

**Validating Modelled, Satellite based, and Reanalysis based Global
Horizontal Irradiance data sets in South African sites.**

by

Brighton Mabasa

A dissertation submitted in accordance with the requirements for the
degree of Master of Science in Physics
at the University of South Africa.

Supervisor : Prof. Sabata. J. Moloi (University of South Africa)
Co-supervisor : Dr. Meena. D. Lysko (Move Beyond Consulting (Pty)
Ltd).

December 2021

Declaration

I declare that “Validating Modelled, Satellite based, and Reanalysis based Global Horizontal Irradiance data sets in South African sites” is my own work and that all sources that I have used or quoted have been indicated and acknowledged by means of references.

I further declare that none of this material has been submitted for any other degree at University of South Africa (Unisa) for another qualification or at any other higher education institution.

Mr. B. Mabasa:

B.Mabasa

Date:

12 May 2022

Abstract

The dissertation presents and compares the alternative methods that can be used to quantify high quality global horizontal irradiance (GHI) in areas with sparse in-situ radiometric measurements like South Africa. This was done by evaluating the performance of six GHI clear sky models, namely: Berger–Duffie, Simple Solis, Haurwitz, Ineichen–Perez, McClear and Bird by comparing 1-minute model outputs to relative clear sky reference 1-minute GHI in-situ data for sites in South Africa. This was also accomplished by calibrating and validating the annual regression coefficients, **a** and **b**, of the Ångström–Prescott (AP) model, which can be used to approximate daily GHI from in-situ sunshine hours. The validation of two hourly reanalysis-derived GHI datasets namely: Modern-Era Retrospective Analysis for Research and Applications (MERRA2), and Fifth generation European Center for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) and three satellite-derived GHI datasets namely Satellite Application Facility on Climate Monitoring (CMSAF SARA), Copernicus Atmosphere Monitoring Service (CAMS) and SOLCAST for sites in South Africa was also done. Graphs, tables and statistical metrics namely relative Root Mean Square Error (rRMSE), relative Mean Bias Error (rMBE), relative Mean Absolute Error (rMAE) and the coefficient of determination (R^2), were used to assess the capability of the datasets in estimating hourly GHI. The results showed that the Bird clear sky model was the overall best for sites in South Africa, with rRMSE of 4 %, rMBE of 2 % and R^2 of 0.998. The AP model results for sites in South Africa were as follows: maximum rMBE was 4 %, rMAE was 9 %, rRMSE was 11 % and the worst correlation was 0.91. The hourly satellite and reanalysis-based results were as follows: A very good correlation for the satellite-derived and a poor correlation for the reanalysis-derived datasets. There was a consistent overestimation of GHI by both satellite and reanalysis-derived datasets. MERRA2 reanalysis-derived was the overall worst performing while SOLCAST satellite-derived was overall best performing. When considering only the freely

available datasets, CAMS and CMSAF performed better. The AP model's performance was similar to CMSAF and CAMS daily estimates. The results suggest that: The Bird clear sky model can be used with quantitative assurance for different applications in South African sites. The AP model coefficients determined can be used to precisely approximate daily GHI data at sites in South Africa where there is in-situ sunshine duration data. CMSAF and CAMS satellite-derived datasets are feasible freely available alternative GHI in South African sites.

Keywords: Solar radiation; Global horizontal irradiance; sunshine duration; Clear sky model; Satellite; Reanalysis; Climatological regions; Ångström–Prescott model; Solar Radiation Data (SoDa); Satellite Application Facility on Climate Monitoring (CM SAF); Fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis(ERA5); SOLCAST; Copernicus Atmosphere Monitoring Service (CAMS); Modern-Era Retrospective Analysis for Research and Applications (MERRA2); Validation; Statistical metrics.

Acknowledgements

- I would like to thank my supervisor Dr. Meena D. Lysko for assisting me with the Recognition of Prior Learning (RPL) applications at University of South Africa (UNISA).
- I would like to thank my supervisors Dr. Meena D. Lysko and Prof. Sabata J. Moloi for guidance, support, making time throughout the course of the study, providing constructive criticism that helped me think outside the box and for supervising and co-authoring the publications generated from the dissertation.
- I would like to thank the South African Weather Services (SAWS) for funding the study and University of South African (UNISA) for giving me the opportunity to study.
- I would like to acknowledge the inputs of the anonymous reviewers, academic editors for the inputs in that helped in shaping the published articles and the co-authors to the published papers Dr. Henerica Tazvinga (SAWS), Ms. Nosipho Zwane (SAWS) and Dr. Sophie Mulaudzi (UNIVEN) for their valuable inputs.
- I would like to thank Prof. Albert A. Lysko for guidance with the coding for the colormaps.
- I would like to thank Mr. Gift Tshifhiwa Rambuwani for transferring Linux skills to extract gridded data sets.
- I would like to thank my partner and my family for encouraging me throughout the study.
- I also would like to thank the South African Weather Services (SAWS) for providing the Global Horizontal Irradiance (GHI) data used as a reference for satellite and reanalysis-derived data validation and for providing auxiliary meteorological parameters: temperature, humidity, sunshine duration and pressure data used in the study; the Copernicus Atmosphere Monitoring Service (CAMS) Solar Radiation Data (SoDa) for freely providing the CAMS hourly GHI data, Linke Turbidity data, ozone data, aerosol optical depth data at 550 nm (AOD_{550}) and at 1240 nm (AOD_{1240}) and McClear model data used in the study on

their website; European Center for Medium-Range Weather Forecasts (ECMWF) for providing freely available ERA5 hourly cloud data and ERA5 hourly radiation on their website; the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) for providing freely available surface Satellite Application Facility on Climate Monitoring (CMSAF SARA) and surface albedo data on their website, and the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) for freely providing hourly Modern-Era Retrospective Analysis for Research and Applications (MERRA2) GHI data; National Aeronautics and Space Administration (NASA) Prediction Of Worldwide Energy Resources (POWER) for providing freely daily GHI data on their website.

Table of Contents

Declaration.....	ii
Abstract.....	iii
Acknowledgements	v
Chapter 1	1
1.1. Introduction and background.....	1
1.2. Global horizontal Irradiance (GHI) Clear Sky Models.....	2
1.2.1. Global horizontal Irradiance (GHI) Clear Sky Models Literature Review.....	3
1.3. The Ångström–Prescott (AP) model	4
1.3.1. The Ångström–Prescott (AP) model Literature Review	5
1.4. Satellite and Reanalysis-derived Global Horizontal Irradiance Datasets.....	6
1.4.1. Satellite and Reanalysis-derived Global Horizontal Irradiance Datasets Literature Review	6
1.5. Problem statement (s)	8
1.6. Research Aims and Objectives.....	9
1.7. Dissertation outline	10
References	11
Chapter 2: Theoretical background.....	17
2.1. The Sun	17
2.2. Extra-terrestrial Solar Irradiance.....	19
2.3. The Solar Radiation Spectrum	20
2.4. The Sun-Earth Geometry	25
2.5. Global Mean radiative balance between Earth-Atmosphere.....	30
2.6. Solar Radiometry	32
2.6.1. Measurement of Global Horizontal Irradiance (GHI)	32
2.6.2. Measurement of Diffuse Horizontal Irradiance (DIF).....	34
2.6.3. Structure of a Pyranometer	34
2.6.4. Measurement of Direct Normal Irradiance (DNI)	38
2.7. Data Quality Control of Data.....	44
2.8. Methods to generate GHI datasets	45
2.8.1. Satellite-derived datasets.....	45
2.8.2. Reanalysis-based datasets.....	47
References	48
Chapter 3: Study Area and observation data.....	54
3.1. Study Area	54
3.2. Pre-Processing of Observation Data	57

3.3. Performance Evaluation Criteria (Statistical Metrics).....	59
References	61
Chapter 4: Calibration and Validation of Clear sky Models Six climatic zones in South Africa	63
4.1. Introduction.....	63
4.2. Description of the Six Models in this study.....	66
4.2.1. McClear Clear Sky Model.....	66
4.2.2. Berger–Duffie Clear Sky Model	67
4.2.3. Haurwitz Clear Sky Model.....	67
4.2.4. Ineichen–Perez Clear Sky Model.....	68
4.2.5. Bird Clear Sky Model	69
4.2.6. Simple Solis Clear Sky Model	70
4.3. Study Area	72
4.3.1. Data	73
4.3.1.1. Observation Data	74
4.3.1.2. Surface Albedo Data	74
4.3.1.3. Cloud Data	74
4.3.1.4. Atmospheric Ozone.....	75
4.3.1.5. Linke Turbidity	75
4.3.1.6. Aerosol Optical Depth	75
4.3.1.7. Water Vapor.....	77
4.3.1.8. Solar Geometry	78
4.3.2. Methodology	78
4.3.3. Statistical Metrics.....	79
4.4 Results and Discussions	80
4.4.1. Clear Sky Detection	80
4.4.2. Validation Results	83
4.4.3. Validation and Comparison of the Results with other Literature.....	87
4.5. Supplementary Materials	91
4.5.1. Statistics of model’s performance per site	91
4.5.2. Scatter plots of measured and modelled clear sky global horizontal irradiance.....	95
4.6. Chapter Summary	112
References	112
Chapter 5: The Ångström-Prescott Regression Coefficients for Six Climatic Zones in South Africa.....	117
5.1. Introduction.....	117
5.2. Study Area	121

5.3. Data	121
5.5. Results	124
5.5.1. Annual Ångström–Prescott (AP) Results.....	124
5.5.2. Validation Results	126
5.6. Chapter Summary	130
References	131
CHAPTER 6: Validating Hourly Satellite Based and Reanalysis Based Global Horizontal Irradiance Datasets over South Africa.....	133
6.1. Introduction.....	133
6.2. Study Area	136
6.3.1. Reanalysis-based Datasets	136
6.3.2. Satellite-based Datasets	137
6.4. Methodology	141
6.4.1. Pre-Processing of Observation Data and Validation Process	143
6.4.2. Statistical Metrics.....	143
6.4.3. Most Feasible Gridded Dataset.....	143
6.5. Results	143
6.5.1. CAMS Results	143
6.5.2. CMSAF Results	146
6.5.3. SOLCAST Results.....	147
6.5.4. ERA5 Results.....	148
6.5.5. MERRA2 Results	149
6.6. Discussion.....	167
6.6.1. CAMS.....	167
6.6.2. CMSAF	168
6.6.3. SOLCAST	168
6.6.4. ERA5	169
6.6.5. MERRA2	170
6.7. Chapter Summary	170
References	172
Chapter 7: Comparison between satellite-based datasets and Ångström–Prescott (AP) model in estimating daily GHI estimates in South African locations	175
7.1. Introduction.....	175
7.2. Study Area	176
7.3. Data	176
7.3.1. NASA SSE data	176

7.4. Methodology	177
7.4.1. Statistical Metrics.....	177
7.4.2. Most Feasible Gridded Dataset.....	177
7.5. Results	177
7.5.1. CAMS.....	177
7.5.2. CMSAF	179
7.5.3. SOLCAST	180
7.5.4. Ångström–Prescott (AP) model	181
7.5.5. NASA SSE.....	182
7.6. Chapter Summary	182
References.....	184
Chapter 8. Conclusion	185
8.1. Problem Statement 1.....	185
8.1.1. Aims and Objectives 1	185
8.1.2. Meeting the Aims and Objectives 1	185
8.2. Problem Statement 2.....	186
8.2.1. Aims and Objectives 2	186
8.2.2. Meeting the Aims and Objectives 2	186
8.3. Problem Statement 3.....	187
8.3.1 Aims and Objectives 3	187
8.3.2 Meeting the Aims and Objectives 3	187
8.4. Problem Statement 4.....	188
8.4.1. Aims and Objectives 4	188
8.4.2. Meeting the Aims and Objectives 4	188
8.5. Study limitations	189
8.6. Future studies.....	190
APPENDIX.....	191
List of Acronyms and Abbreviations	191
Python scripts.....	194
Data Analysis Script	194
1 minute to 15-minute averages	194
15-minute average to 1 hour average	195
Hourly averages to Daily averages	195
Clear Sky Model configuration.....	196
Merging Clear Sky Model Output.....	198
Clear Sky Detection	198

Calculating the theoretical sunshine duration (N) an input to the Ångström–Prescott (AP) model	200
Calculating the ratio between GHI and GHI at the top of the atmosphere ($Y = GH_{IHTOA}$) and ($X = nN$) the ratio between in-situ sunshine duration (n) and the theoretical sunshine duration (N) that will be used to calculate a and b	200
Calculating and displaying the Ångström–Prescott (AP) model coefficients a and b	201
Statistical Metrics.....	202
Plotting (Data Visulisation).....	203
Correlation Graphs.....	203
Line Graph	204
Solar Geometry	206
BSRN data quality control of GHI data.....	207
Visual Analysis of BSRN quality controlled GHI data.....	208

Chapter 1

In this chapter an introduction and background to Clear Sky, Ångström–Prescott (AP), Satellite-derived, and Reanalysis-derived datasets models are introduced. The models are used to quantify/approximate the amount of global horizontal irradiance (GHI) on the ground surface based on the clouds in the atmosphere. Literature review regarding each methodology, the gaps from each literature review section formulated the research questions, aims and the objectives of the dissertation presented in this Chapter.

1.1. Introduction and background

Solar radiation is the electromagnetic radiation emitted from the Sun because of the nuclear fusion (i.e., the fusion of four hydrogen atoms to make helium nucleus) inside the Sun [1]. GHI is the sum of diffuse horizontal irradiance (DIF), which is because of the dispersing of radiation by the atmospheric components and direct normal irradiance (DNI), which is the radiation that reaches directly on the ground surface without being affected by the atmospheric components [2]. Accurate GHI information is of great value for the economic and technical assessment of solar technologies, engineering designs, [3–11] and in the development, calibration, and validation of pragmatic models [6]. GHI is also of great value in environmental studies, climate change studies, agricultural sciences, ecology, and meteorology [6,7,9,10,12,13]. Therefore, acquiring verifiable GHI at a site is essential.

Clear sky GHI data, that is GHI collected under the cloudless atmosphere using clear sky models are also of great significance for several applications for example calculating the potential turnover of solar technologies and in solar radiation forecasting [14–22], cloudiness index calculation [15,16,18,20] and data quality control [16–20,23].

GHI measurements taken from in-situ solar radiation stations using a pyranometer are the most precise method to gather GHI information. Nevertheless, GHI recording sites are few, and far between, costly to establish, maintain and calibrate. For that reason, GHI in-situ measured data is in short supply [6,8–13,24-27]. To make up for this, dependable measurements taken from a scattered network are required to develop, calibrate, and validate pragmatic models that can be used to approximate and forecast the exploitable solar energy at other sites [28]. GHI data can also be created by interpolation of recorded in situ GHI [29]. The disadvantage of this technique is the uncertainties in generated data. Reanalysis and satellite-derived GHI datasets can provide dependable alternate GHI datasets and make up for the unavailability of in-situ radiometric sites by increasing the quantity of GHI data. The study also validates six GHI clear sky models to obtain proof of their reliability. The density of available clear sky irradiance in South African locations would, therefore, be increased.

