A Geographic Information System Mapping Method for Creating Improved Satellite Solar Radiation Dataset Over Qatar

Sachin Jain[a], Daniel Perez-Astudillo[a], Dunia A. Bachour[a], Antonio P. Sanfilippo[a]
[a] *Qatar Environment & Energy Research Institute, P.O.Box 5825, Doha, Qatar.*

***Abstract***

The future of solar energy in Qatar is evolving steadily. Hence, high-quality spatial solar radiation data is of the uttermost requirement for any planning and commissioning of solar technology. Generally, two types of solar radiation data are available: satellite data and ground observations. Satellite solar radiation data is developed by the physical and statistical model. Ground data is collected by solar radiation measurement stations. The ground data is of high quality. However, they are limited to distributed point locations with the high cost of installation and maintenance for the ground stations. On the other hand, satellite solar radiation data is continuous and available throughout geographical locations, but they are relatively less accurate than ground data. To utilize the advantage of both data, a product has been developed here which provides spatial continuity and higher accuracy than any of the data alone.

The popular satellite databases: National Solar radiation Data Base, NSRDB (PSM V3 model, spatial resolution: 4 km) is chosen here for merging with ground-measured solar radiation measurement in Qatar. The spatial distribution of ground solar radiation measurement stations is comprehensive in Qatar, with a network of 13 ground stations. The monthly average of the daily total Global Horizontal Irradiation (GHI) component from ground and satellite data is used for error analysis. The normalized root means square error (NRMSE) values of 3.31%, 6.53%, and 6.63% for October, November, and December 2019 were observed respectively when comparing in-situ and NSRDB data.

The method is based on the Empirical Bayesian Kriging Regression Prediction model available in ArcGIS, ESRI. The workflow of the algorithm is based on the combination of regression and kriging methods. A regression model (OLS, ordinary least square) is fitted between the ground and NSBRD data points. A semi-variogram is fitted into the experimental semi-variogram obtained from the residuals. The kriging residuals obtained after fitting the semi-variogram model were added to NSRBD data predicted values obtained from the regression model to obtain the final predicted values. The NRMSE values obtained after merging are respectively 1.84%, 1.28%, and 1.81% for October, November, and December 2019. One more explanatory variable, that is the ground elevation, has been incorporated in the regression and kriging methods to reduce the error and to provide higher spatial resolution (30 m). The final GHI maps have been created after merging, and NRMSE values of 1.24%, 1.28%, and 1.28% have been observed for October, November, and December 2019, respectively. The proposed merging method has proven as a highly accurate method. An additional method is also proposed here to generate calibrated maps by using regression and kriging model and further to use the calibrated model to generate solar radiation maps from the explanatory variable only when not enough historical ground data is available for long-term analysis. The NRMSE values obtained after the comparison of the calibrated maps with ground data are 5.60% and 5.31% for November and December 2019 month respectively.

***Keywords:*** Global Horizontal Irradiation, Geographic Information System (GIS), Empirical Bayesian Kriging (EBK) Regression Prediction, Regression Kriging (RK), NSRDB.

## I. Introduction

Solar radiation resource assessment is an essential part of site suitability analysis before the installation of the solar power plants. However, there are other several factors too such as availability of land, sunshine duration, relative humidity, wind speed, and temperature data. But here our focus is mainly on the most prioritized factor which is solar radiation [1]. Solar radiation, often called the solar resource, refers to all the electromagnetic radiation that the sun emits. The amount of solar radiation reaching the earth's surface varies according to Geographical location, time of the day, season, and local weather. Instruments such as pyranometers and pyrheliometers are used to measure solar irradiance at the ground. Various satellite data models (physical and statistical) are also available that estimate the solar irradiance received at the earth's surface, such as the National Solar radiation Data Base, NSRDB (Physical Solar

Model, PSM V3 model, spatial resolution: 4 km, hourly temporal resolution) and NSRDB (SUNY Semi-Empirical model, spatial resolution: 10 km, hourly temporal resolution) [2]. To develop a suitable solar radiation database reference for site suitability it is recommended that data should represent 10-11 years (at least one solar cycle) of historical data with high accuracy [3]. The ground station measurements are highly accurate and available at the high temporal resolution, but it is observed at the discrete point location, and available ground data may not be enough for long-term analysis. Hence, satellite solar radiation data plays an important role, and it can be utilized for the development of highly accurate maps with the combination of satellite and ground data for a particular region [4]-[6] and further can be used in calibration process to develop solar radiation maps of historical data, where historical ground data is not available for long term analysis. Here, regression and kriging (RK) algorithms are used for the aforesaid process. Several studies had been carried out on RK to explore its potential in spatial analysis and interpolation [7]-[8]. The regression and kriging algorithm is a combination of both regression and kriging approaches: the regression model is fitted between the dependent and explanatory variables to obtain the prediction. The kriging model with the expected value zero is used to fit the residual. Hence, the final prediction will be more accurate than either regression or the kriging algorithm alone [9]. RK and its different versions with advanced modification can be implemented using commercial and open-source available tools such as ArcGIS and QGIS respectively. Here, Empirical Bayesian Kriging (EBK) Regression Prediction method [10]-[11] has been implemented using ArcGIS to create the combined product from ground measurement and NSRDB satellite data and RK has been implemented using open source RK model available in pykrige package of python for calibration purpose [12].

## II. Data

Global horizontal irradiation (GHI) component of solar radiation from ground measurements for 10-13 Solar radiation resource assessment stations. The ground stations are distributed over the Qatar region as shown in Fig. 1. Monthly average of daily total GHI values has been used here for experiment purposes.

GHI component from NSRDB (Physical Solar Model, PSM V3 model, spatial resolution: 4 km, hourly temporal resolution) solar radiation database has been accessed for merging and calibration process. NSRDB database has hourly temporal resolution and 4 km spatial resolution. The monthly average of the daily total has been calculated from hourly values. The distributed grid point over the Qatar region is shown in Fig. 2.

