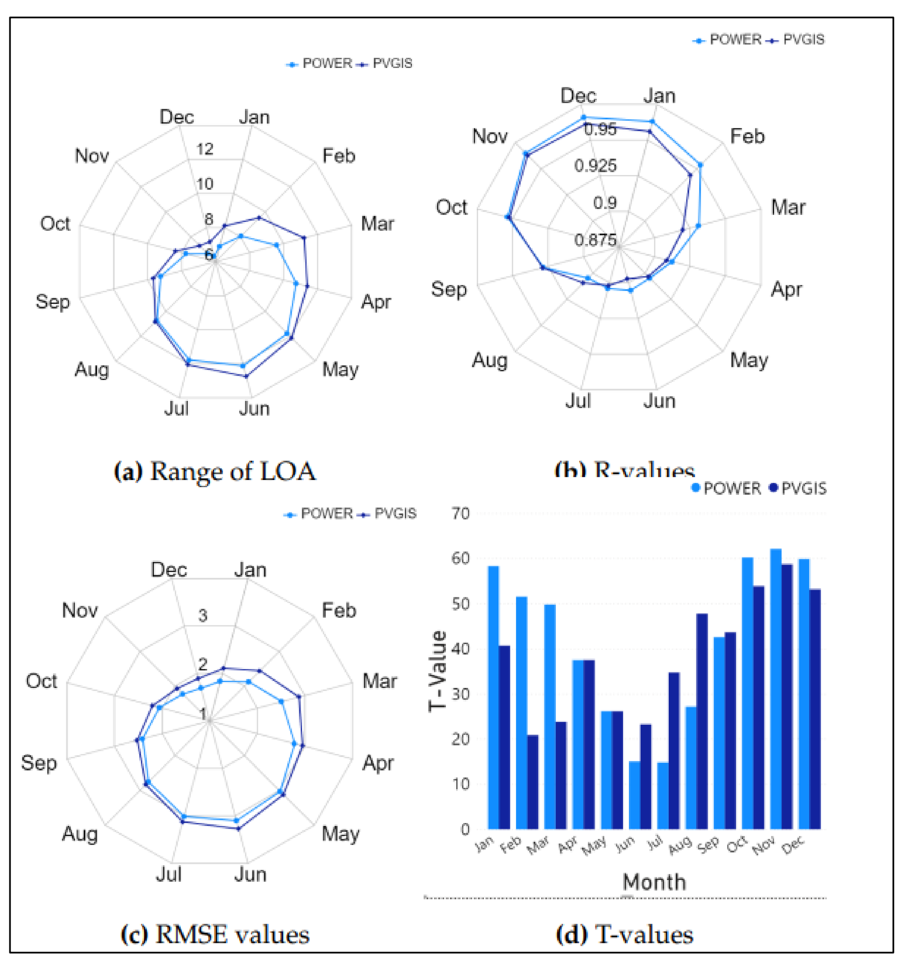


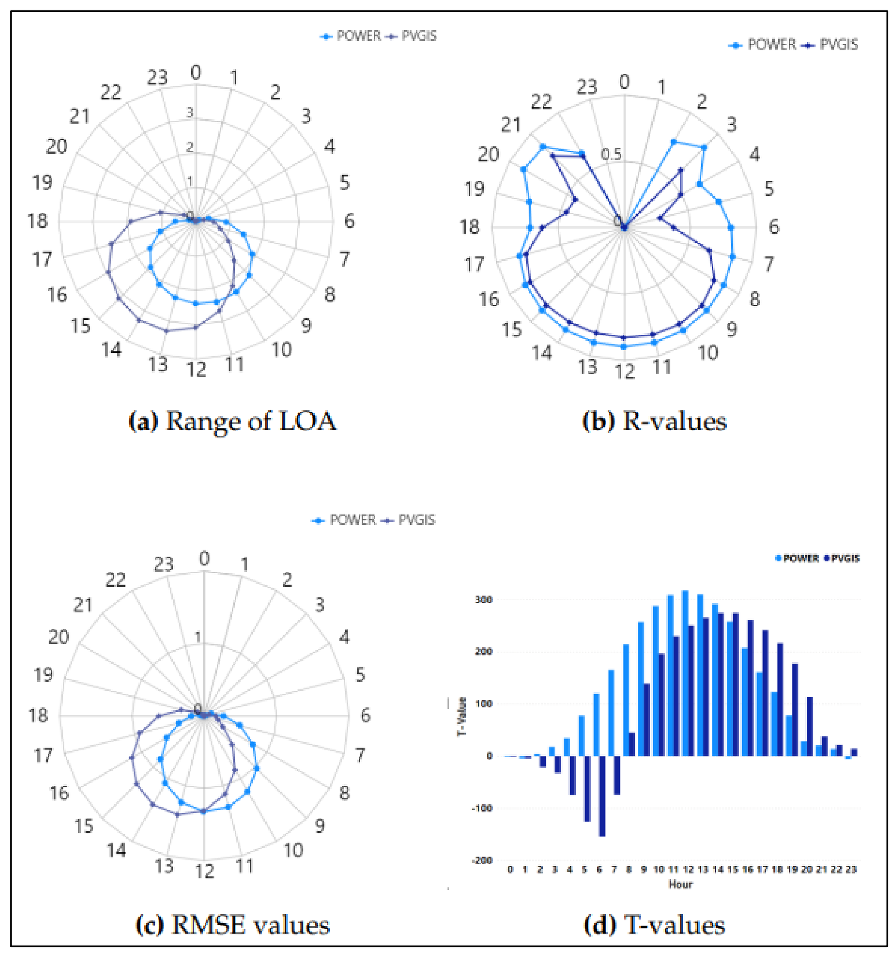
Satisticial measure	POWER SD Data Standard BA (% BA)	PVGIS SD Data Standard BA (% BA)
Bias (MJ/m²/day)	0.45 (7.81)	0.47 (2.69)
Std Dev. (MJ/m²/day)	2.44 (24.89)	2.61 (26.80)
Upper LOA limit (MJ/m²/day)	5.23 (56.59)	5.59 (55.21)
Lower LOA limit (MJ/m²/day)	-4.32 (-40.98)	-4.65 (-49.83)
LOA range (MJ/m²/day)	9.55 (97.57)	10.23 (105.04)
Percentage of data within LOA	94.25	94.38
RMSE (MJ/m²/day)	2.48	2.65
T-value	128.98	125.63
R-value	0.96	0.95

The results revealed significant agreement between SD and GM data, with the POWER dataset showing narrower LOA ranges (9.55 MJ/m²/day) compared to PVGIS (10.23 MJ/m²/day), indicating stronger agreement with GM data



The plots illustrate the influence of latitude on the agreement between SD and GM data. The LOA range plot shows no clear correlation with latitude, while the R-value plot indicates stronger correlations (values close to 1) at higher latitudes. Similarly, the RMSE plot shows a decrease in RMSE with increasing latitude, suggesting better compatibility between SD and WRDC data at elevated latitudes. However, the t-value plot does not reveal any discernible pattern relative to latitude.





The R, range, and RMSE plots all indicate that the SD data agreed better with the GM data from November to February, likely due to factors such as reduced cloud cover and precipitation affecting remote sensing accuracy during this period. However, the R plot shows a slightly different trend, though the difference between maximum and minimum R-values is minimal, less than 0.1.

The agreement between SD and GM data is strongest during early morning and evening hours, as indicated by lower RMSE, narrower LOA ranges, and high R-values. Midday hours (10 AM to 3 PM) show greater variability and reduced agreement, likely due to challenges in measuring or modeling peak solar radiation. Despite this, overall correlation remains consistently high throughout the day.

CONCLUSION

The Bland-Altman analysis demonstrated that NASA's POWER dataset aligns more closely with ground-measured data than PVGIS, particularly at higher latitudes and during certain seasons. The findings highlight the importance of considering latitude, seasonality, and diurnal variations when using satellite-derived data for solar energy applications.

FUTURE WORK / REFERENCES

- Expand the analysis to include other SD datasets (e.g., Solcast, Solargis) for broader comparison.
- Investigate additional factors like atmospheric conditions and sensor performance under extreme weather.