1.2. Global horizontal Irradiance (GHI) Clear Sky Models

Clear sky models approximate the quantity of GHI that reaches the ground surface under uncloud atmospheres. Clear sky irradiance data can only be generated by testing various clear sky models at various locations or by taking only the values recorded by a pyranometer in the absence of clouds in the atmosphere. There is no universal clear sky model or sensors to measure clear sky irradiance information. Several models have been developed, calibrated, improved, and validated to approximate irradiance under clear skies. Even though there are modern techniques to generate GHI data, in-situ observation remains the most precise method to gather data. However, in-situ GHI measurements are inadequate because it is expensive to install, maintain and calibrate the solar radiation sensors. The transmittance of the atmosphere is complex [30], making it complicated to precisely model clear sky irradiance. The constant change nature of the atmosphere results in errors when the estimated clear sky GHI is compared

to data recorded using a reference in-situ GHI. To quantify the errors clear sky models must be validated.

1.2.1. Global horizontal Irradiance (GHI) Clear Sky Models Literature Review

Five GHI clear sky models were validated in Algeria, Bourareah and Ghardaia sites [16]. In this study, the European Solar Radiation Atlas (ESRA) and Ineichen–Perez models were the best performing models. Seventy-five GHI clear sky irradiance models were validated against data from seventy-five stations, various models were found to perform differently in each region [17]. Nine clear sky models were validated using 1-minute data from 14 stations in Australia, ESRA clear sky model was the best model [18].

ESRA clear sky model was the best model in Uruguay [23]. Antonanzas-Torres et al., 2019 [20] investigated seventy clear sky models against data from stations in France, the models performed differently from the two stations. Badescu et al., 2013 [21] investigated the precision of fifty-four models against data from three stations in Romania. No model showed an excellent performance, but some models had better performance than other models. The authors found that the precision of the models is influenced by the inputs and location from which the inputs were sourced. Mikofski et al., 2017 [22] investigated whether using in-situ water vapor and atmospheric optical depth as inputs to the models instead of reanalysis-derived information influence the precision of estimated irradiance. The results showed that using reanalysis-derived information does not affect the model performance. McClear model was investigated against in-situ data from eleven stations in different climates [31]. The model performance was reasonable but varied with climates. Ineichen (2006) [32] investigated eight models in twenty sites using inputs from sixteen different databases. The author found that linke turbidity had the major significance on model precision, and that not using in-situ measured inputs results in the underestimation of GHI by the models.

The main obstacles in the validation of models, according to [14] are the absence of high-resolution meteorological input parameters and the challenge in the detection of cloudless periods. In this study, measured meteorological parameters (pressure, humidity, and temperature) have been used as model inputs, specialised databases were accessed to get input parameters that were not measured from stations, albedo data from satellite was used. Clear sky periods were detected by using cloud cover information ERA5 [33]. This will enable more clear sky times to be identified. In addition to ERA5 data, Reno and Hansen (2016) [34] have been applied to identify some additional outliers.

In South Africa, the assessment was performed by earlier researchers [35–37]. The authors used a shorter reference in-situ GHI data to assess the performance of the models, they also used few stations that did not cover all climatic zones of South Africa, they also validated few models using a few numbers of clear sky days. Javu et al., 2019 [37] stressed out that it was insufficient to make thorough investigation of the performance of the models. In this study the performance of the models is investigated, using more sites and more clear sky days than in the previous studies, this allows models to be fully assessed. The choice of the models was based on the input parameters availability, executability with Python software and their existence in the studies to allow the comparison with other validation results.

1.3. The Ångström–Prescott (AP) model

The AP model approximates daily GHI using daily top of the atmosphere GHI radiation (GHI_{TOA}), theoretical day length (N), in situ sunshine duration (n), and AP model coefficients a and b . The model was initially suggested by Ångström [38] in 1924 up to then Prescott [39] improved it in 1940 by putting on GHI_{TOA} to take the place of cloudless GHI. The primary AP coefficients were $b = 0.75$ and $a = 0.25$; these were considered using meteorological

information from Stockholm [40]. The AP coefficients are location-dependent; therefore, they must be calibrated at sites where they will be applied to approximate GHI [9,13,40,41].

1.3.1. The Ångström–Prescott (AP) model Literature Review

The AP coefficients with verified precision in one climatic zone should not be assumed to be valid in the other climatic zone without further proof [41]. Researchers for instance from India, China, Spain, and Brazil [9–13] calibrated AP regression coefficients to their own climatic zones. The researchers concentrated on sites in the same climatic zones. In relation to the results by Zhang et al. [10], the AP regression coefficients were not the same in various climate regions. The variations in AP regression coefficients in different climatic zones were suspected to be due to differences in ground albedo, latitude, water vapor and aerosols [41]. The study by Tsung et al. [42] concentrated on one location using GHI and sun duration in-situ gathered from two different sites due to the unavailability of both sunshine duration and GHI from the same site. In this study, eight sites located in all six climatic zones of South Africa, and where both GHI and sunshine duration in-situ data were gathered from the same site, are considered so that the specific AP coefficients are indicative of a climatic region in the country and not the entire country.

Based on the finding from five independent research groups [9,10,12,41,43], sunshine-based models provided precise GHI approximations when compared to temperature and cloud-based models. This was suspected to be due to the close interrelation between GHI and sunshine duration reaching the ground's surface [10]. The availability of dependable in-situ sunshine duration data and GHI in all climatic zones of the study area and the excellent performance of sunshine-based models in the literature motivated the focus on sunshine-based models in this study. The AP linear regression model was chosen because of its simplicity and executability using Python software.

In South Africa, two research to assess the AP coefficients were accomplished by [44] and [45]. The obstacle, as per [45], was the unavailability of reference that covers all the climatic zones to calibrate and validate the AP coefficients. In this study, the data that covers all the regions of the study was available.

1.4. Satellite and Reanalysis-derived Global Horizontal Irradiance Datasets

Satellite and reanalysis-derived datasets are produced by using different methods or equations and different input meteorological parameters from different places are used as inputs to the equations. As a result, different performance in different areas by different datasets is probable. There is a need to validate different datasets in various stations/sites.

1.4.1. Satellite and Reanalysis-derived Global Horizontal Irradiance Datasets Literature Review

The study by Bright [46] investigated hourly SOLCAST satellite-derived GHI data, against forty-eight sites. SOLCAST showed excellent performance. The author stressed that a relative study against satellite and reanalysis-derived datasets is necessary to benchmark SOLCAST's performance. Yang and Bright [47] investigated six satellites and two reanalysis-derived datasets against data from fifty-seven stations over all continents. They found that the reanalysis-derived datasets performed poorly when compared to the satellite-derived datasets. They overestimated GHI in most locations. The authors established that each dataset considered performed better at a certain site, so the evaluation of various datasets at various locations was highlighted. It was also indicated that SOLCAST, which is not freely available dataset, performed better overall but it did not outperform the freely available datasets over all stations.

Merchand et al. [48] investigated the CAMS satellite-derived dataset against in-situ data from sixteen sites from the Netherlands. They established that CAMS satellite-derived datasets could precisely approximate GHI. The small errors reported were suspected to be because of the biases in the aerosol data used as an input to the McClear model [31], an input to the Heliosat-4 method [49].

Thomas et al. [50] assessed CAMS satellite-derived against in-situ data from forty-two sites in Brazil. Prior to their investigation, CAMS satellite-derived datasets were only validated in North Africa and Europe. The errors increased with an increase in humidity and there was an overall overestimation of GHI.

Ameen et al. [51] investigated CAMS satellite-derived dataset in Iraq, against in-situ data from nine sites. The author found that CAMS satellite-derived dataset could approximate the trends of the in-situ GHI data. Assessment of CAMS satellite-derived dataset in other regions was emphasized.

Marchand et al. [24] investigated CAMS satellite-derived dataset against in situ GHI from five sites in Morocco. The author found that CAMS satellite-derived dataset slightly overestimated GHI, but it was capable of approximating GHI.

Trollet et al. [52] assessed reanalysis-derived (ERA5 and MERRA2) and satellite-derived datasets (CMSAF SARA and CAMS) against GHI data in the tropical Atlantic Ocean [53]. The author found that satellite-derived datasets had an excellent performance while poor performance by reanalysis-derived datasets were reported.

1.5. Problem statement (s)

1. Which clear sky model performs better in estimating minute global horizontal irradiance (GHI) in each location/station under study and in each macro-climatological region of South Africa? Which GHI clear sky model is better overall in estimating minute clear sky GHI in South Africa?
2. Are the Ångström–Prescott (AP) coefficients which are used as inputs to the model that is used daily estimate GHI using sunshine duration (SD) and Top of the atmosphere (TOA) GHI as an input different from default values $a=0.25$ and $b=0.75$? What are the values of AP coefficients in in each location/station under study and in each macro-climatological region of South Africa? How accurate are the calibrated AP coefficients in estimating daily GHI in each location/station under study and in each macro-climatological region of South Africa?
3. Which is the best freely available (for research purposes) GHI satellite-based dataset and GHI reanalysis-based dataset in each location/station under study and overall, in South Africa? What are the benchmarks for hourly (rMBE, rRMSE, rMAE and R^2) in the study area?
4. Which is the overall best between the satellite-based datasets and Ångström–Prescott (AP) model in estimating daily GHI estimates in South African locations? What are the benchmarks for daily (rMBE, rRMSE, rMAE and R^2) in the study area?

1.6. Research Aims and Objectives

The overall aim of the study is to determine viable alternative methods or datasets that can be used to quantify the amount of GHI (on clear sky conditions and all sky conditions) at the Earth's surface in areas where there are no in-situ radiometric measurements.

The set objectives are to:

1. Calibrate (using meteorological input parameters) and validate six global horizontal irradiance (GHI) clear sky models namely: Bird [54,55], Simple Solis [56,57], McClell [30], Ineichen–Perez [58,59], Haurwitz [60,61] and Berger–Duffie [62] against GHI data recorded by a reference pyranometer only on clear sky days, to determine the overall of the best performing clear sky model for sites in South Africa.
2. Calibrate the Ångström–Prescott (AP) model regression coefficients a and b that could be used to approximate GHI in different climatic zones of South Africa, thus increasing the density of available GHI in South Africa.
3. Validate and rank (based on the performance) hourly satellite-based global horizontal irradiance (GHI) datasets: SOLCAST [63], Copernicus Atmosphere Monitoring Service (CAMS) [64,65], and Satellite Application Facility on Climate Monitoring (CMSAF SARA) [66] and two reanalysis-based: Fifth generation European Center for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) [33] and Modern-Era Retrospective Analysis for Research and Applications (MERRA2) [67] for sites in South Africa.
4. Compare daily satellite-based global horizontal irradiance (GHI) datasets estimates and Ångström–Prescott (AP) model estimates against daily measured in-situ GHI to

determine the overall best method in approximating daily GHI data for South African sites.

Hypothesis: In the study GHI models, model coefficients, GHI satellite-based, GHI reanalysis-based will be assessed and the most suitable data set in estimating GHI in South African locations will be identified.

1.7. Dissertation outline

Chapter 1 presents the introduction and background, literature review, problem statement(s), and research and aims of the study. Chapter 2 is about theoretical overview covering the section on the structure of the Sun and how the Sun generates solar radiation, solar radiation spectrum, Sun-Earth geometry, and the Global Mean radiative balance between Earth-Atmosphere. Chapter 2 covers the measurements of different solar radiation parameters and the instruments that are used, uncertainties of the instrument and data quality control to eliminate some of the uncertainties. Chapter 2 covers the theoretical background of the satellites and reanalysis-derived datasets. Chapter 3 is about the study area, the Processing of the in-situ data, data Quality Control and the Performance Evaluation Criteria used to benchmark the performance of estimated data against observed data. Chapter 4 is about the assessment of clear sky models which approximate the quantity of GHI under cloudless atmospheres. Chapter 5 is about the calibration and validation of Ångström–Prescott (AP) model, which can be used to approximate daily GHI from in-situ sunshine duration hours. Chapter 6 is about the validation of the satellite and reanalysis-derived GHI datasets that can be used to approximate GHI data, which can be used to make up for the scarcity of the in-situ radiation stations by increasing the GHI density. Chapter 7 compares the satellite-derived datasets and the Ångström–Prescott (AP) model in estimating daily GHI. Chapter 8 gives the overall conclusion of the findings. The APPENDIX provides the Python scripts used in data analysis and plotting.

References

1. Liou, K. N. (2002). *An introduction to Atmospheric Radiation*. Elsevier.
2. Dal Pai, A.; Escobedo, J.F.; Martins, D.; Teramoto, É.T. Analysis of hourly global, direct and diffuse solar radiations attenuation as a function of optical air mass. *Energy Procedia* **2014**, 57, 1060–1069.
3. Schwandt, M.; Chhatbar, K.; Meyer, R.; Mitra, I.; Vashistha, R.; Giridhar, G.; Gomathinayagam, S.; Kumar, A. Quality check procedures and statistics for the Indian SRRA solar radiation measurement network. *Energy Procedia* **2014**, 57, 1227–1236.
4. Zawilska, E.; Brooks, M. An assessment of the solar resource for Durban, South Africa. *Renew. Energy* **2011**, 36, 3433–3438.
5. Urraca, R.; Gracia-Amillo, A.M.; Huld, T.; Martinez-De-Pison, F.J.; Trentmann, J.; Lindfors, A.V.; Riihelä, A.; Sanz-Garcia, A. Quality control of global solar radiation data with satellite-based products. *Sol. Energy* **2017**, 158, 49–62.
6. Zawilska, E.; Brooks, M.J.; Meyer, A.J. A review of solar resource assessment initiatives in South Africa: The case for a national network. *In Proceedings of the World Renewable Energy Forum, Denver, CO, USA, 2012*.
7. Jiang, H.; Yang, Y.; Bai, Y.; Wang, H. Evaluation of the total, direct, and diffuse solar radiations from the ERA5 reanalysis data in China. *IEEE Geosci. Remote Sens. Lett.* **2020**, 17, 47–51.
8. Sianturi, Y.; Marjuki; Sartika, K. Evaluation of ERA5 and MERRA2 reanalyses to estimate solar irradiance using ground observations over Indonesia region. *AIP Conf. Proc.* **2020**, 2223, 020002.
9. Tang, W.; Yang, K.; He, J.; Qin, J. Quality control and estimation of global solar radiation in China. *Sol. Energy* **2010**, 84, 466–475.
10. Zhang, Q.; Cui, N.; Feng, Y.; Jia, Y.; Li, Z.; Gong, D. Comparative Analysis of Global Solar Radiation Models in Different Regions of China. *Adv. Meteorol.* **2018**, 1–21.
11. Almorox, J.; Benito, M.; Hontoria, C. Estimation of monthly Angström-Prescott equation coefficients from measured daily data in Toledo, Spain. *Renew. Energy* **2005**, 30, 931–936.
12. Adamala, S.; Reddy, Y.V.K. Evaluation of Different Solar Radiation Estimation Methods for Indian Locations. *Water Resour. Environ. Eng. II.* **2018**, 47–56.
13. De Medeiros, F.J.; Silva, C.M.; Bezerra, B.G. Calibration of Ångström-Prescott Equation to Estimate Daily Solar Radiation on Rio Grande do Norte State, Brazil. *Rev. Bras. Meteorol.* **2017**, 32, 409–416.