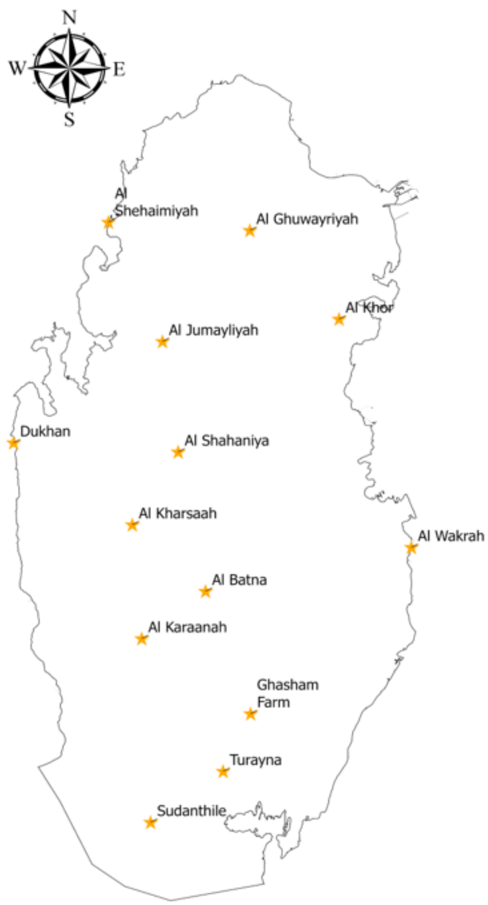



*Fig. 1 Solar radiation resource assessment station*

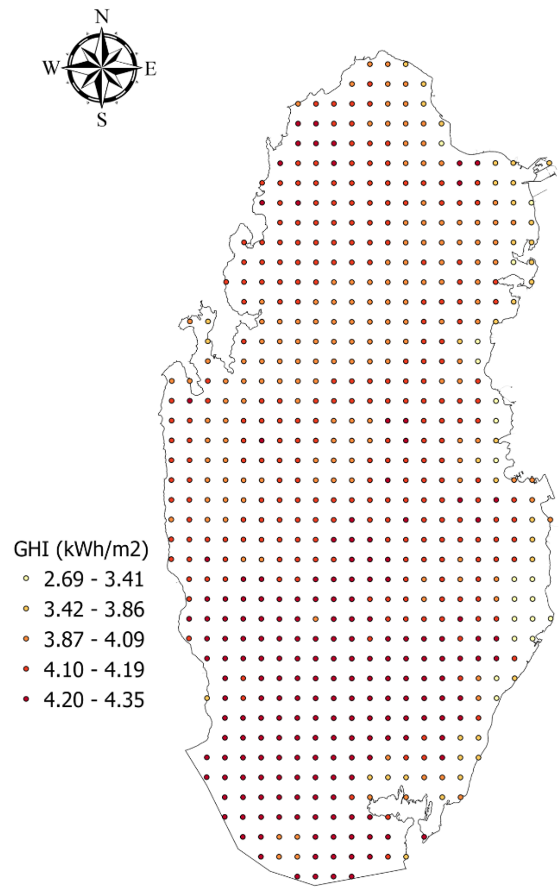



*Fig. 2 Nov 2019, grid point for GHI, NSRDB*

The GHI data from satellite data and ground measurements of 3 months i.e., October 2019, November 2019, and December 2019 has been used here for the development of merged product and calibration of the RK model.

## III. Method

### a. Merged Product

The EBK regression prediction [13] tool from ArcGIS, ESRI toolsets has been used here for the development of the merged product. The input parameter has been fed into the model. Model run over the ground and satellite data and stored the result in raster format. The working algorithm is based on the regression and kriging approach. The principal component analysis (PCA) has been carried out on explanatory variables before any regression. Apart from NSRDB data, the digital surface model (30 m spatial resolution) from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) global digital elevation data has been used here as an explanatory variable. Ordinary least square regression has been carried out between the dependent and explanatory variables (PCA). The Model divides the data into local subsets and apply the algorithm to different local subset further, these local RK models are mixed to produce the final prediction map. The model computes many semi-variograms by doing default numbers of simulation from the data in the subset and further utilizes those semi-variograms for each prediction location. Fig.3 shows the subset created by the algorithm and Fig. 4 shows the simulation carried out at al shahaniya solar radiation station. The fact that the models are computed locally gives EBK Regression Prediction a significant advantage over other regression kriging models. This enables the model to adapt to diverse situations and take into consideration local influences.

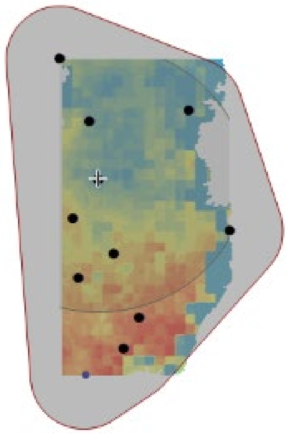


*Fig.3 Subset polygon, reference from ArcGIS*

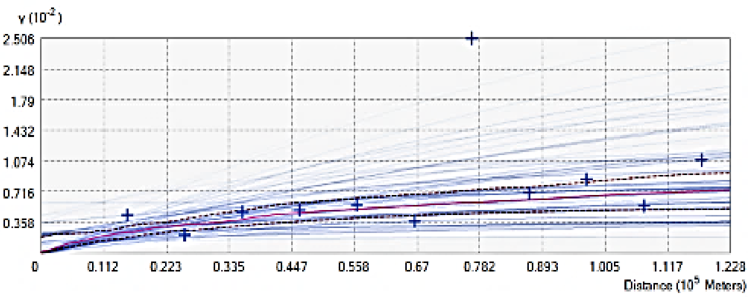



*Fig. 4 Simulation at Al Shahaniya station, semi variance at y axis and distance at x axis reference from ArcGIS*

b. Calibrated product

The regression and kriging (RK) model from the pykrige package of python is used here for calibration purposes. RK model is calibrated by the inclusion of satellite and ground data. A linear regression model is first fitted between the ground and NSBRD data points. The residuals obtained from the regression model are interpolated based on the semi-variogram model, and further interpolated residual values were added to NSRBD predicted data obtained from the fitted regression model, to obtain the final predicted values. RK model is first fitted over the October 2109 data and further prediction has been carried out on November, and December 2019 NSRDB data to obtain the calibrated map for November, and December 2019.

