EuroSun 2026

Evaluation of Site Adaptation Techniques for Satellite-derived solar Irradiance Data in Hot Desert Climate

Summary

Satellite-derived solar radiation data provides a practical alternative for obtaining solar irradiance information, especially in regions where ground-based measurements are limited, require frequent maintenance, or lack long-term reliability. Unlike ground observations, satellite data offer extensive spatial and temporal coverage. However, these datasets may be affected by uncertainties caused by inaccurate atmospheric parameters—such as cloud cover and aerosols—differences in temporal resolution, and the fact that satellite models represent area-based pixels rather than specific point measurements. Consequently, using satellite data without proper calibration and correction against ground observations may result in significant errors. This study evaluates and compares several calibration methods for satellite-based solar irradiance using ground measurements. We utilize five years of minute-resolution ground data from 13 sites across Qatar, comparing traditional bias-correction approaches such as Empirical Quantile Mapping (EQM) to advanced machine learning models like Gradient Boosting Machines (GBM). Model performance for both global horizontal irradiance (GHI) and direct normal irradiance (DNI) is assessed using relative root mean square error (rRMSE) and relative mean bias error (rMBE). Our findings show that the GBM model achieves substantial error reduction for both GHI and DNI, with rRMSE improvements of approximately 37.04%, while EQM offers only marginal enhancements. This abstract presents preliminary results from a more extensive study utilizing additional machine learning techniques and includes a techno-economic analysis.

1. Introduction

The accelerated deployment of solar energy systems has intensified the need for precise and high-resolution solar radiation datasets to support applications including resource assessment, optimal site selection, system sizing, operational monitoring, and energy yield forecasting. This need is particularly pronounced in desert environments, where solar irradiance is subject to substantial temporal and spatial variability. Such variability is primarily driven by elevated aerosol concentrations, recurrent dust events, and highly dynamic atmospheric conditions [1]. Under these circumstances, even minor systematic deviations in irradiance estimation can propagate into significant errors in predicted energy output, highlighting the importance of region-specific calibration of solar radiation data.

In recent years, satellite-derived and reanalysis-based solar radiation products have emerged as essential tools for solar resource evaluation, especially in regions with limited ground-based observations [2, 3]. Datasets provided by platforms such as the Copernicus Atmosphere Monitoring Service (CAMS), HelioClim, and SolarGIS deliver consistent temporal resolution alongside extensive spatial coverage and long-term data availability. Despite these advantages, such products often exhibit systematic biases linked to challenges in accurately representing cloud dynamics, aerosol properties, and atmospheric processes [4, 5, 6, 7]. These limitations are particularly evident in arid and semi-arid regions, where localized climatic factors are not adequately captured by globally parameterized models [8]. Consequently, the direct application of satellite-based irradiance datasets may result in persistent overestimation or underestimation of surface solar radiation [6, 7]. To this end, we evaluate two well known site-adaptation methods namely Empirical Quantile method (EQM) and Gradient Boosting Machine (GBM), under the hot desert climate of Qatar.

2. Methodology

This study analyzes 1-minute GHI and DNI data from 13 sites across Qatar (2020–2024), with 30 km spacing and calibrated sensors. CAMS-Rad satellite data are evaluated using two bias-correction methods—EQM and GBM—trained on 2020–2022 and tested on 2023–2024.

EQM adjusts modeled irradiance by matching its empirical distribution to observations [9]. GBM is an ensemble method that sequentially fits decision trees to minimize residual errors [10]. Its ability to capture nonlinear relationships and iteratively reduce bias makes it effective for improving solar irradiance estimates.

3. Priliminary Results

Table 1 presents rRMSE, rMBE, and percentage improvements over the CAMS-Rad baseline for the two bias correction models evaluated during the 2023–2024 test period.

The CAMS-Rad baseline shows notable systematic underestimation, with rMBE values of -19.05% (DNI) and -6.1% (GHI). Error levels are also high, with rRMSE reaching 35.44% for DNI and 19.14% for GHI.

Both EQM and GBM reduce both bias and error. GBM performs best, achieving rRMSE improvements of 37.04% (GHI) and 29.94% (DNI).

Tab. 1: Overall Performance of CAMS-Rad components on the test split (SZA $\leq$ 70deg)

	GHI				DNI			
Method	rRMSE (%)	rMBE (%)	Impr. rRMSE (%)	Impr. rMBE (%)	RMSE (%)	MBE (%)	Impr. rRMSE (%)	Impr. rMBE (%)
CAMS-Rad (Baseline)	19.14	-6.1	-	-	35.44	-19.05	-	-
Empirical Quantile Mapping (EQM)	14.29	-1.18	25.34	80.66	26.94	-5.39	23.98	71.71
Gradient Boosting Machine (GBM)	12.05	**0.30**	**37.04**	**95.08**	24.83	-3.24	**29.94**	**82.99**

To evaluate spatial robustness, both the models were further analyzed across individual stations (Figure 1). Despite Qatar's relatively uniform terrain, notable differences exist between individual site. Stations such as Dukhan and Shahaniya exhibit consistently low residual errors, whereas Sudanthile shows persistently higher errors across models. This spatial variability suggests that localized factors, such as aerosol loading and coastal humidity, influence model accuracy.