14. Polo, J.; Martín-Pomares, L.; Gueymard, C.A.; Balenzategui, J.L.; Fabero, F.; Silva, J.P. Fundamentals: Quantities, Definitions, and Units. In *Solar Resource Mapping—Fundamentals and Applications*; Green Energy and Technology; Polo, J., Martín-Pomares, L., Sanfilippo, A., Eds.; Springer: Zurich, Switzerland, **2019**; 1–14.
15. Reno, M.J.; Hansen, C.W.; Stein, J.S. *Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis*; SAND2012-2389; Sandia National Laboratories: Albuquerque, NM, USA; Livermore, CA, USA, 2012. Available online: https://scholar.google.com/scholar?q=related:NRwgbk5tC6sJ:scholar.google.com/&scioq=&hl=en&as_sdt=0,5 (accessed on 23 June 2020).
16. Gairaa, K.; Benkaciali, S.; Guermoui, M. Clear-sky models evaluation of two sites over Algeria for PV forecasting purpose. *Eur. Phys. J. Plus* **2019**, *134*, 534.
17. Sun, X.; Bright, J.M.; Gueymard, C.A.; Acord, B.; Wang, P.; Engerer, N.A. Worldwide performance assessment of 75 global clear sky irradiance models using Principal Component Analysis. *Renew. Sustain. Energy Rev.* **2019**, *111*, 550–570.
18. Engerer, N.A.; Mills, F.P. Validating nine clear sky radiation models in Australia. *Sol. Energy* **2015**, *120*, 9–24.
19. Dazhia, Y.; Jirutitijaroen, P.; Walsh, W.M. The Estimation of Clear Sky Global Horizontal Irradiance at the Equator. *Energy Procedia* **2012**, *25*, 141–148.
20. Antonanzas-Torres, F.; Urraca, R.; Polo, J.; Perpiñán-Lamigueiro, O.; Escobar, R. Clear sky solar irradiance models: A review of seventy models. *Renew. Sus. Energ. Rev.* **2019**, *107*, 374–387.
21. Badescu, V.; Gueymard, C.A.; Cheval, S.; Oprea, C.; Baci, M.; Dumitrescu, A.; Iacobescu, F.; Milos, I.; Rada, C. Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. *Renew. Energy* **2013**, *55*, 85–103.
22. Mikofski, M.M.; Hansen, C.W.; Holmgren, W.F.; Kimbal, G.M. Use of measured aerosol optical depth and precipitable water to model clear sky irradiance. In Proceedings of the 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC), Washington, DC, USA, 25–30 June 2017; pp. 110–116. Available online: <https://ieeexplore.ieee.org/abstract/document/8366314> (accessed on 15 June 2020).
23. Laguarda, A.; Abal, G. Clear-sky broadband irradiance: First model assessment in Uruguay. In Proceedings of the Publicado en las Actas del ISES Solar World Congress, Abu Dhabi, United Arab Emirates, 2017, doi:10.18086/swc.2017.21.05. Available online: <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/21620?mode=full> (accessed on 16 September 2020).
24. Marchand, M.; Ghennioui, A.; Wey, E.; Wald, L. Comparison of several satellite-derived databases of surface solar radiation against ground measurement in Morocco. *Adv. Sci. Res.* **2018**, *15*, 21–29.

25. Martín-Pomares, L.; Romeo, M.G.; Polo, J.; Frías-Paredes, L.; Fernández-Peruchena, C.M. Sampling design optimization of ground radiometric stations. In *Solar Resources Mapping*; Springer International Publishing: Cham, Switzerland, **2019**; 253–281.
26. Wilbert, S.; Stoffel, T.; Myers, D.; Wilcox, S.; Habte, A.; Vignola, F.; Wood, J.; Pomares, L.M. Measuring solar radiation and relevant atmospheric parameters. In *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications*; National Renewable Energy Laboratory: Golden, CO, USA, 2017. Available online: <https://hal-mines-paristech.archives-ouvertes.fr/hal-01184753> (accessed on 12 March 2020).
27. Moradi, I. Quality control of global solar radiation using sunshine duration hours. *Energy* **2009**, 34, 1–6.
28. Garg, H.P.; Garg, S.N. Measurement of solar radiation—II. Calibration and standardization. *Renew. Energy* **1993**, 3, 335–348.
29. Žák, M.; Mikšovský, J.; Pišoft, P. CMSAF radiation data: New possibilities for climatological applications in the Czech Re-public. *Remote Sens.* **2015**, 7, 14445–14457.
30. Masoom, A.; Kashap, Y.; Bansal, A. Solar Radiation Assessment and Forecasting Using Satellite Data; Tyagi, H., Agarwal, A.K., Chakraborty, P.R., Powar, S., Eds.; Springer: Singapore, 2019;45–71.
31. Lefèvre, M.; Oumbe, A.; Blanc, P.; Espinar, B.; Gschwind, B.; Qu, Z.; Wald, L.; Schroedter-Homscheidt, M.; Hoyer-Klick, C.; Arola, A.; et al. McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmos. Meas. Tech.* **2013**, 6, 2403–2418. Available online: www.atmos-meas-tech.net/6/2403/2013/doi:10.5194/amt-6-2403-2013 (accessed on 14 July 2020).
32. Ineichen, P.; Perez, R. A new airmass independent formulation for the Linke turbidity coefficient. *Sol. Energy* **2002**, 73, 151–157.
33. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, 146, 1999–2049.
34. Reno, M.J.; Hansen, C.W. Identification of periods of clear sky irradiance in time series of GHI measurements. *Renew. Energy* **2016**, 90, 520–531.
35. Zhandire, E. Predicting clear-sky global horizontal irradiance at eight locations in South Africa using four models. *J. Energy S. Afr.* **2017**, 28, 77–86.
36. Patel, S.S.; Rix, A.J. Water surface albedo modelling for floating photovoltaic plants. In *Proceedings of the South African Solar Energy Conference (SASEC)*, Mpekwini Beach Resort, Eastern Cape Province, South Africa, 25–27 November 2019. Available online: <https://www.sasec.org.za/papers2019/8.pdf> (accessed on 27 February 2020).

37. Javu, L.; Winkler, H.; Roro, K. Validating clear-sky irradiance models in five South African locations. In Proceedings of the South African Solar Energy Conference (SASEC), Mpekweni Beach Resort, Eastern Cape Province, South Africa, 25–27 November 2019. Available online: <https://www.sasec.org.za/papers2019/6.pdf> (accessed on 15 February 2020).
38. Angstrom, A. Solar and terrestrial radiation: Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quart. J. Roy. Meteorol. Soc.* 1924, 50, 121–126.
39. Prescott, J.A. Evaporation from a water surface in relation to solar radiation. *Trans. Roy. Soc. S. Aust.* 1940, 46, 114–118.
40. Iqbal, M. *An Introduction to Solar Radiation*; Academic Press: Toronto, ON, Canada, 1983.
41. Martínez-Lozano, J.A.; Tena, F.; Onrubia, J.E.; De La Rubia, J. The historical evolution of the [57] Angström formula and its modifications: A review and bibliography. *Agric. For. Meteorol.* 1984, 33, 109–128.
42. Tsung, K.Y.; Tan, R.; Ii, G.Y. Estimating the global solar radiation in Putrajaya using the Angstrom-Prescott model. *IOP Conf. Series: Earth Environ. Sci.* 2019, 268, 012056.
43. Besharat, F.; Dehghan, A.A.; Faghih, A.R. Empirical models for estimating global solar radiation: A review and case study. *Renew. Sustain. Energy Rev.* 2013, 21, 798–821.
44. Eberhard, A.A. *A Solar Radiation Data Handbook for Southern Africa*; Elan Press: Cape Town, South Africa, 1990.
45. Mulaudzi, S.T.; Sankaran, V.; Lysko, M.D. Solar radiation analysis and regression coefficients for the Vhembe Region, Limpopo Province, South Africa. *J. Energy South. Afr.* 2017, 24, 2–7.
46. Bright, J.M. Solcast: Validation of a satellite-derived solar irradiance dataset. *Sol. Energy* 2019, 189, 435–449.
47. Yang, D.; Bright, J.M. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. *Sol. Energy* 2020, 210, 3–19.
48. Marchand, M.; Lefèvre, M.; Saboret, L.; Wey, E.; Wald, L. 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. *Adv. Sci. Res.* 2019, 16, 103–111.
49. Qu, Z.; Oumbe, A.; Blanc, P.; Espinar, B.; Gesell, G.; Gschwind, B.; Klüser, L.; Lefèvre, M.; Saboret, L.; Schroedter-Homscheidt, M.; et al. Fast radiative transfer

- parameterisation for assessing the surface solar irradiance: The Heliosat 4 method. *Meteorol. Z.* 2017, 26, 33–57.
50. Thomas, C.; Wey, E.; Blanc, P.; Wald, L. Validation of three satellite-derived databases of surface solar radiation using measurements performed at 42 stations in Brazil. *Adv. Sci. Res.* 2016, 13, 81–86.
 51. Ameen, B.; Balzter, H.; Jarvis, C.; Wey, E.; Thomas, C.; Marchand, M. Validation of hourly global horizontal irradiance for two satellite-derived datasets in Northeast Iraq. *Remote Sens.* 2018, 10, 1651.
 52. Trolliet, M.; Walawender, J.P.; Bourlès, B.; Boilley, A.; Trentmann, J.; Blanc, P.; Lefèvre, M.; Wald, L. Downwelling surface solar irradiance in the tropical Atlantic Ocean: A comparison of re-analyses and satellite-derived data sets to PIRATA measurements. *Ocean Sci.* 2018, 14, 1021–1056.
 53. PIRATA Network. Available online: <https://www.pmel.noaa.gov/gtmba/pmel-theme/atlantic-ocean-pirata> online (accessed on 15 May 2021).
 54. Bird, R.E.; Hulstrom, R.L. Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces; No. SERI/TR-642-761; Solar Energy Research Inst.: Golden, CO, USA, 1981.
 55. Daryl, R.M. Solar radiation, Practical Modeling for Renewable Energy Applications. *Energy and the Environment*; CRC Press: Boca Raton, FL, USA, 2013.
 56. Mueller, R.; Dagestad, K.; Ineichen, P.; Schroedter-Homscheidt, M.; Cros, S.; Dumortier, D.; Kuhlemann, R.; Olseth, J.; Piernavieja, G.; Resie, C.; et al. Rethinking satellite based solar irradiance modelling. The SOLIS clear-sky module. *Remote Sens. Environ.* 2004, 91, 160–174. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0034425704000690> (accessed on 27 September 2020).
 57. Ineichen, P. A broadband simplified version of the Solis clear sky model. *Sol. Energy* 2008, 82, 758–762.
 58. Ineichen, P.; Perez, R. A new airmass independent formulation for the Linke turbidity coefficient. *Sol. Energy* 2002, 73, 151–157.
 59. Perez, R.; Ineichen, P.; Moore, K.; Kmiecik, M.; George, R.; Vignola, F. A new operational model for satellite-derived irradiances: Description and validation. *Sol. Energy* 2002, 73, 307–317.
 60. Haurwitz, B. Insolation in relation to cloudiness and cloud density. *J. Meteorol.* 1945, 2, 154–166.

61. Haurwitz, B. Isolation in relation to cloud type. *J. Meteorol.* 1948, 5, 110–113.
62. Berger, X. *Etude du Climat en Region Nicoise en vue d'Applications a l'Habitat Solaire*; CNRS: Paris, France, 1979.
63. SOLCAST. Available online: <https://solcast.com> (accessed on 15 April 2020).
64. Copernicus Portal. Available online: <https://atmosphere.copernicus.eu/data> (accessed on 16 March 2020).
65. Solar Radiation Data (SoDa) Service. Available online: <http://solar.atmosphere.copernicus.eu/cams-radiation-service> (accessed on 02 March 2020).
66. Pfeifroth, U.; Kothe, S.; Trentmann, J.; Hollmann, R.; Fuchs, P.; Kaiser, J.; Werscheck, M. *Surface Radiation Data Set—Heliosat (SARAH), 2.1 Ed; Satellite Application Facility on Climate Monitoring*: Offenbach, Germany, 2019.
67. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* 2017, 30, 5419–5454.

Chapter 2: Theoretical background

2.1. The Sun

In this section the structure of the Sun and how the Sun generates solar radiation are discussed.

The Sun is a sphere of intensely hot gaseous matter with a diameter of 13.9×10^5 km and is located 1.5×10^8 km away from the Earth [1-5]. The Sun is composed of the core, the interior, the convection zone, the photosphere, the reversing layer, the chromosphere, and the corona [1,4]. The core is characterised by extreme temperature ($8-40 \times 10^6$ K) and pressure (1×10^5 kg/m³) with a density of 150 g/cm³ [1,4]. The core is primarily composed of 70 % hydrogen, 28 % helium and small amounts of iron, silicon, neon, and carbon. The quota changes over time as hydrogen is transformed to helium, as shown in Figure 2.1. Extreme density and temperature in the core of the Sun create the ideal conditions for the thermonuclear reaction, resulting in 4×10^6 tonnes of hydrogen to be converted into helium every second.

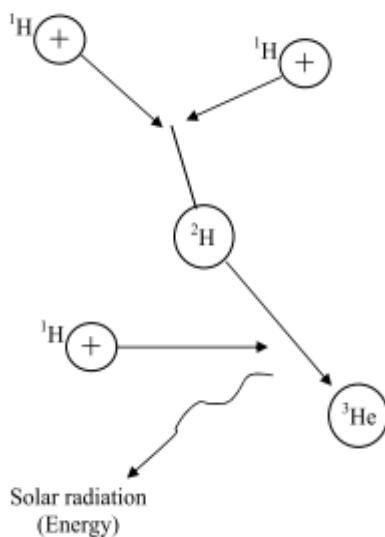


Figure 2.1: Reaction of Hydrogen (H) and Helium (He) in the Sun's core which result in the generation of solar radiation (Adapted from Sen, Zekai 2008 [4]).

Thermonuclear reaction results in the release of a great amount of energy, which then traverses from the core to the outer regions of the Sun. The core, with a radius of $0.2R$ (where $R = 13.9 \times 10^5$ km, is the diameter of the sun) is the source of all the energy from the Sun [1,4].

The region next to the core is the interior, it is also known as the radiation zone because the energy in the interior is conveyed by radiation. The radiation zone's temperature is about 1.3×10^4 K, and its density is about 70 kg/m^3 [1,4]. The convection zone is the region that is characterised by the movement of gases, it is formed once the energy exit the radiation zone. The convection zone extends from $0.7R$ to $1R$, its density is about 10^{-5} kg/m^3 and its temperature is about 5×10^3 K. The photosphere is the densest and the lowest part of the atmosphere. The chromosphere is the gas layer above the photosphere, it is approximately 1×10^4 km thick. The energy in this layer is transported by radiation, convective patterns and the reddish flames which extend thousands of kilometres before dropping down [4]. The corona is the outermost layer of the Sun, it radiates radiation that emanates from the Sun [4].

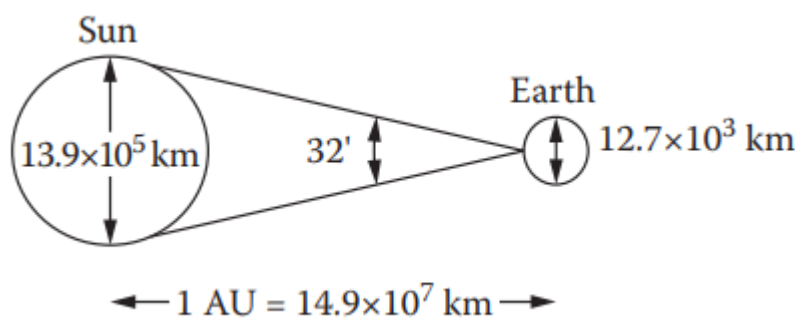


Figure 2.2: Sun-Earth distance (Adapted from Foster, R et al., 2009) [3]).