The major difference between the commercially available tool EBK regression prediction and the open-source RK model is in terms of semi-variogram calculation. For RK model prediction semi variogram is calculated from the stable model automatically once and fitted throughout the data and used further for interpolation. The variation in values between sample points with close distances usually tends to be minimal. Hence, the semi-variance is minimal as shown in Fig. 5. However, the likelihood of similarity decreases as sample point distances increase. As a result, the semi-variance increases significantly. As the distance increases away from the sample points, there is no longer a relationship between the sample points and their variance starts to flatten out. There is no local subset creation for multiple simulations as carried out in EBK regression prediction.

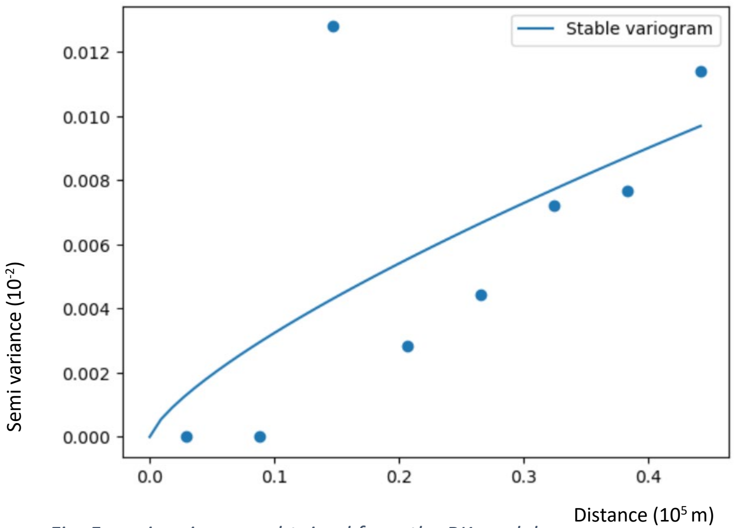



*Fig. 5 semi variogram obtained from the RK model.*

## IV. Results

After comparing in-situ and NSRDB data, the normalized root means square error (NRMSE) values for October, November, and December 2019 were found to be 3.31%, 6.53%, and 6.63%, respectively as shown in Fig. 6, Fig. 7, and Fig. 8.