These findings underscore that station-level contributions to error reduction are not uniform, which has important implications for targeted calibration.

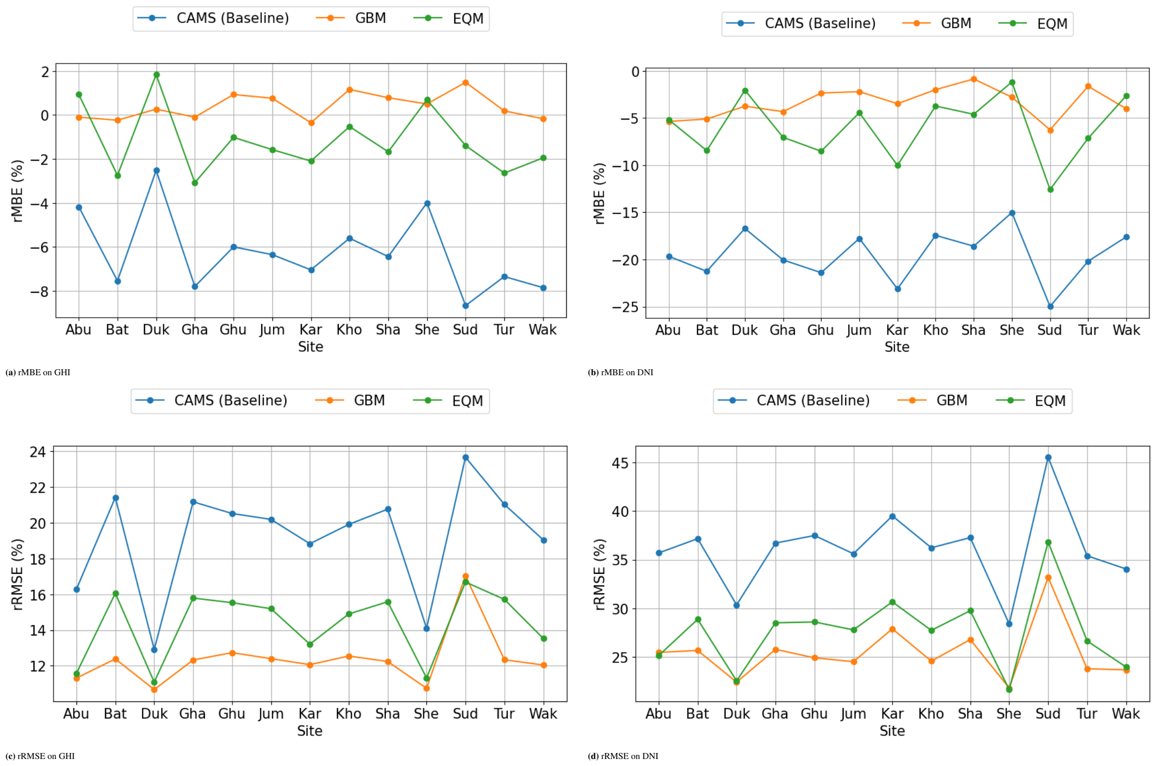

Fig. 1: Site-based comparison of CAMS solar radiation components.

References

[1] Christos Fountoukis, Harshvardhan Harshvardhan, Ivan Gladich, Luis Ackermann, and Mohammed A Ayoub. Anatomy of a severe dust storm in the Middle East: Impacts on aerosol optical properties and radiation budget. *Aerosol and Air Quality Research*, 20(1):155–165, 2020.

[2] Richard Perez, Richard Seals, and Joseph J Michalsky. All-weather model for sky luminance distribution—preliminary configuration and validation. *Solar Energy*, 50(3):235–245, 1993.

[3] ART Jonkman. Satellite solar radiation datasets for solar energy applications. *Solar Energy*, 130:1–7, 2016.

[4] Mathilde Marchand, Mireille Lefèvre, Laurent Saboret, Etienne Wey, and Lucien Wald. Verifying the spatial consistency of the CAMS Radiation Service and HelioClim-3 satellite-derived databases of solar radiation using a dense network of measuring stations: the case of The Netherlands. *Advances in Science and Research*, 16:103–111, 2019.

[5] Abdul Wahab Ziaullah, Dunia Bachour, Daniel Perez-Astudillo, Lionel Menard, and Philippe Blanc. Performance evaluation of cams reanalysis for improved solar resource assessment and forecasting in qatar's desert climate. In *EU PVSEC 2025: 42th EuropeanPhotovoltaic Solar Energy Conference*. WIP-Munich, 2025.

[6] Pierre Ineichen. Comparison of broadband solar radiation models. *Solar Energy*, 57(2):125–132, 1996.

[7] Richard Perez, Richard Seals, and André Zelenka. A new operational model for satellite-derived irradiances: Description and validation. *Solar Energy*, 73(5):307–317, 2002.

[8] Achraf Chakri, Nour-Eddine Laftouhi, Lahcen Zouhri, Hassan Ibouh, and Mounsif Ibnoussina. Assessment of satellite and reanalysis precipitation data using statistical and wavelet analysis in semi-arid, Morocco. *Water*, 17(11):1714, 2025.

[9] Douglas Maraun. Bias correction, quantile mapping, and downscaling: Revisiting the inflation problem. *Journal of Climate*, 28(17):6710–6722, 2015.

[10] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.