The Sun is the source of energy that affects atmospheric motions and the Earth's climate [1-5]. The energy released by the Sun is called solar radiation. The quantity and allocation of incoming solar radiation influence the weather and climate on the Earth. The quantity and strength of solar radiation that reaches the Earth vary in time and space [6]. The distance

between the Sun and Earth, the presence of clouds and other atmospheric constituents affect the amount of solar radiation that reaches at the Earth's surface. The shorter the Sun-Earth distance, the greater the quantity of solar energy received on a plane perpendicular to the Sun at the top of the atmosphere. Even though the Earth-Sun distance is large, about 14.9×10^7 km (Figure 2.2), there is a notable quantity of solar energy that reaches the Earth's surface. The solar radiation, per unit time, received on a plane normal to the direction of radiation, at mean Earth-Sun distance located at the top of the atmosphere and integrated over the whole spectrum is known as the solar constant (I_{SC}) [1-6]. The variation of the I_{SC} during a year is about $\pm 3\%$ as the Earth orbits the Sun and approximately 0.01% over 30 calendar days [1]. The I_{SC} value was initially determined from ground measurements, then from rockets a I_{SC} with a value of 1322 W/m^2 was determined. The availability of aircrafts, balloons, spacecrafts, satellites, and absolute radiometers enabled the I_{SC} value to be re-examined, to date the revised solar constant value of 1367 W/m^2 was determined by the World Radiation Centre (WRC) with an uncertainty of 1% [1]. According to Gueymard (2004) [7], a solar constant value of 1367 W/m^2 was also proposed by the World Meteorological Organization (WMO) in 1981. This is the I_{SC} value that was used in this study. I_{SC} value is given by using equation (2.1) from (Gueymard (2004) [7].

$$I_{SC} = 1367 \frac{\text{W}}{\text{m}^2} \quad (2.1)$$

2.2. Extra-terrestrial Solar Irradiance

Extra-terrestrial solar irradiance estimates the quantity of solar energy reaching at the top of the Earth's atmosphere. The irradiance that is normal to the direction of solar rays is called direct normal extra-terrestrial irradiance (DNI_{TOA}), and is given by Equation (2.2), as in [1].

$$DNI_{TOA} = \left(\frac{r}{R^2} \right) I_{SC} \quad (2.2)$$

$\left(\frac{r}{R^2} \right)$ is the eccentricity factor, E_o , it is approximated by equation (2.3) as:

$$E_o = 1 + 0.0333 \cos \left[\left(\frac{2\pi D}{365} \right) \right] \quad (2.3)$$

where D is the Julian day. So, equation (2.3) is given as:

$$DNI_{TOA} = \left[1 + 0.0333 \cdot \cos \left(\frac{2\pi D}{365} \right) \right] \cdot I_{SC} \quad (2.4)$$

The value of the DNI_{TOA} changes throughout the year because the Sun-Earth distance changes during the year due to the elliptical rotation of the Earth. To determine the amount of total Extra-terrestrial radiation received by a horizontal surface (GHI_{TOA}), DNI_{TOA} is multiplied by the cosine of zenith angle $\cos(\theta_z)$, it is given by equation (2.5 and 2.6), as in [3,5].

$$GHI_{TOA} = DNI_{TOA} \cdot \cos(\theta_z) \quad (2.5)$$

$$GHI_{TOA} = \left[1 + 0.0333 \cdot \cos \left(\frac{2\pi D}{365} \right) \right] \cdot I_{SC} \cdot \cos(\theta_z) \quad (2.6)$$

2.3. The Solar Radiation Spectrum

It has been proven that anybody with a temperature over the absolute zero Kelvin (-273.15°C) release radiation. The Sun has an essential surface temperature of approximately 5 800 K ($5\,526.85^\circ\text{C}$) because of atom fusion, therefore, it emits electromagnetic radiation with a distribution that is similar to that envisioned by Planck's law (Figure 2.3) [8]. The radiation released by the Sun extends over a wide range of wavelengths from radio waves (1 mm) through the visible to X-rays and gamma rays (less than 0.01 nm) [2]. The distribution of the radiation forms the solar radiation spectrum. The wavelength range of the solar radiation spectrum at the Earth's surface that is important for solar energy purposes ranges from about 250 nm to 300 nm [1]. The spectrum that is important for solar energy purposes covers three regions (Figure 2.3) of solar radiation spectrum near ultraviolet ($\lambda < 380$ nm), visible (380 nm to 780 nm) and near-infrared ($\lambda > 780$ nm) [1]. Different wavelength carries different

amount of energy, half of the solar radiation spectrum energy is in the visible region (about 48 % or 656 W/m^2) the remaining lies in the ultraviolet (about 6.4 % or 87 W/m^2) and infrared regions (45.6 % or 623 W/m^2) [1]. The solar radiation spectrum peaks at about 550 nm in the visible region. The spectral region (250 nm to 3000 nm) carries about 98 % energy of the solar spectrum [1]. This study will focus on the spectral region between 290 nm to 4000 nm, which according to [8], represents about 98.5 % energy of the solar spectrum. Pyranometers and pyrhemimeters usually measure between 280 nm and 2800 nm; this range represents approximately 97 % of the total solar irradiance.

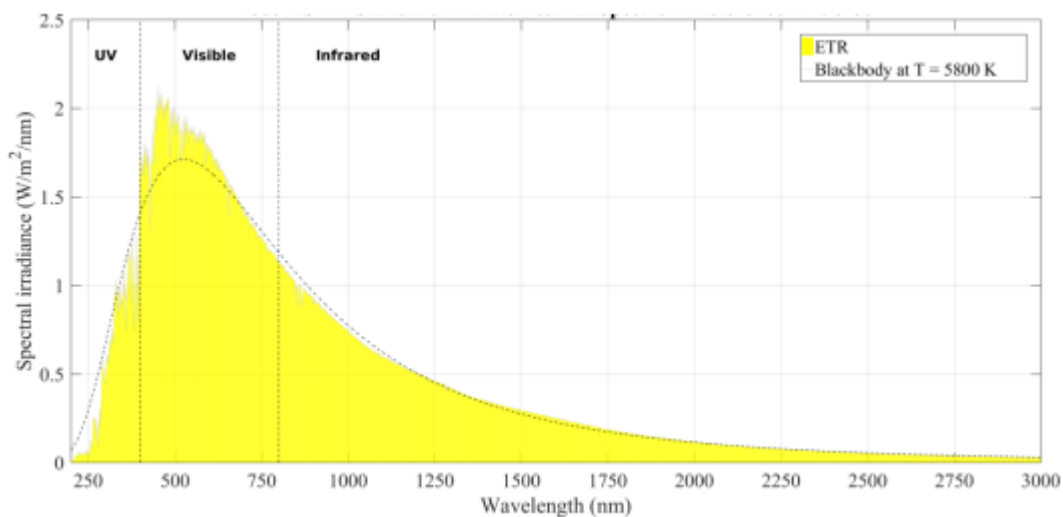


Figure 2.3: Extra-terrestrial solar radiation spectrum (Adapted from Habte, A., et.al 2008)[8].

The solar radiation spectral dispersal demonstrates how much solar radiation is available for each wavelength. It varies with time, and due to change air mass, which varies with zenith angle. Good knowledge of the solar radiation spectral distribution and the transmittance of the atmosphere is important in understanding and quantifying the amount of solar radiation that is lost in the atmosphere and the one that reaches the surface of the Earth.

The Atmosphere acts as a sieve by blocking some of the ultraviolet (UV) radiation from making it to the Earth's surface. In the first part of the solar radiation spectrum (less than 29 nm), almost

all the UV radiation is totally absorbed by ozone (O_3), oxygen (O_2) and nitrogen. This is the part of the radiation that is biologically dangerous and degrades many materials and colours [2].

The visible and infrared radiation can reach the Earth's surface even though they are dissipated by atmospheric components. There is slight absorption in the second part of the solar spectrum, about 40 % of solar energy measured at the Earth's surface is from the visible region. The infrared region is the most highly reflected radiation of the spectrum because of the long wavelengths.

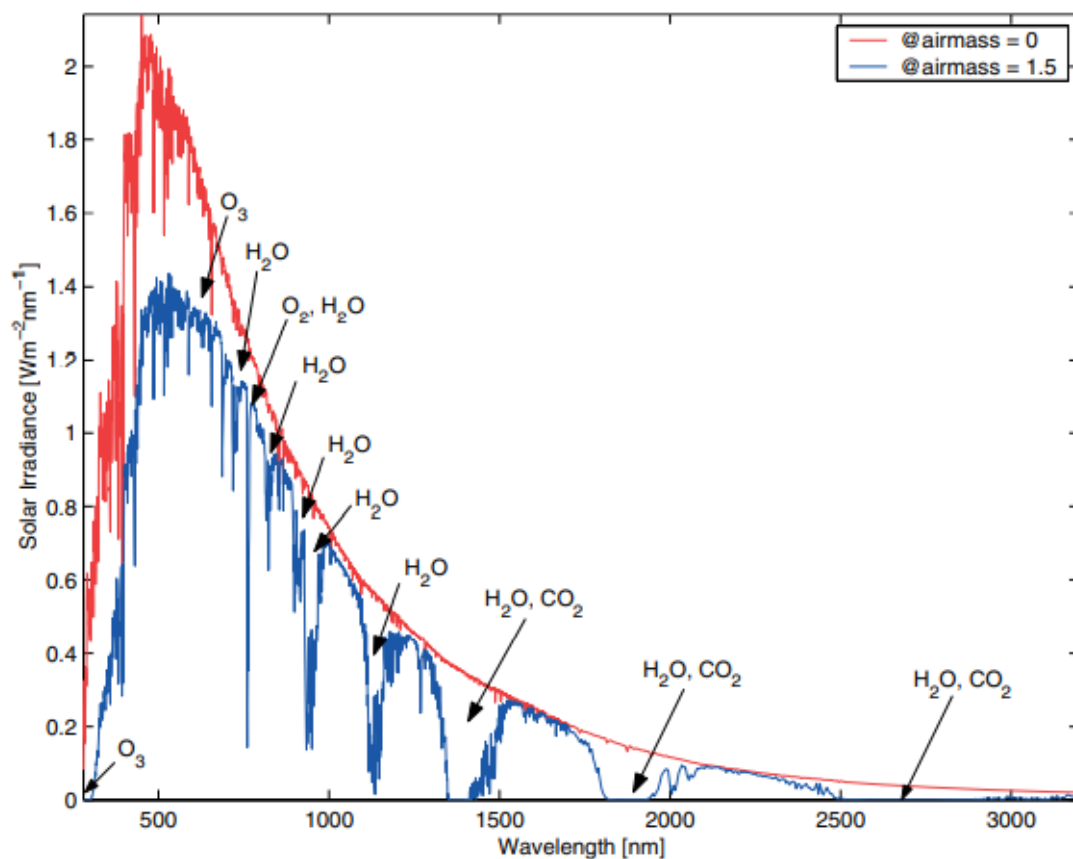


Figure 2.4: Spectral dispersal of solar radiation at the top of the Earth's surface (red) at airmass zero and at the Earth's surface (blue) when the airmass is equal to 1.5 (Adapted from Lysko 2006) [9].

The absorption of solar radiation in the visible and infrared solar radiation spectrum according to Iqbal, 2002 [2], and as shown in Figure 2.4, is as follows:

- Ozone absorbs some visible radiation at wavelengths from 450 nm to 770 nm.
- Oxygen absorbs solar radiation at the following wavelengths 630 nm, 690 nm, and 760 nm
- Water vapor absorbs solar radiation at 720 nm, 820 nm, 940 nm, 1100 nm, 1380 nm, 1870 nm, and 2700 nm.
- Carbon dioxide absorbs solar radiation at 1450 nm, 1600 nm, 2000 nm, and 2700 nm

The effect of nitrogen oxide, carbon monoxide and methane on the amount of solar radiation arriving at the Earth's surface is very low because of their lower concentration in the atmosphere [2]. Solar radiation arrives at the Earth's surface at approximately 1200 nm, 1600 nm, and 2200 nm windows [2]. At wavelengths greater than 3000 nm in the infrared range almost all the energy is absorbed by atmospheric carbon dioxide and water vapour [1,2].

Solar radiation from the top of the atmosphere to the Earth's surface is also altered by the relative air mass (AM) in the solar radiation spectrum. The relative AM is the optical passage length where solar radiation is absorbed and scattered when the Sun is not directly high up in the sky [1]. AM is given by equation (2.7) and the concept of air mass is demonstrated by Figure 2.5.

$$AM = \frac{1}{\cos \theta_z} = \sec \theta_z, \text{ where } \theta_z \text{ is the zenith angle.} \quad (2.7)$$

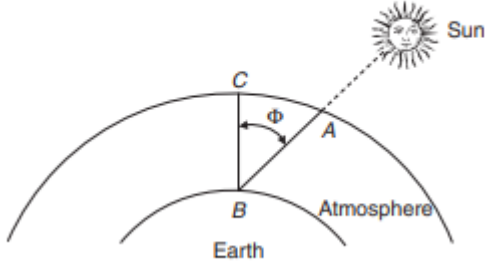


Figure 2.5: Fraction of the Atmosphere that the solar radiation must traverse before reaching the Earth respective to its overhead path length (A/C). {Adapted from Kalogirou, S. A. (2013) [5]}.

When the Sun is high up in the sky (i.e when $\theta_z = 0$, $AM = 1$), the minimum amount of AM or shorter path length occurs (solar noon), resulting in the higher quantity of solar energy reaching the Earth. Around sunset and sunrise when the Sun is further down in the sky with respect to a detector on the Earth's surface the further the Sun-Earth distance.

The AM estimation air mass approximation is moderately precise for $\theta_z < 80^\circ$, when the $\theta_z > 80^\circ$ more precise and complex models are needed. Kasten and Young (1989) [10], proposed the AM model equation (2.7) be improved to allow for the Earth's arch. The improved version is given by equation (2.8).

$$AM = \left[\frac{1}{\cos \theta_z + 0.50572 \cdot (6.07995 + (90 - \theta_z))^{-1.6364} + \cos \theta_z} \right] \quad (2.8)$$

Taking into consideration the effects of altitude Bird et al (1981) [11], recommended that the AM should further be improved by multiplying it with station pressure and then dividing it by the pressure at sea level (101325). The improved AM is known as absolute air mass (AM_a) and it is given by equation (2.9).

$$AM_a = p \cdot \frac{AM}{101325}, \text{ where } p \text{ is the pressure (in pascals).} \quad (2.9)$$

2.4. The Sun-Earth Geometry

The Earth revolves around the constant Sun in a slightly elliptical orbit, with an eccentricity (ϵ) of 0.01675 [2]. The eccentricity of the orbit produces changes in the Earth-Sun distance (this is because the Sun is located at one of the centres of the ellipse, not at its centre). The average Earth-Sun distance is about 14.9×10^7 km, and it varies around the year by about $\pm 1.7\%$ [1-3, 5,6]. The minimum Sun-Earth distance is about 14.7×10^7 km and the maximum is about 15.2×10^7 km, as shown in Figure 2.6. Duffie and Beckman (1991) [1] proposed an expression to calculate the Earth-Sun distance (E_0), equation (2.3).