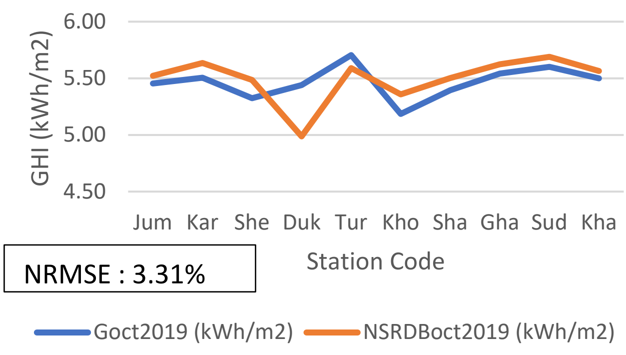



*Fig. 6 Ground vs NSRDB for October 2019*

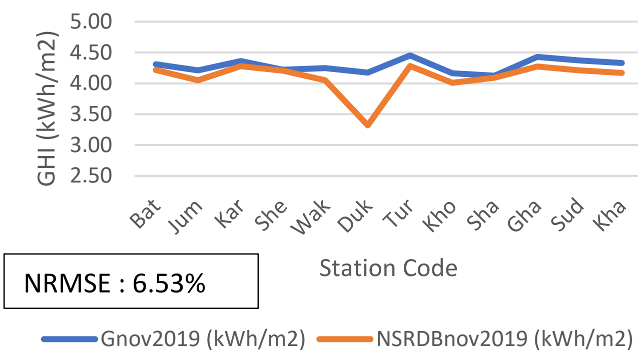



*Fig. 7 Ground vs NSRDB for November 2019*

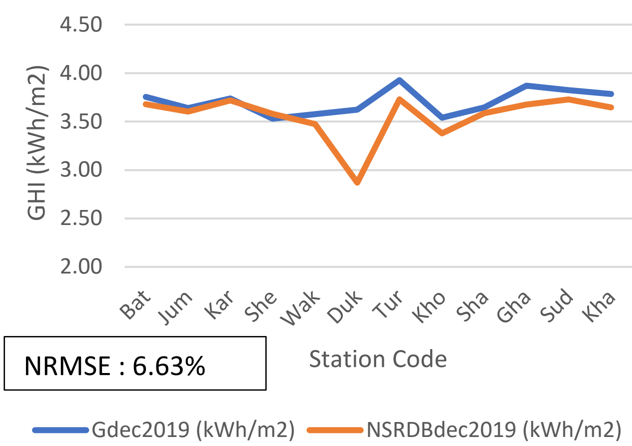
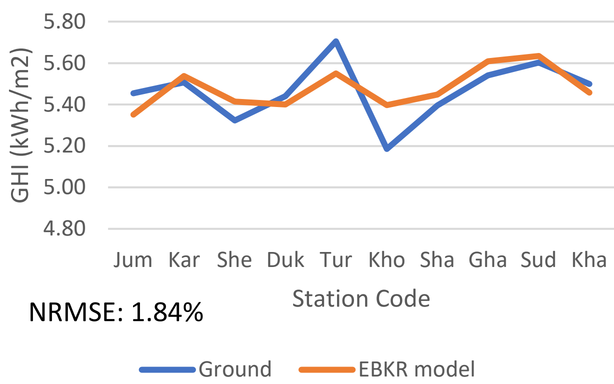
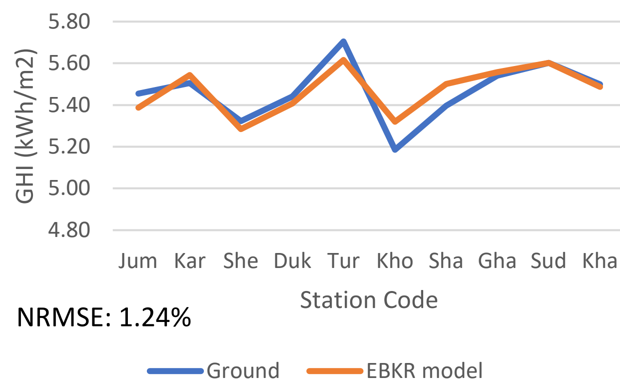
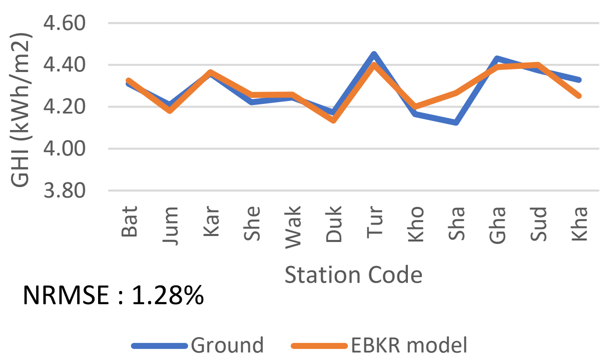
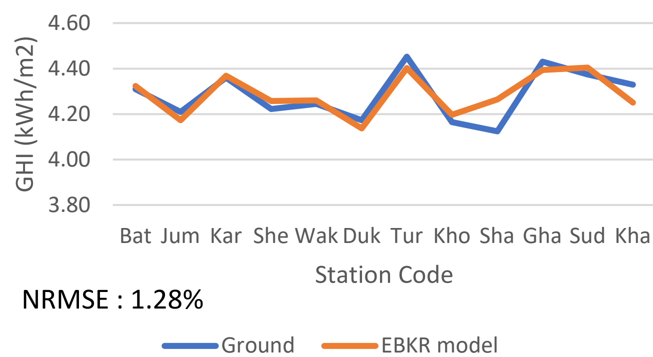
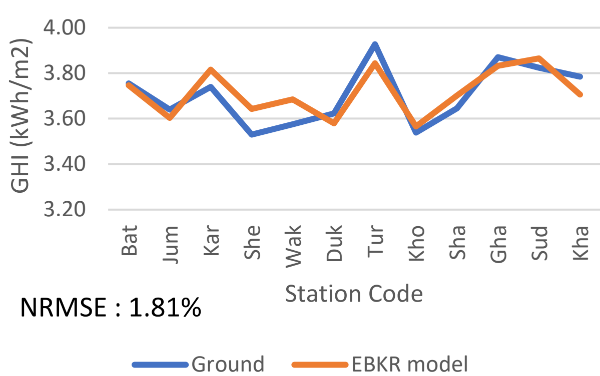
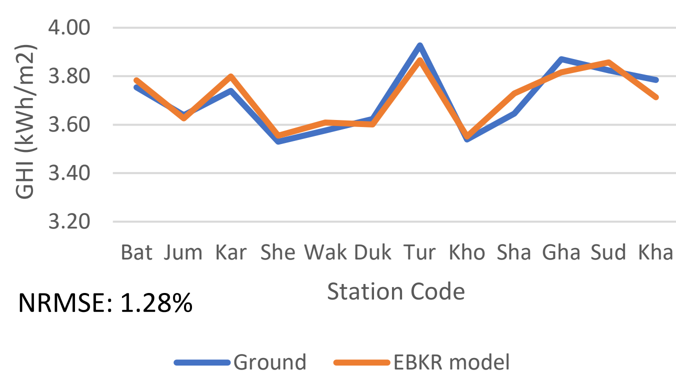



*Fig. 8 Ground vs NSRDB for December 2019*

a. Merged product.

The leave one out cross-validation error (LOOCV) analysis has been carried out for the merged product and ground data, and NRMSE values of 1.84%, 1.28%, and 1.81% has been observed for October, November, and December 2019, respectively as shown in Fig. 9, Fig.10 and Fig.11. These results are comparatively lower than the errors obtained before applying the merging method. One more explanatory variable, that is the ground elevation, has been incorporated to reduce the error and to provide high-resolution (30 m) spatial continuous data. The final GHI maps has been created with 30 m spatial resolution, and error analysis has been carried out for which NRMSE values of 1.24%, 1.28%, and 1.28% has been observed for October, November, and December 2019, respectively as shown in Fig. 12, Fig.13 and Fig.14. Table 1. Represent the LOOCV error analysis for merged product and merged product with additional variable.

Table 1. LOOCV Error analysis for merged product and merged product with additional variable

| Merged Product | Merged Product with additional variable. |
| --- | --- |
| <br>*Fig. 9 Ground vs EBKR model, October 2019* | <br>*Fig. 12 Ground vs EBKR model, October 2019* |
| <br>*Fig. 10 Ground vs EBKR model, November 2019* | <br>*Fig. 13 Ground vs EBKR model, November 2019* |
| <br>*Fig. 11 Ground vs EBKR model, December 2019* | <br>*Fig. 14 Ground vs EBKR model, December 2019* |

The GHI maps for October, November, and December 2019 are shown in Fig. 15,16, and 17 respectively.

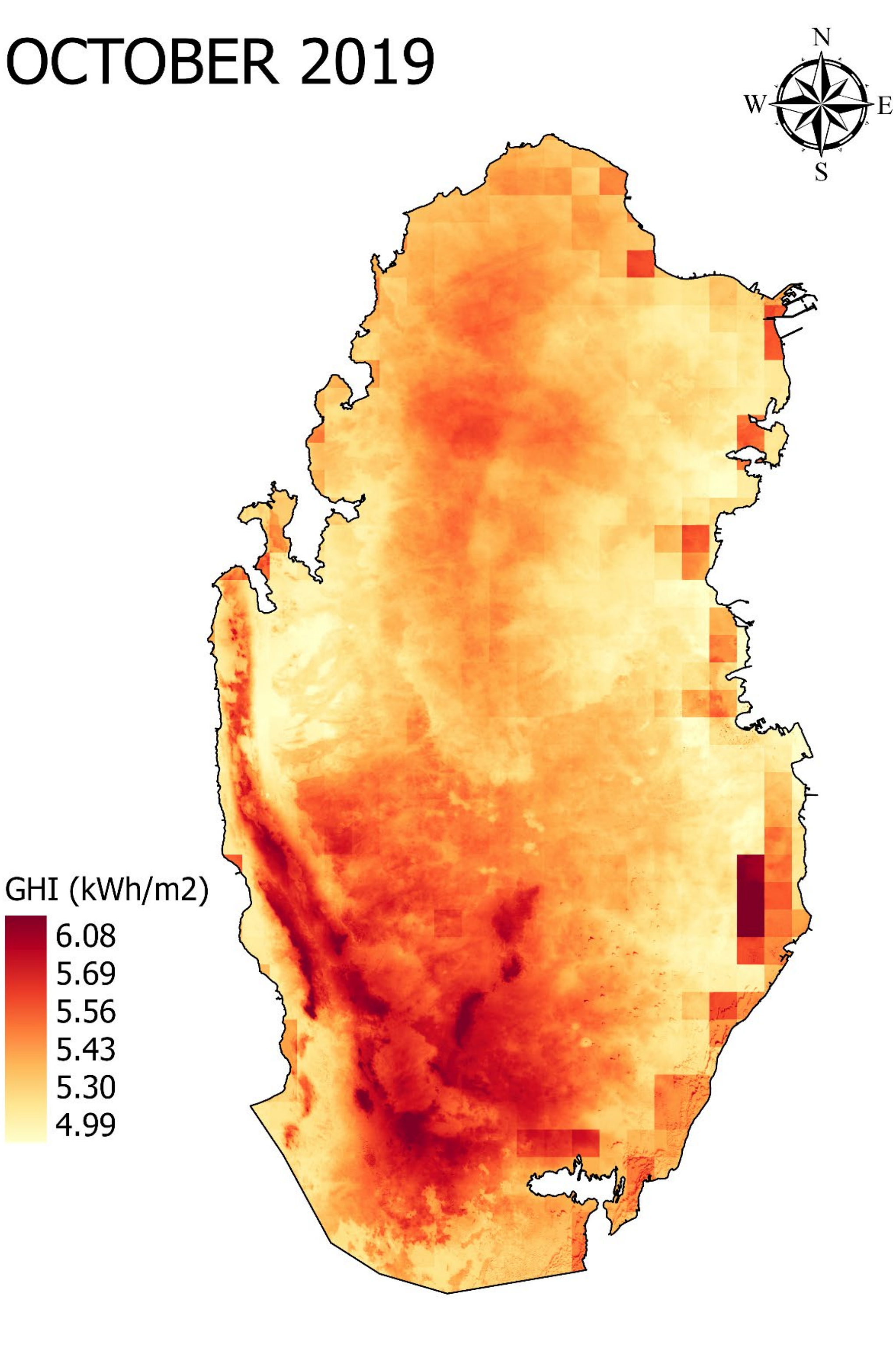



*Fig. 15 GHI map for the month of October (30 m resolution), generated from the calibrated EBKR model.*

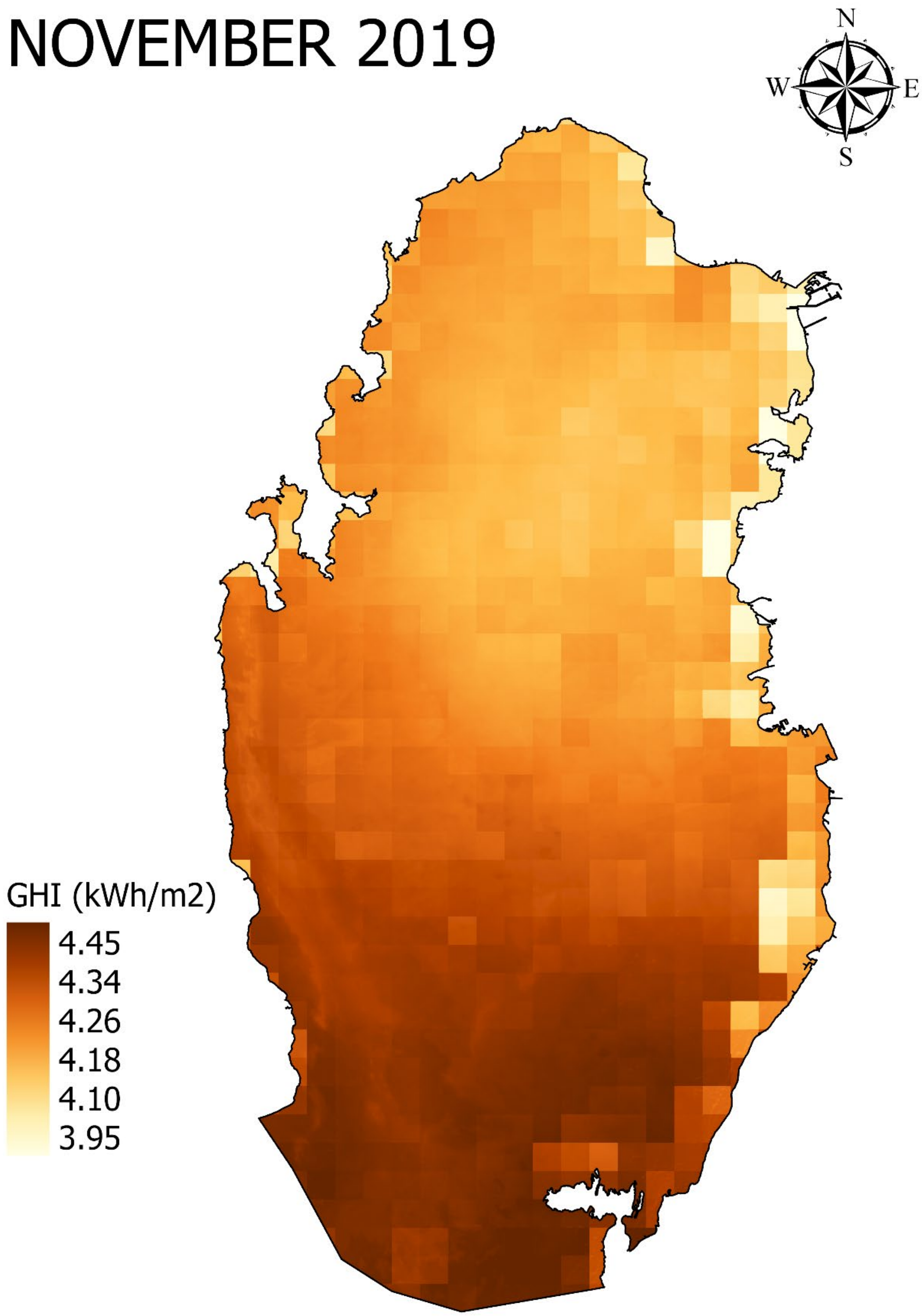



*Fig. 16 GHI map for the month of November (30 m resolution), generated from the calibrated EBKR model.*