The speed of the Earth varies within its elliptical orbit {higher the perihelion, the closest Earth-Sun distance in early January (14.7×10^7 km) than at the aphelion (the farthest Earth-Sun distance in early July (15.2×10^7 km))}. The duration of an orbit is about 365.25 days. The Earth's rotation around the Sun and the Earth's orbit rotation over its axis causes the astronomical seasonal changes. The Earth rotates on its axis once a day, this causes a change in day and night hours throughout the year [3,6].

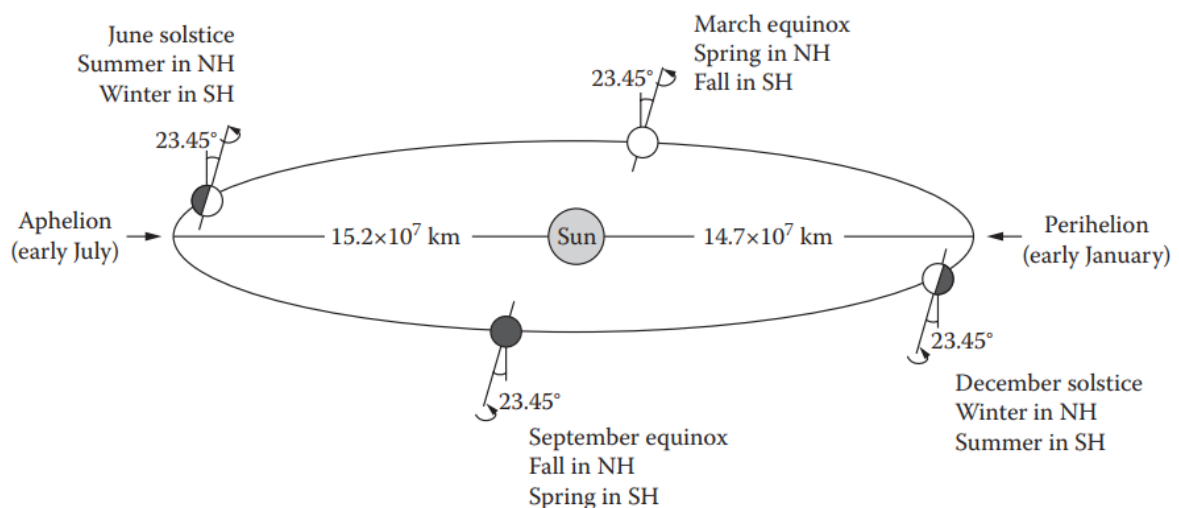


Figure 2.6: Earth's path about the Sun. (Adapted from Foster, R et al., 2009) [3].

The relative position of the Earth and the Sun at different times of the year is shown in Figure 2.6. In the Southern hemisphere (SH) the summer solstice occurs in December, winter solstice occurs in June, the equinoxes occur in March and September. In the SH, the shortest day of the year occurs around 21 June. The Sun keeps on rising in the sky and around 21 September the spring equinox is reached. The Sun reaches the highest point in the sky around 21 December resulting in the longest day (12 hours) of the year (summer solstice), the Sun lies directly over the Tropic of Capricorn. Around March 21 the longest day (12 hours) takes place again (fall equinoxes) [3].

Due to forces among the planets, the direction of the Sun is not always parallel to the equatorial plane of the Earth. The equatorial plane of the Earth is curved approximately 23.45 degrees relative to the Earth's rotation around the Sun. The angle that is formed between the horizontal line normal to a line between the Sun-Earth and the equatorial plane is known as the solar declination angle (δ). It varies with time from -23.45° to $+23.45^\circ$ [6] and it is defined in equation (2.10) from [12]:

$$\delta = -23.45 \sin \left[\frac{360(D+284)}{365} \right], \quad (2.10)$$

To enable accurate quantification or modelling of the quantity and intensity of solar energy striking any position of the Earth at any time, it is important to understand the locality of the Sun relative to a location of any orientation on the Earth's surface. To elucidate the position of the Sun relative to a location of any orientation at any time on the Earth's surface the following angles solar azimuth (γ_s), solar altitude (α_s), θ_z , and hour (ω) angles which are based on the Earth's coordinates must be known. These angles are important in the calculation of the day length and the amount of solar radiation allocation at any time at a given point. Figure 2.7 depicts the geometric interrelation of the angles which establish the orientation of the Sun in the sky.

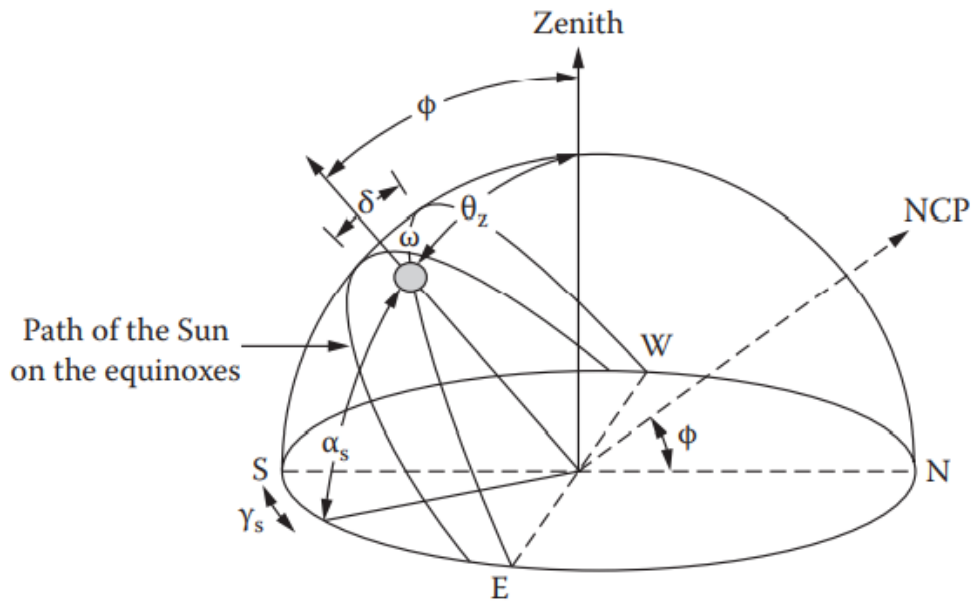


Figure 2.7: Position of the Sun in the sky in relation to the solar angles (Adapted from Foster, R et al., 2009) [3].

The solar altitude angle (α_s) is defined as the angle between the projected radiation ray and the location of the Sun [1]. When the Sun is above the skyline, $\alpha_s = 0^\circ$ and when it is at zenith, $\alpha_s = 90^\circ$.

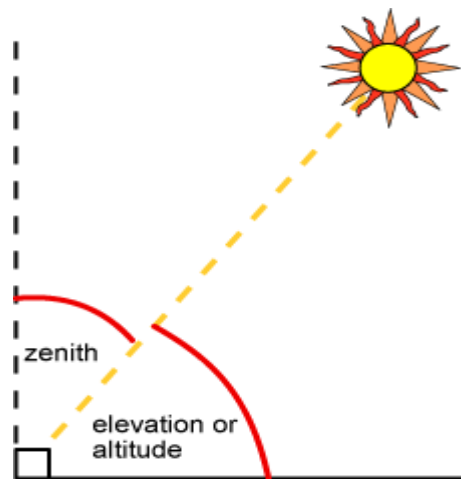


Figure 2.8: The correlation between the sun, zenith angle and the solar elevation angle. (Adapted from [13]).

Zenith angle (θ_z) is the angle between the Sun and the vertical [1]. The zenith angle complements the solar altitude ($\theta_z + \alpha_s = 90^\circ$) as shown in Figure 2.8, adapted from [13]. The θ_z is given in [1,2] as:

$$\cos \theta_z = \sin \phi \cdot \sin \delta + \cos \phi \cdot \cos \delta \cdot \cos \omega \quad (2.11)$$

In Equation (11), ϕ is the latitude of the location, and δ is the solar declination angle, as given by equation (2.9). In equation (2.11) ω is the hour angle, given by [1,2,12,14], as in equation (2.12):

$$\omega = 15^\circ \cdot (ST - 12), \quad (2.12)$$

$$\text{with, } ST = STD + 4 \cdot (L_{st} - L_{local}) + EOT$$

(2.13) where ST is the Solar Time, STD is the local standard time, L_{st} is the standard meridian for the local time zone, L_{local} is the local longitude, and EOT is the equation of time given by Equation (2.14):

$$EOT = 229.18(0.000075 + 0.001868 \cos B - 0.032077 \sin B - 0.014615 \cos 2B - 0.04089 \sin 2B) \quad (2.14)$$

$$\text{With, } B = (D - 1) \cdot \frac{360}{365} \quad (2.15)$$

The hour angle ω is the angular distance between the Sun's location at a certain time and its top point for that day when crossing the local meridian at the solar noon (distance that the Earth moves in a day) [1,2]. The hour angle is established on the estimated time required (24 hours) for the Earth to rotate (i.e., complete 360°) in one day. The ω changes by 15° ($360/24$) per hour. The ω is zero at solar noon and it is negative before noon and a positive in the afternoon.

Other important equations are hour angles at sunrise and sunset. The equation was derived by assuming that $\theta_z = 90^\circ$ at sunset, therefore $\cos \theta_z = 0$. These angles are a result of solar latitude angle (ϕ) and the declination angle (δ). The sunset angle (ω_s) is given by equation (2.17) from [2].

From equation (11) $\cos \theta_z = \sin \phi \cdot \sin \delta + \cos \phi \cdot \cos \delta \cdot \cos \omega$, at sunset $\cos \theta_z = 0$.

$$\sin \phi \cdot \sin \delta + \cos \phi \cdot \cos \delta \cdot \cos \omega_s = 0 \quad (2.16)$$

$$\omega_s = \cos^{-1} [(-\tan(\phi) \cdot \tan(\delta))] \quad (2.17)$$

The sunrise and the sunset hour angles are equal in magnitude but having different signs [2]. Since the sunset angle is equal to the length of day from solar noon, the daylength (N_d) is equivalent equal to the double the sunset angle from solar noon [2]. N_d in hours, is given by equation (2.18) from [2].

$$N_d = -2(\omega_s) = \frac{2}{15} \cos^{-1} [(-\tan(\phi) \cdot \tan(\delta))] \quad (2.18)$$

The solar azimuth (γ_s) is the angle on the ground surface between the projection of the Sun rays and the north-south direction line [1,2]. It is given by Equation (2.19) from [2]:

$$\cos(\gamma_s) = -\frac{[\sin(\delta) - \cos(\theta_z) \cdot \sin(\phi)]}{[\cos(\phi) \cdot \sin(\theta_z)]} \quad (2.19a)$$

$$\gamma_s = \cos^{-1} \left(\frac{\cos(\theta_z) \cdot \sin(\phi) - \sin(\delta)}{\cos(\phi) \cdot \sin(\theta_z)} \right) \quad (2.19)$$

Solar time (ST) is established due to the movement of the Earth around the polar axis and on its movement around the Sun, it does not correspond with local standard time (STD) [1]. A clock day is precisely 24 hours while a solar day deviates a bit from 24 hours [2]. The differences are due to the Earth's ellipticity movement about the Sun resulting in the Earth's rotation variation over a year. The offset between ST and the noon STD is called the Equation of Time (EOT). The empirical equation that is used to correct time differences is given by equation (2.14).

To account for and correct for the differences in longitude between ST and STD, the fact that it takes 4 minutes for the Sun to complete 1° of longitude across the sky is considered [5]. To determine the specific time at the longitude where the detector is located, that is the conversion of the clock time (STD) to solar time (ST) is given by equation (2.13).

2.5. Global Mean radiative balance between Earth-Atmosphere

During its path in the atmosphere to the ground surface, the transmittance of the atmosphere attenuates the incoming solar radiation before it reaches the ground surface. The quantity of solar radiation at the ground surface is about 50 % compared to the original quantity at the top of the Earth before it interacts with the atmosphere [4]. The interaction of the solar radiation from the top of the atmosphere, the atmosphere and the ground surface is summarised in Figure 2.9, adapted from [16].

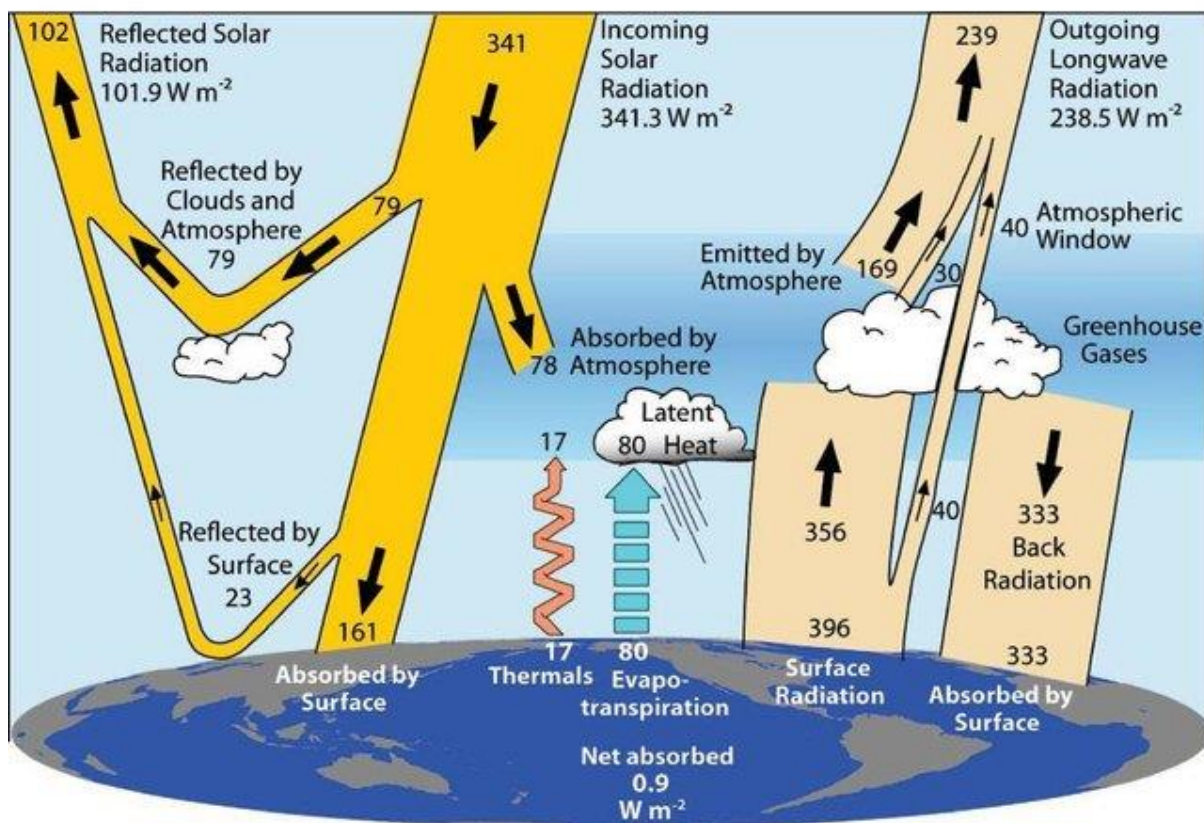


Figure 2.9: Global Mean radiative balance between Earth (ground surface) and Atmosphere, short wave radiation on the left. (Adapted from Trenberth et al., 2009) [16].