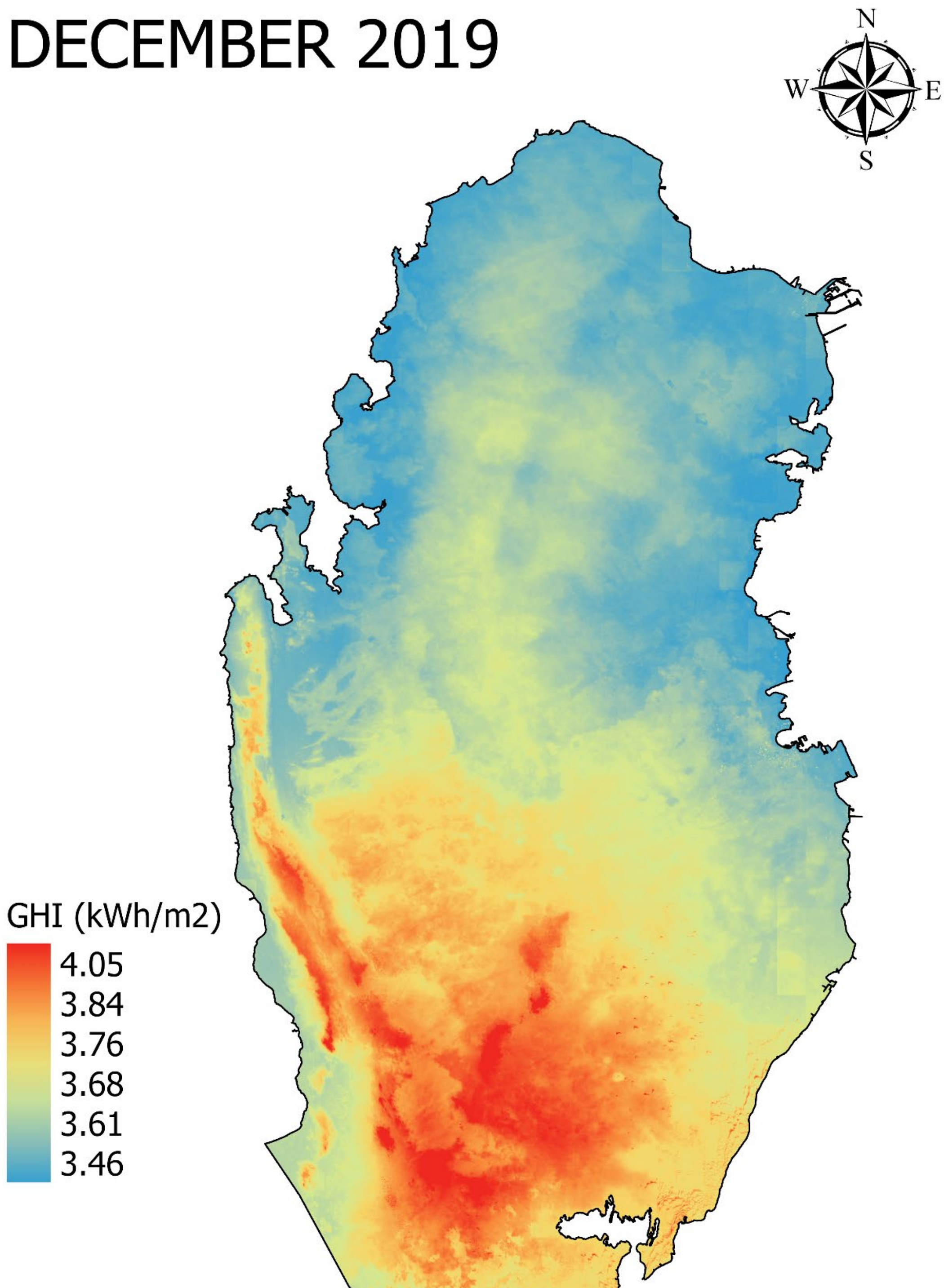



*Fig. 17 GHI map for the month of December (30 m resolution), generated from the calibrated EBKR model.*

b. Calibrated product

The error analysis for the calibrated product and ground data has been carried out and NRMSE value of 5.60% and 5.63% has been observed for November and December 2019 respectively. Fig. 18, and 19 shows the error analysis.

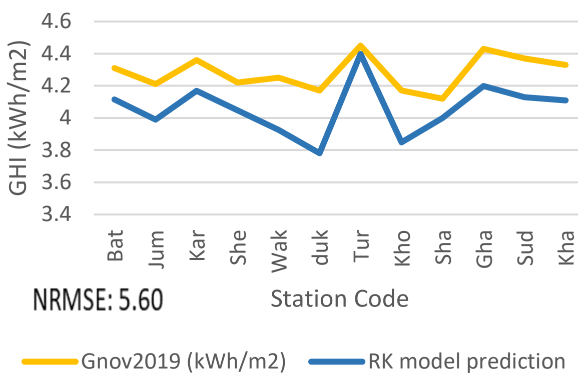



*Fig. 18 Ground vs RK model prediction for November 2019*

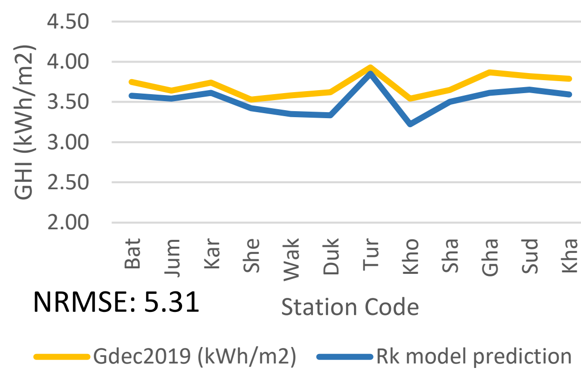



*Fig. 19 Ground vs RK model prediction for December 2019*

The GHI maps for November and December 2019 are shown in Fig. 20 and 21 respectively.