From Figure 2.9, the quantity of shortwave solar radiation at the top of the Earth's surface on average is about 341.3 W/m^2 . Clouds, aerosols, and molecules absorb about 79 W/m^2 (about 23 %) of the incoming solar radiation. Reflection and deflection of shortwave solar radiation

into space because of the presence of clouds represent 78 W/m^2 (about 23 %) of the incoming shortwave solar radiation. The quantity of solar radiation that reaches the ground surface depending on the ground albedo of the ground surface is about 184 W/m^2 ($161+23$) which is more than 50 %. 23 W/m^2 which is (about 7 %) of the solar radiation that makes it to the ground surface is returned by the ground to the atmosphere. In total 102 W/m^2 ($79 + 23$), i.e., about 30 % is returned from the ground surface due to the ground albedo and clouds. From 184 W/m^2 ($161+23$) which is more than 50 % of the shortwave solar radiation that reaches the ground surface, 55 W/m^2 (about 30 %) occurs as direct short-wave radiation and 49 W/m^2 (about 27 %) occurs as diffuse shortwave radiation. Direct normal irradiance (DNI) makes it to the ground from the Sun without being disturbed by the atmosphere, while diffuse horizontal irradiance (DIF) which is due to the interaction with atmospheric attenuations and as a result its direction has been changed. The summation of DNI and DIF at the ground surface is referred to as the global horizontal irradiance (GHI) [1]. The distribution of shortwave radiation components is illustrated in Figure 2.10. The three components GHI, DNI and DIF at the Earth's surface are related as follows:

$$GHI = DNI \cdot \cos \theta_z + DIF \quad (2.20)$$

where, θ_z is the zenith angle, as shown in equation (2.11)

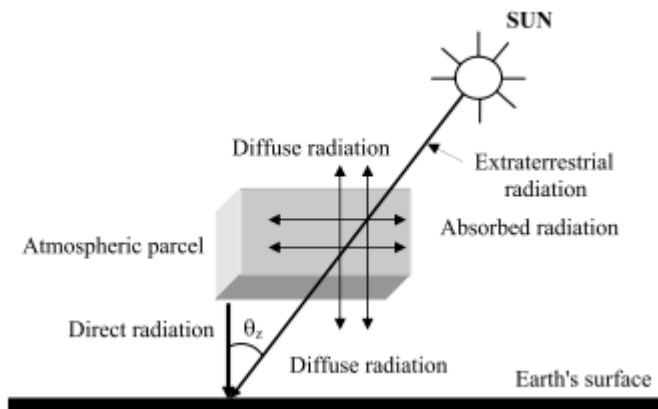


Figure 2.10: Distribution of diffuse radiation, direct radiation, and absorbed radiation (Adapted from Sen, Zekai 2008 [4]).

2.6. Solar Radiometry

Radiometry is the science of measuring electromagnetic radiation. An instrument that is used to measure electromagnetic radiation is known as a radiometer [17-19]. A radiometer uses a detector or a thermopile to sense the electromagnetic radiation and then generates a signal that is a representative of the quantity being monitored [17,18,21]. The main purpose of radiometry is to quantify the electromagnetic radiation that makes it to the ground surface which is of high great importance to various applications. This section focuses on the radiometers that were used to measure solar radiation data used in the study, these are pyranometers and pyrhemliometers. A pyranometer is used for measuring GHI arriving from the whole hemisphere. A pyranometer with a shading ring or shading ball to block the direct beams is used to measure DIF. A pyrhemliometer is used to measure DNI. A pyrhemliometer is usually attached to an automatic Sun-tracking device that follows the sun throughout the day.

2.6.1. Measurement of Global Horizontal Irradiance (GHI)

Pyranometers (Kipp & Zonen CMP10, CMP11, CMP21) measure shortwave global radiation irradiance (GHI) arriving from the whole hemisphere received by a horizontal plane surface. They are classified as Secondary standard pyranometers according to [21-23] and as Spectrally Flat Class A according to [24].

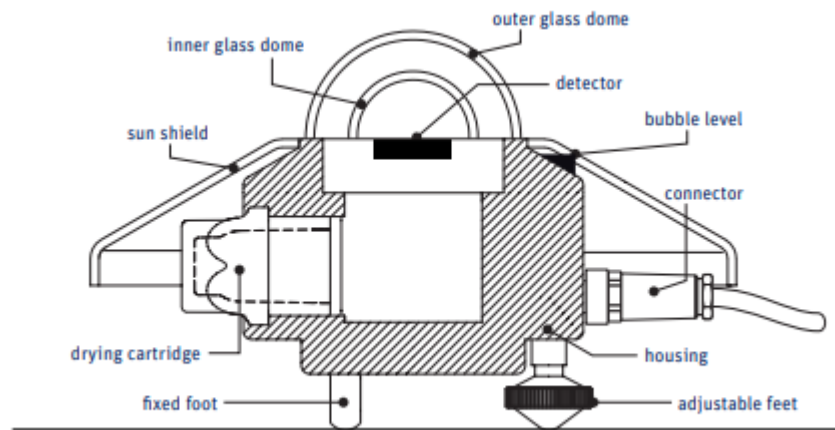


Figure 2.11: Schematic of a double dome pyranometer Kipp & Zonen CMP10, CMP11, CMP21) (Adapted from Kipp & Zonen Instruction Manual CMP Series Pyranometer [23].



Figure 2.12: Unshaded pyranometer (Kipp and Zonen CMP11 pyranometer) used for measuring GHI at the Polokwane International Airport solar radiation station (SAWS station).

2.6.2. Measurement of Diffuse Horizontal Irradiance (DIF)

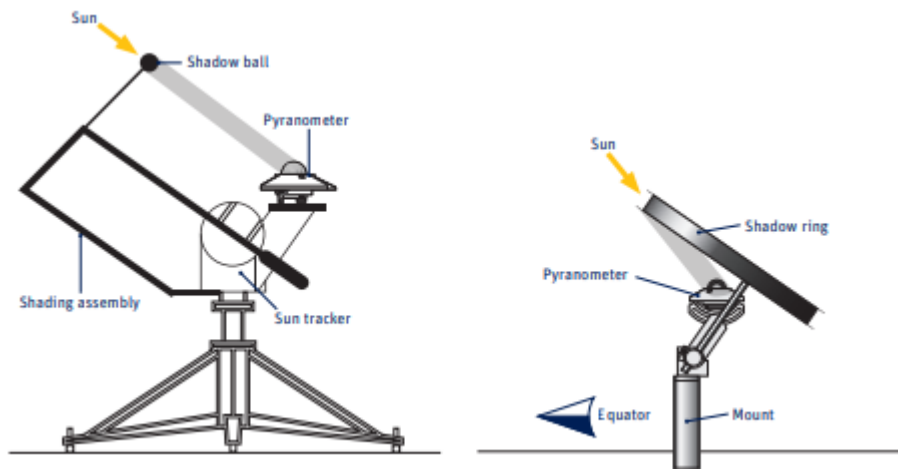


Figure 2.13: Pyranometer shaded with a shadow ball (left) and shaded with a static shadow ring (right) (Adapted from [23]).

Pyranometers (Kipp & Zonen CMP10, CMP11, CMP21) can also be used to measure DIF by using a shadow ball that moves with the Sun tracker to screen the DNI from making its way to the pyranometer Figure 2.13 (left) or by using the static shadow ring to screen the DNI from making its way to the pyranometer as shown in Figure 2.13 (right). The static shadow ring requires periodic manual alteration to compensate for the change in the direction of the sun's path with change in seasons. The static shadow ring might obstruct some DIF from reaching the pyranometer resulting in unreliable DIF measurements, to eliminate some of the uncertainties in measured DIF, subsequent treatment of the measured DIF data is recommended [2,19,22,23].

2.6.3. Structure of a Pyranometer

Pyranometers have two adjustable feet and one fixed foot (Figure 2.11 and Figure 2.12), the feet are levelled when installing the pyranometer on a horizontal surface to align the built-in

bubble level. The bubble level alignment enables the pyranometer's thermopile to be aligned horizontally for accurate GHI and DIF measurements.

Pyranometers have a drying cartridge part (Figure 2.11), this is where a silica gel or desiccants which prevents condensation of moisture inside the sensor domes are inserted. The build-up in moisture can cause rust inside the pyranometer, the moisture might also decrease solar radiation values recorded by a pyranometer [18,23]. The desiccants are monitored if a change in colour from yellow to colourless is detected they must be replaced as soon as possible.

Pyranometers have two glass domes (Figure 2.11 and Figure 2.12), the inner dome and the outer glass dome. The outer glass dome is about 50 mm in diameter its main purpose is to prevent the instrument from dust particles. The inner glass dome is about 30 mm in diameter its function is to define the spectral range of the pyranometer. The inner glass dome is made of a material that allows about 97 to 98 % of the solar radiation spectrum to be conveyed to the thermopile of a pyranometer where it will be absorbed [18,23]. According to [21,22] a pyranometer is designed to measure the hemispherical (GHI or DIF) irradiance integrated over the wavelength range from 300 to 3000 nm. The pyranometer domes are covered by a removable dome cover or sun shield (Figure 2.11). The dome cover shields the thermopile from thermal convection, rain, and wind, which may influence its temperature. However, the dome cover does not influence the transference of the irradiance equally from all directions of the pyranometer [18,23].

The detector or the thermopile (Figure 2.11), is the sensing element of a pyranometer. The detector is assembled by many thermocouple junction pairs connected electrically in series. The top part of the thermopile and the bottom part are separated by an insulated surface. When the pyranometer is exposed to the irradiance, the top part of the detector will react to the change in temperature while the bottom part will not react to the change in temperature because of the

thermal resistance layers which prevent the heat to be transferred from the top layer to the bottom layer. The differences in temperature between the top part and the bottom part of the detector will produce an electromotive force (emf) that is directly proportional to the temperature difference created. This is called the thermoelectric effect [19,20,23].

To quantify GHI recorded by the pyranometer, the emf created by the temperature differences must be divided by the sensitivity or a calibration factor of the pyranometer. The sensitivity of each detector is unique, as a result each radiometer has unique calibration factor [23]. The equation that relates GHI, voltage and sensitivity of a pyranometer is given below equation (2.21), derived from [23].

$$GHI = \frac{U_{emf}}{S} \quad (2.21)$$

Where GHI is the Global Horizontal irradiance (GHI) in (W/m²), U_{emf} is the output of a pyranometer in microvolts (μV) and S is the sensitivity or calibration factor of the pyranometer's thermopile in μV/ W/m².

The sensitivity of the pyranometers' thermopile degrades when it is constantly exposed to the sun, this results in change in the sensitivity or a calibration factor of a pyranometer which is used in the quantification of GHI or DIF. To correct for the drift the pyranometer's sensitivity, Kipp & Zonen recommends the recalibration of pyranometers every 2 years [23]. The pyranometer can be recalibrated using the following methods, according to [19,22].

- ISO 9846: 1993 Calibration of a pyranometer using a pyrhelimeter [25], that is by comparison with a standard pyrhelimeter (DNI signal) and calibrated shaded pyranometer (DIF signal).
- ISO 9846: 1993 Calibration of a pyranometer using a pyrhelimeter [25], that is by comparison with a standard pyrhelimeter (DNI signal) using the Sun as a source of

solar radiation with a movable shading disc from the pyranometer (to create alternating GHI and DIF signals).

- ISO 9846: 1993 Calibration of a pyranometer using a pyrhelimeter [25], that is by comparison with a standard pyrhelimeter (DNI signal) using the Sun as a source of solar radiation and two pyranometers to be calibrated measuring GHI and DIF.
- ISO 9847: 1992 Calibration of field pyranometers by comparison to a reference pyranometer [26], that is by comparison with a standard pyranometer using the Sun as a source of solar radiation.
- ISO 9847: 1992 Calibration of field pyranometers by comparison to a reference pyranometer [26], that is, indoors using the stable artificial Sun by comparing a pyranometer to a similar standard reference pyranometer calibrated outdoors.
- Indoors, using an integrated chamber simulating DIF by comparing with similar standard reference pyranometer.
- Comparison with a standard pyrhelimeter using a pyranometer covered with a collimating tube [19].

The reference pyranometer must be of the same class or higher-class, same type, same model and from the same manufacturer as the field pyranometer [23]. The reference pyrhelimeter must be a secondary standard or the primary standard, self-calibrating Absolute Cavity Radiometer (ACR) [25]. The reference pyranometer or pyrhelimeter must have a calibration certificate with information detailing the traceability to the World Radiometric Reference (WRR) from World Radiation Centre (WRC) in Davos, Switzerland where the standard World Radiometric Reference (WRR) is maintained.

2.6.4. Measurement of Direct Normal Irradiance (DNI)

A pyrheliometer (Kipp and Zonen CHP1 pyrheliometer) is an instrument that is used to measure the DNI. The instrument is usually attached to an electrically driven Sun-tracking device (e.g Kipp & Zonen's SOLYS 2 Sun Tracker) that follows the Sun and continuously points the pyrheliometer to the Sun throughout the day. Pyrheliometers (CHP1) are classified as Secondary standard pyranometers according to [21,22,27] and as Spectrally Flat Class A according to ISO 9060:2018 [24].

A schematic of a pyrheliometer is shown in Figure 2.15 and the actual pyrheliometer is shown in Figure 2.14. The pyrheliometer is specifically designed to measure the DNI with a narrow field of view (FOV) limited to 5° [18,19,22,27]. The solar tracker points it to the Sun to enable that the solar disk always falls within the 5° FOV of the pyrheliometer. DNI from the pyrheliometer decreases rapidly to zero in cloudy conditions because of the narrow FOV.

The front window is fitted with a quartz to protect the pyrheliometer and to act like a sieve that allows only solar radiation spectrum between 200 nm and 4000 nm to pass [18,27]. The quartz window of the pyrheliometer, allows, approximately 97-98 % of the solar radiation spectrum to be conveyed through the acceptance angle, passes through the window and the collimation tube to the thermopile where it will be absorbed. The thermopile, which is a sensing element of a pyrheliometer is located placed at the end of the tube as shown in the schematic (Figure 2.15).



Figure 2.14: A pyrheliometer used to measure the DNI, mounted on the Kipp & Zonen's SOLYS 2 Sun Tracker at Thulamela Information Centre, Thohoyandou, (SAWS station).

The thermopile consists of active junctions and reference junctions. Active junction increases its temperature when the pyrheliometer absorbs solar radiation in a FOV when the pyrheliometer is pointed directly to the Sun while the reference junction keeps a constant stable temperature. The differences in temperatures between the active and reference junctions when

exposed to the Sun produces an emf that is directly proportional to the different temperature established, this process is called thermoelectric effect [27].

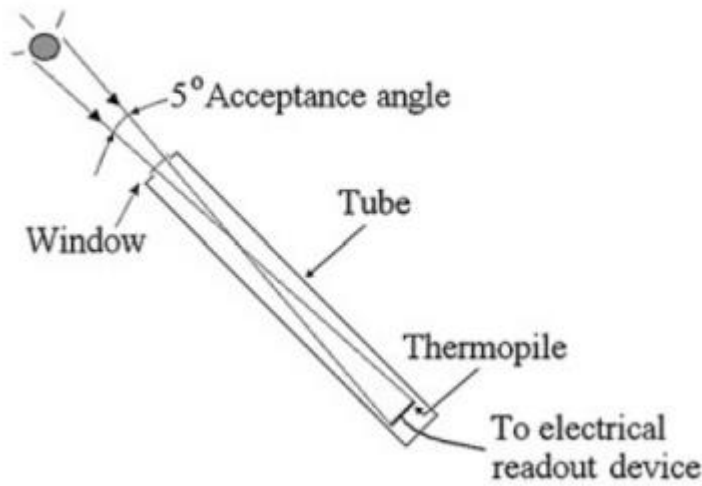


Figure 2.15: A Schematic of a pyrliometer. {Adapted from Paulescu et al. (2012) [18]}.