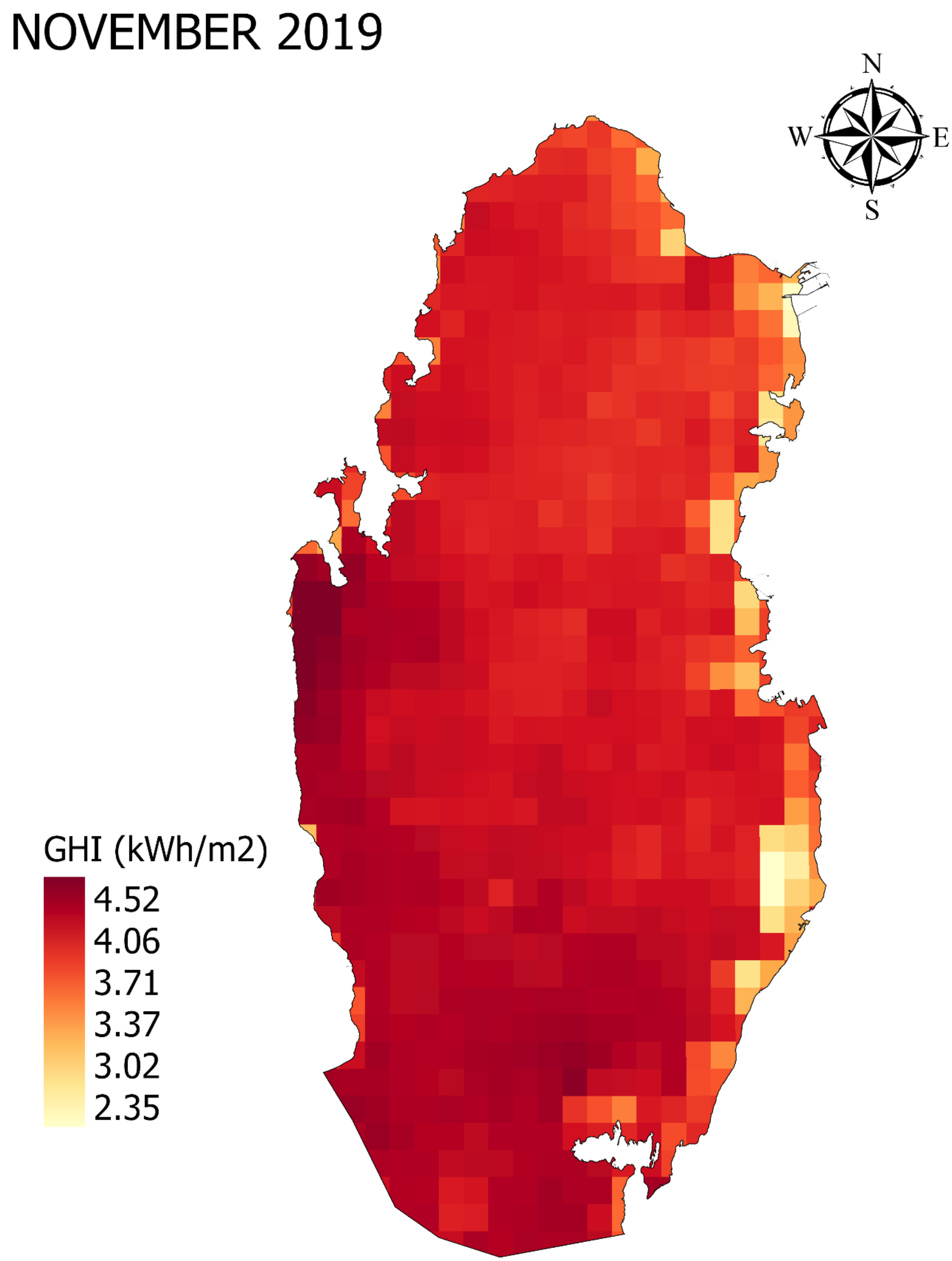



*Fig. 20 GHI map for the month of November (4 km resolution), generated from the calibrated RK model.*