To quantity DNI recorded by the pyrliometer, the electromotive force created by the temperature differences must be divided by the sensitivity or a calibration factor of a pyrliometer. The equation that relates DNI, voltage and sensitivity of a pyrliometer is given below equation (2.22), derived from [27].

$$DNI = \frac{U_{emf}}{S} \quad (2.22)$$

Where DNI is the Direct Normal Irradiance in (W/m^2), U_{emf} is the output of a pyrliometer in microvolts (μV) and S is the sensitivity or calibration factor of the pyrliometer's thermopile in $\mu V/ W/m^2$.

The thermopile sensitivity which quantifies DNI recorded by the pyrliometer degrades with time when exposed to solar radiation, calibration every two years to determine the new sensitivity is advised by Kipp & Zonen [11]. The pyrliometer is recalibrated using ISO 9059:

1990 [12] methodology, which is the outdoor (using the Sun as the source of solar radiation) calibration of field pyrheliometers (e.g. secondary standard Kipp & Zonen CHP1) by comparison to a reference pyrheliometer. The reference pyrheliometer may be the same model (secondary standard Kipp & Zonen CHP1) or the primary standard, self-calibrating Absolute Cavity Radiometer (ACR).

Table 2.1: Specification of pyranometer (CMP series) and pyrheliometer (CHP1) based on [23,27].

Sensor Model	Kipp & Zonen CMP10/CMP11/ CMP21 pyranometer	Kipp & Zonen CHP1 pyrheliometer
ISO 9060:1990 Classification	Secondary Standard	Secondary Standard
ISO 9060:2018 Classification	Spectrally Flat Class A Pyranometer	Spectrally Flat Class A Pyrheliometer
Spectral Range	285 to 2800 nm	200 to 4000 nm
Sensitivity	7 to 14 $\mu\text{V}/\text{W}/\text{m}^2$	7 to 14 $\mu\text{V}/\text{W}/\text{m}^2$
Expected output range (0 to 1500 W/m^2)	0 to 20 mV	0 to 20 mV
Zero offsets (a) thermal radiation (at 200 W/m^2) (b) temperature change (5 K/hr)	< 7 W/m^2 < 2 W/m^2	- < 1 W/m^2
Detector type	Thermopile	Thermopile
Temperature Dependence of Sensitivity	< 1 % (-10 °C to +40 °C)	-
Response Time	< 5 s	< 5 s
Non-Stability (change/year)	< 0.5 %	< 0.5 %
Non-Linearity	< 0.2 % (0 to 1000 W/m^2)	-
Directional Error	< 10 W/m^2 (up to 80 ° with 1000 W/m^2 beam)	-
Tilt Error	< 0.2 % (0 to 90 ° at 1000 W/m^2)	-
Level Accuracy	0.1 °	-

Operating Temperature Range	-40 to +80 °C	-40 to +80 °C
Maximum Irradiance	4000 W/m ²	4000 W/m ²
Expected Daily Uncertainty	< 2 %	< 2 %
Dome Diameter	50 mm	38 mm
Width	150 mm with shield	76 mm
Height	68 mm	314 mm
Weight	0.6 kg	0.9 kg
Required tracking accuracy	-	< 0.5 ° from ideal
Field of view	180 °	5 ° ±0.2 °
Accuracy of bubble level	< 0.1 °	-
Slope angle		1 ±0.2 °

From (Table 2.1), a summary of the specification of pyranometers and pyrhemometers (thermopile sensors) is given. It is evident that the thermopile sensors are not perfect, they contribute to some uncertainties in the recorded data. These uncertainties are changes in sensitivity of the detector with constant exposure to solar radiation. This can be corrected by calibration, spectral selectivity, pyranometer directional error, pyranometer temperature dependence of sensitivity, response with time, pyranometer tilt effects, pyranometer cosine error, level accuracy, azimuthal error, zero offset due to thermal radiation and temperature change, non-stability, and also operational errors.

The pyranometer type was used in the study (Kipp & Zonen CMP11) is a thermopile type pyranometer. Photodiode or Silicon or Semiconductor pyranometer and Photovoltaic (PV) pyranometers are other types of pyranometers that can be used to record GHI data.

Table 2.2: Summary of different types of pyranometers based on detector type, spectral range, and their uses.

Pyranometer type	Thermopile type	Photodiode/Silicon /Semiconductor	Photovoltaic (PV)/ Reference PV cell
Detector	Thermopile (convert thermal energy into electrical energy)	Photodiode (convert light into current)	-
Spectral range	285 to 2800 nm	400 to 900 nm	Varies
Advantages	Long-term stability, Precise reading of the total available solar radiation	Cheap and easy to use	Cheap and easy to use.
Disadvantages	Expensive	Large temperature, azimuth and cosine errors and long response time. Low accuracy because it does not respond to the full solar spectrum.	Spectrally selective response, less precise than the thermopile, very sensitive with a fast response.
Applications	Scientific research, Minute, hourly and daily measurements in weather stations, Solar energy utilization, agriculture	Daily measurements in weather stations and in field testing. Acceptable for integrating solar radiation over periods of a day or longer.	To measure the cell's power and PV modules. Applied in simulators and with PV systems for calculating power and system performance.

Other pyranometers in the market that can be used to measure GHI and DIF are Apogee SP-510 and SP-610 pyranometers [29], Middleton EQ08-B pyranometer [30], SR30-M2-D1 Hukseflux pyranometers [31], EKO MS-80S Pyranometer [32], Eppley Global Precision Pyranometer [33], LI-200R Pyranometer [34] and Delta-T SPN1 Sunshine Pyranometer [35].

Other pyrhelimeters in the market that can be used to measure DNI are Delta-T SPN1 Sunshine Pyranometer [35], Eppley Normal Incidence Pyrhelimeter (sNIP) [36], EKO MS-

57 pyrheliometer [37], Middleton Solar DN5-B pyrheliometer [38] and Hukseflux DR30-D1 pyrheliometer [39].

Other Sun-tracking devices in the market that can be used to point DNI sensors to the Sun and to support a shadow ball that moves with the Sun tracker to restrict DNI making it to the pyranometer when measuring DIF are Kipp & Zonen SOLYS Gear Drive Sun Tracker [40], Middleton AST-03T Active Solar Trackers [41], Geonica Sun Tracker Sun Tracker 3000 SMS (solar monitoring systems) [42], EKO STR-21G-S2 Sun tracker [43], The Eppley Automatic Solar Tracker (Model SMT) [44].

2.7. Data Quality Control of Data

Measuring solar radiation is subject to errors [45], this is because radiometers produce electric voltage when it reacts with radiation, which is then converted into measurement of solar radiation [46]. The origin of errors is classified as equipment and operation related errors [47]. Equipment errors are introduced by type of sensor used (as summarised in Table 3.1) while operational errors are not introduced by the type of instrument used and include factors such as shading of the sensors, dust covering the sensor domes, inaccurate levelling, power failures, detached sensor cables station shut-downs, electric fields in the vicinity of cables or a malfunction in the data acquisition hardware [47]. Therefore, radiometry instruments should be taken care of and maintained for accurate measurements.

Some of the practices that can be done to eliminate errors includes:

- Cleaning the sensors' domes to reduce the effect of strains on the recorded values.
- Checking the alignment or levelling of the sensors.

- Monitoring the desiccants and replacing them when they change colour from yellow to colourless to remove moisture within the dome.
- Regular inspection of recording instruments and document exact times of the maintenance events.
- Perform regular calibrations as recommended by the instrument manufacturer.
- Station status live monitoring to detect problems and solve them as soon as possible.

Even if the best practices are applied the use of the thermopiles exposed to solar radiation in the field introduces unavoidable errors. To pick out errors in the recorded data, quality control checks (QC) should be applied to solar radiation data [48]. Various QC methods in the literature includes Baseline Solar Radiation Network (BSRN) QC [49], NREL SERI QC [50], QCRad methodology [51], MESOR recommendations [52,53], ENDORSE [54], RMIB [55], MDMS [56], AQC [57] and HelioClim model QC [58]. In the study BSRN QC test and HelioClim model QC were applied to the recorded data.

2.8. Methods to generate GHI datasets

Satellite and reanalysis-derived data sources are alternative methods to estimate GHI in areas where there are no in-situ measurements because of their wider coverage.

2.8.1. Satellite-derived datasets

Satellite-derived datasets estimate GHI at the Earth's surface from meteorological satellites using two approaches namely statistical and physical [59]. Statistical method is established on the relationship between satellite and ground data. Physical method is based on the radiative transfer models which create a correlation between satellite and ground data [59]. A training set of GHI at the ground surface is required to determine the parameters of the regression

model, which in turn will predict the GHI. The basis of the methodology assumes that amount of the cloud cover determines the amount of GHI [59]. Cloud cover index is first determined for each site of the satellite image and then the cloud cover index information is used to estimate GHI statistically for that site [59]. The correlation between satellite and ground data is given by equation (2.23) derived from [59].

$$E(t) = \rho(t)I_{0j}(\sin v)^{(1+a)}, \quad (2.23)$$

Where albedo $\rho(t)$ is the clear sky-ground albedo, I_{0j} is the solar irradiance at top of atmosphere, v is the solar elevation angle and a is a parameter to characterise the total atmospheric transmittance which depends on geographical site, it ranges between 0.14-0.17 depending upon the geographical site [59].

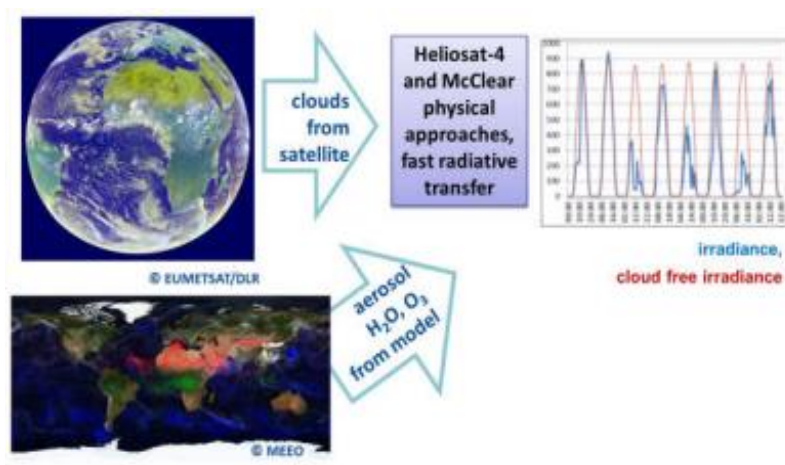


Figure 2.16: Process used to estimate GHI from satellite images (Adapted from Lefèvre and Wald [60]).

Basically, the interrelation between GHI at the ground surface and the satellite images is created by a formula that merges cloud data as shown in Figure 2.16. Cloud information is created from a cloud index i.e., the difference between a cloud pixel and a clear sky pixel and clear sky model. Linke turbidity, water vapour, ozone and aerosol data are added as inputs to a formula which will then approximate the quantity of GHI at the ground surface. Different

satellite-based datasets are generated using different sources of meteorological information. Data sources used in the study are described in chapter 6 (subsection 6.3.2) and chapter 7 (subsection 7.3.1).

2.8.2. Reanalysis-based datasets

Reanalysis-based datasets were named based on the way they are generated, i.e., reanalysis which is a systematic approach to produce data sets [61]. Reanalyses-based datasets are generated using a consistent data homogenization algorithm and model(s) which assimilate all available observations every 6-12 hours over the period being analysed. The consistent algorithms provide a constant change of the climate condition at each time interval. The only parameter that varies is the input data collected from observational network (radiosonde, satellite, buoy, aircraft, and ship reports) [61]. Reanalysis-based datasets have proven to be useful alternative data sources, however the uncertainties in observation data from various sources might introduce incorrect trends. The lack of input observational data in some areas affects the accuracy of reanalysis in a different location, time period, and variable considered [61]. The reanalysis-derived datasets assessed in this study are elucidated in chapter 6 (section 6.3.1).

References

1. Duffie, John A., William A. Beckman, and Nathan Blair. *Solar engineering of thermal processes, photovoltaics and wind*. John Wiley & Sons, 2020.
2. Iqbal, Muhammad. *An introduction to solar radiation*. Elsevier, 2012.
3. Foster, R., Ghassemi, M. and Cota, A., 2009. *Solar energy: renewable energy and the environment*. CRC press.
4. Sen, Zekai. *Solar energy fundamentals and modeling techniques: atmosphere, environment, climate change and renewable energy*. Springer Science & Business Media, 2008.
5. Kalogirou, S. A. (2013). *Solar energy engineering: processes and systems*. Academic Press.
6. Lucien Wald. *BASICS IN SOLAR RADIATION AT EARTH SURFACE*. 2018. <https://hal-minesparistech.archives-ouvertes.fr/hal-01676634>
7. Gueymard, C.A. The sun's total and spectral irradiance for solar energy applications and solar radiation models. *Sol. Energy* 2004, 76, 423–453, doi:10.1016/j.solener.2003.08.039.
8. Habte, A., Stoffel, T., Perez, R., Myers, D., Gueymard, C.A., Blanc, P. and Wilbert, S., 2017. Overview of solar radiation resource concepts
9. M. D. Lysko, (2006). *Measurements and Models of Solar Irradiance*, Doctoral Dissertation, Norwegian University of Science and Technology, Faculty of Natural Sciences and Technology, Department of Physics, Trondheim
10. Kasten, F.; Young, A.T. Revised optical air mass tables and approximation formula. *Appl. Opt.* 1989, 28, 4735–4738, doi:10.1364/AO.28.004735.
11. Bird, R.E.; Hulstrom, R.L. Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces; No. SERI/TR-642-761; Solar Energy Research Inst.: Golden, CO, USA, 1981.
12. Cooper, P.I. The absorption of radiation in solar stills. *Sol. Energy* 1961, 12, 333–346, doi:10.1016/0038-092X(69)90047-4.

13. C.B.Honsberg and S.G.Bowden, “Photovoltaics Education Website,”<https://www.pveducation.org/pvcdrom/properties-of-sunlight/elevation-angle>”, 2019
14. Spencer, J.W. Fourier series representation of the position of the sun. Search 1971, 2, 172. Available online: <https://ci.nii.ac.jp/naid/10005108360/> (accessed on 2 April 2020).
15. C.B.Honsberg and S.G.Bowden, “Photovoltaics Education Website,”<https://www.pveducation.org/pvcdrom/properties-of-sunlight/azimuth-angle>”, 2019. [16] Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, 90(3), 311-324.
16. C.B.Honsberg and S.G.Bowden, “Photovoltaics Education Website,”<https://www.pveducation.org/pvcdrom/properties-of-sunlight/azimuth-angle>”, 2019. [16] Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, 90(3), 311-324.
17. Datla, R. U., & Parr, A. C. (2005). 1. Introduction to Optical Radiometry. *Optical Radiometry. Series: Experimental Methods in the Physical Sciences*, 41, 1-34.
18. Paulescu, Marius, Eugenia Paulescu, Paul Gravila, and Viorel Badescu. *Weather modeling and forecasting of PV systems operation*. Springer Science & Business Media, 2012.
19. Balenzategui, J.L., Fabero, F. and Silva, J.P., 2019. Solar Radiation Measurement and Solar Radiometers. In *Solar Resources Mapping* (pp. 15-69). Springer, Cham.
20. Garg HP, Garg SN. Measurement of solar radiation. Pt. 1; Radiation instruments. *Renewable Energy*;(United Kingdom). 1993 Jan 1;3.
21. ISO 9060:1990 Solar energy — Specification and classification of instruments for measuring hemispherical solar and direct solar radiation.
22. World Meteorological Organization (WMO)‘s Guide to Meteorological Instruments and Methods of Observation, WMO-No. 8, 7th edn. In: The CIMO guide Chapter 7 for solar radiation measurements https://www.seedmech.com/documents_folder/wmo_no_8.pdf (last accessed on 15 April 2020).
23. Kipp & Zonen Instruction Manual CMP Series Pyranometer CMA Series Albedometer 2016 . Available online: <https://cn.kippzonen.com/Download/72/Manual-Pyranometers-CMP-series-English> (accessed on 7 October 2020).