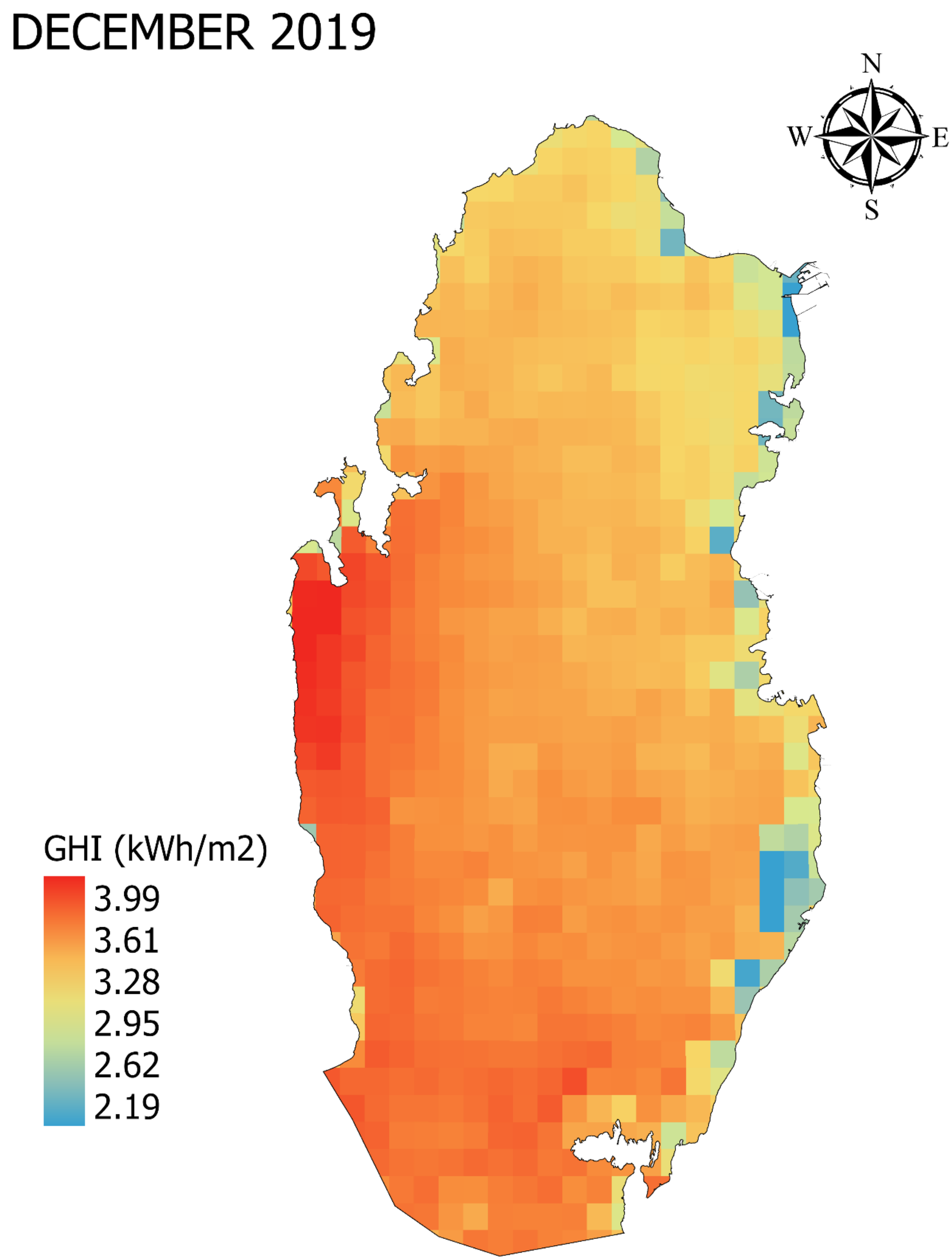



*Fig. 21 GHI map for the month of December (4 km resolution), generated from the calibrated RK model.*

## V. Conclusion

Solar radiation resource assessment is essential and plays an important role at different stages of the installation and commissioning of the solar power plant. Hence, it is important to have an accurate historical solar radiation database for reference. Various satellite and ground data is available to fulfill this need but with their limitation. It is beneficial to have a merged product as proposed in this paper when both data are available because of high spatial resolution and high accuracy. However, as solar resource assessment require long-term historical data and it may not be possible to have long-term ground data, in that case the open-source RK model can be used for calibration and further can be used to generate historical data from explanatory variable only. The results presented here are considered preliminary, as it is only based on 3 months of data.

**References:**


[1] Mahmoud A. Hassaan, Ahmed Hassan, Hassan Al-Dashti, GIS-based suitability analysis for siting solar power plants in Kuwait, The Egyptian Journal of Remote Sensing and Space Science, Volume 24, Issue 3, Part 1, 2021, Pages 453-461, ISSN 1110-9823, https://doi.org/10.1016/j.ejrs.2020.11.004

[2] National Solar radiation Data Base, NSRDB. (Online). https://nsrdb.nrel.gov/data-sets/international-data

[3] Solar resource database: Validation of Solargis solar radiation model Report 170-11/2019. (Online). Available: https://solargis.com/docs/accuracy-and-comparisons/overview

[4] Chua, Z.-W., Kuleshov, Y., Watkins, A.B., Choy, S., Sun, C. A, Comparison of Various Correction and Blending Techniques for Creating an Improved Satellite-Gauge Rainfall Dataset over Australia. Remote Sens. 2022, 14, 261.https://doi.org/10.3390/rs14020261

[5] Kirubel Mekonnen, Assefa M. Melesse, Tekalegn Ayele Woldesenbet. Merging satellite rainfall estimates and daily rain gauge observations for improved flood simulation in Melka Kuntire catchment, upper Awash Basin, Ethiopia, Remote Sensing Applications: Society and Environment, Volume 25, 2022,100701, ISSN 2352-9385, https://doi.org/10.1016/j.rsase.2022.100701.

[6] Brinckmann, S., Klameth, A., and Trentmann, J.: Merging of satellite and ground measurements of hourly surface solar radiation variables in Germany, EMS Annual Meeting 2021, online, 6–10 Sep 2021, EMS2021-128, https://doi.org/10.5194/ems2021-128, 2021.

[7] Tomislav Hengl, Gerard B.M. Heuvelink, David G. Rossiter, About regression-kriging: From equations to case studies, Computers & Geosciences, Volume 33, Issue 10, 2007, Pages 1301-1315, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2007.05.001.

[8] Elijah Akwarandu Njoku, Patrick Etim Akpan, Augustine Edet Effiong, Isaac Oluwatosin Babatunde, The effects of station density in geostatistical prediction of air temperatures in Sweden: A comparison of two interpolation techniques, Resources, Environment and Sustainability, Volume 11, 2023, 100092, ISSN 2666-9161, https://doi.org/10.1016/j.resenv.2022.100092

[9] Ge, Y., Thomasson, J.A., Sui, R. et al. Regression-kriging for characterizing soils with remote sensing data. Front. Earth Sci. 5, 239–244 (2011). https://doi.org/10.1007/s11707-011-0174-1

[10] Krivoruchko K. and Gribov A. (2019). "Evaluation of empirical Bayesian kriging," Spatial Statistics Volume 32. https://doi.org/10.1016/j.spasta.2019.100368

[11] Pilz, J., and G. Spöck (2007). "Why Do We Need and How Should We Implement Bayesian Kriging Methods," Stochastic Environmental Research and Risk Assessment 22 (5):621–632.

[12] Murphy, B.: PyKrige: development of a kriging toolkit for Python, in: AGU fall meeting abstracts, 2014.

[13] Krivoruchko K. (2012). "Empirical Bayesian Kriging," ArcUser Fall 2012