24. ISO 9060:2018 Solar energy — Specification and classification of instruments for measuring hemispherical solar and direct solar radiation.
25. ISO 9846: 1993—Solar energy—Calibration of a pyranometer using a pyrhelimeter.
26. ISO 9847: 1992—Solar energy—Calibration of field pyranometers by comparison to a reference pyranometer.
27. Kipp & Zonen Normal Incidence Pyrhelimeter. CHP1 Pyrhelimeter Manual 2008 <https://www.kippzonen.com/Product/18/CHP1-Pyrhelimeter#.YZIoTU5ByUk> (accessed on 7 October 2020).
28. ISO 9059: 1990—Solar energy—Calibration of field pyrhelimeters by comparison to a reference pyrhelimeter.
29. Apogee pyranometers <https://www.apogeeinstruments.com/pyranometer/> (accessed on 15 November 2021).
30. Middleton pyranometers <https://www.middletonsolar.com/products/> (accessed on 15 November 2021).
31. Hukseflux pyranometers <https://www.hukseflux.com/products/solar-radiation-sensors/pyranometers> (accessed on 15 November 2021).
32. EKO Pyranometers <https://eko-eu.com/products/solar-energy/pyranometers/ms-80s-pyranometer> (accessed on 15 November 2021).
33. Eppley Global Precision Pyranometer <http://www.eppleylab.com/instrument-list/global-precision-pyranometer/> (accessed on 15 November 2021).
34. LI-200R Pyranometer <https://www.licor.com/env/products/light/pyranometer.html> (accessed on 15 November 2021).
35. Delta-T SPN1 Sunshine Pyranometer <https://delta-t.co.uk/product/spn1/> (accessed on 15 November 2021).
36. Eppley Normal Incidence Pyrhelimeter <http://www.eppleylab.com/instrument-list/normal-incidence-pyrhelimeter/> (accessed on 15 November 2021).

37. EKO MS-57 pyrheliometer <https://eko-eu.com/products/solar-energy/pyrheliometers/ms-57-pyrheliometer> ((accessed on 15 November 2021).
38. Middleton Solar DN5-B pyrheliometer <https://www.middletonsolar.com/products/product1.htm> (accessed on 15 November 2021).
39. Hukseflux pyrheliometer <https://www.hukseflux.com/products/solar-radiation-sensors/pyrheliometers> (accessed on 15 November 2021).
40. Kipp & Zonen SOLYS Gear Drive Sun Tracker : <https://www.kippzonen.com/Product/284/SOLYS-Gear-Drive-Sun-Tracker#.YXhrKppBzIU> (accessed on 15 November 2021).
41. Middleton AST-03T Active Solar Trackers <https://www.middletonsolar.com/products/product12.htm> (accessed on 15 November 2021).
42. Geonica Sun Tracker Sun Tracker 3000 SMS (solar monitoring systems) <http://www.huksefluxusa.com/wp-content/uploads/2017/03/Sun-Trackers.pdf> (accessed on 15 November 2021).
43. EKO STR-21G-S2 Sun tracker <https://eko-eu.com/products/solar-energy/solar-monitoring-stations/str-21g-s2-solar-monitoring-systems> (accessed on 15 November 2021).
44. EEF The Eppley Automatic Solar Tracker (Model SMT) <http://www.eppleylab.com/instrument-list/automatic-solar-tracker/> (accessed on 15 November 2021).
45. Moradi, I., 2009. Quality control of global solar radiation using sunshine hours. *Energy* 34, 1–6. <http://dx.doi.org/10.1016/j.energy.2008.09.006>.
46. Zo, Il-Sung, et al. "Baseline Surface Radiation Network (BSRN) quality control of solar radiation data on the Gangneung-Wonju National University radiation station." *Asia-Pacific Journal of Atmospheric Sciences* 53.1 (2017): 11-19.
47. Urraca, Ruben, et al. "Quality control of global solar radiation data with satellite-based products." *Solar Energy* 158 (2017): 49-62.
48. Schwandt, Marko, et al. "Quality check procedures and statistics for the Indian SRRA solar radiation measurement network." *Energy procedia* 57 (2014): 1227-1236.

49. Long, C.N., Dutton, E.G., 2002. BSRN Global Network Recommended QC Tests, V2.0. BSRN Technical Report <
<http://ezksun3.ethz.ch/bsrn/admin/dokus/qualitycheck.pdf> > .
50. Maxwell, E, S Wilcox, M Rymes. 1993. User's Manual for SERI QC Software, Assessing the Quality of Solar Radiation Data. Solar Energy Research Institute, Golden, CO.
51. Long, C.N., Shi, Y., 2008. An automated quality assessment and control algorithm for surface radiation measurements. *Open Atmos. Sci. J.* 2, 23–27. <http://dx.doi.org/10.2174/1874282300802010023>.
52. Hoyer-Klick, C., H.G. Beyer, D. Dumortier, M. Schroedter-Homscheidt, L. Wald, M. Martinoli, C. Schillings, et al. 2008. Management and Exploitation of Solar Resource Knowledge. In *EUROSUN 2008, 1st International Conference on Solar Heating, Cooling and Buildings*, Lisbon, Portugal.
53. Hoyer-Klick, Carsten, H Beyer, D Dumortier, L Wald, C Schillings, B Gschwind, L Menard, et al. 2009. MESoR – MANAGEMENT AND EXPLOITATION OF SOLAR RESOURCE KNOWLEDGE. In *SolarPACES 2009*, Berlin : Germany (2009).
54. Espinar, Bella, Lucien Wald, Philippe Blanc, Carsten Hoyer-Klick, Marion Schroedter Homscheidt, Thomas Wanderer. 2011. Project ENDORSE - Excerpt of the Report on the Harmonization and Qualification of Meteorological Data: Procedures for Quality Check of Meteorological Data. <https://hal-minesparistech.archives-ouvertes.fr/hal-01493608>.
55. Journée, Michel, Cédric Bertrand. 2011. Quality Control of Solar Radiation Data within the RMIB Solar Measurements Network. *Solar Energy* 85 (1): 72–86. <https://doi.org/10.1016/j.solener.2010.10.021>.
56. Geuder, N., F. Wolfertstetter, S. Wilbert, D. Schüller, R. Affolter, B. Kraas, E. Lüpfer, B. Espinar. 2015. Screening and Flagging of Solar Irradiation and Ancillary Meteorological Data. *Energy Procedia* 69 (May): 1989–98. <https://doi.org/10.1016/J.EGYPRO.2015.03.205>.
57. Molineaux, B., Ineichen, P., 2003. Automatic Quality Control of Daylight Measurement: Software for IDMP Stations, Technical Report. Vaulx en Velin, France: International Daylight Measurement Programme, École National des Travaux Publics <<http://idmp.ente.fr>>.

58. Geiger, M.; Diabaté, L.; Ménard, L.; Wald, L. A web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy* **2002**, 73, 475–480, doi:10.1016/S0038-092X(02)00121-4.
59. Cano, D.; Monget, J-M; Albuisson, M.; Guillard, H; Regas, N; Wald, L. A method for the determination of the global solar radiation from meteorological satellite data. *Solar Energy*, Elsevier, 1986, 37(1), pp.31-39. Elsevier, 1986, 37 (1), pp.31-39. <hal-00464051>
60. Lefèvre, M., Wald, L., CAMS-72 Solar radiation products, D72.2.3.1 Regular Validation Report Issue #20 S-O-N 2017, 2018, https://atmosphere.copernicus.eu/sites/default/files/2018-07/20_CAMS72_2015SC3_D72.2.3.1-SON2017_RAD_validation_report-201806_v1_0.pdf
61. Dee, Dick, Fasullo, John, Shea, Dennis, Walsh, John & National Center for Atmospheric Research Staff (Eds). Last modified 12 Dec 2016. "The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables." Retrieved from <https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables> (accessed on 7 October 2021).

Chapter 3: Study Area and observation data

3.1. Study Area

The study area was determined by considering the locations of the South African Weather Service (SAWS) radiometric stations which measures the 1-minute GHI data. The area of study is thirteen sites located in all six macro-climate regions in South Africa. Climatic zones are the regions with similar climatic conditions [1] and according to [2] climatic zones were established in order to classify different areas based on their maximum energy demand and maximum energy consumption. These macro-climate regions and locations of the 13 sites study area are shown on the map in Figure 3.1. The colour shaded climate zones shown in Figure 3.1 is based on SAWS macro classification and the climate codes which are indicated in square brackets are the “micro” zones based on the Council for Scientific and Industrial Research (CSIR) Köppen-Geiger climate classification (KGCC) as given by Conradie in [2]. Table 3.1 gives the sites’ KGCC, latitude, longitude, altitude, the period of assimilated data and the climate zone description. The 13 stations measure GHI by means of a CMP11 secondary-standard pyranometer from Kipp & Zonen. GHI measurements are acquired using a Campbell Scientific CR1000 data logger with a sampling rate of 5 s and integrated to 1 minute.

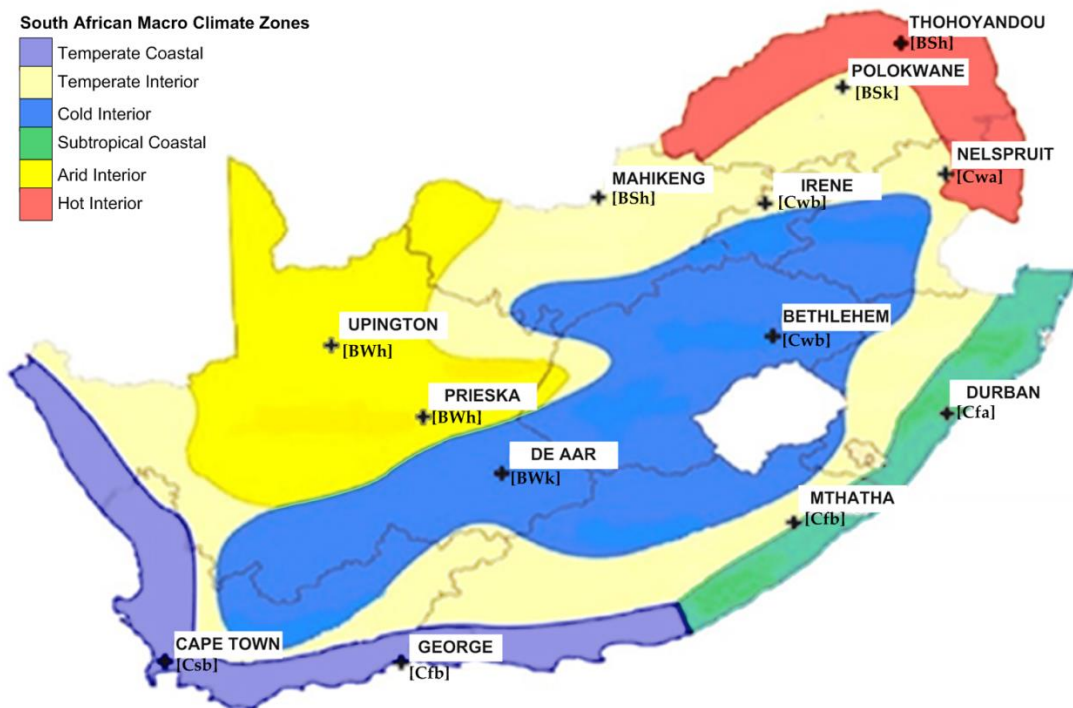


Figure 3.1: A map showing South African Weather Service's radiometric station location, SAWS macro climate zones and with CSIR Köppen-Geiger "micro" climate in square brackets.

As shown in Table 3.1 and Figure 3.1, the thirteen stations are evenly distributed in six different climatic regions of South Africa. The station altitude ranges from 86 m to almost 1688 m. The study area is bounded by latitudes 23.08° to 34.35° South and longitudes 18.48° to 31.11° East. The study area is bounded by the South Indian Ocean on the east coast (nearby where Durban and Mthatha stations are located), the South Atlantic on the west coast and the Southern Ocean to the south (nearby where George and Cape point stations are located). The rest of the stations are located inland.

Where each radiometric station is located there is an automatic weather station (AWS) measuring hourly sunshine duration (SD), hourly temperature, rainfall, pressure, humidity, wind speed and wind direction providing auxiliary meteorological parameters.

Table 3.1. South African Weather Services radiometric stations with Köppen–Geiger climate classification (KGCC), altitude, latitude, longitude, observation period, average number of clear sky days per year (percentage of clear sky days per year), annual aggregated diffuse fraction, humidity, and the percentage of data outliers removed per station (colors are described in Table 3.2).

Station	KGCC	Altitude (m)	Longitude (°)	Latitude (°)	GHI Observation Period	Clear sky days	Diffuse Fraction	Humidity	Outliers (%)
Upington	BWh	848	21.12	−28.48	01-02-2014 to 30-11-2019	97 (27)	0.18	35.4	4.47
Prieska	BWh	989	22.71	−29.68	01-09-2013 to 31-08-2019	78 (21)	0.18	38	3.88
De Aar	BWk	1284	23.99	−30.67	01-05-2014 to 31-12-2019	58 (16)	0.2	44.5	4.23
Bethlehem	Cwb	1688	28.3	−28.25	01-01-2015 to 31-12-2019	43 (12)	0.31	59.1	6.57
Irene	Cwb	1524	28.21	−25.91	01-03-2014 to 31-12-2019	62 (17)	0.3	54.9	2.99
Mahikeng	BSh	1289	25.54	−25.81	01-01-2016 to 31-12-2019	77 (21)	0.24	43.9	6.4
Polokwane	BSk	1233	29.45	−23.86	01-03-2015 to 31-12-2019	49 (13)	0.31	58.2	5.12
Nelspruit	Cwa	870	31.1	−25.39	01-02-2014 to 31-12-2019	39 (11)	0.4	62	5.85
Thohoyandou	BSh	619	30.38	−23.08	01-03-2015 to 31-10-2017	50 (14)	0.34	60.8	4.06
Mthatha	Cfb	744	28.67	−31.55	01-07-2014 to 31-12-2019	19 (5)	0.33	68.1	4.98
Durban	Cfa	91	31.11	−29.61	01-03-2015 to 31-12-2019	20 (5)	0.39	72.8	5.34
Cape Point	Csb	86	18.48	−34.35	01-02-2015 to 31-12-2019	12 (3)	0.34	77.2	4.96
George	Cfb	192	22.38	−34.01	01-01-2015 to 31-12-2019	11 (3)	0.36	79.2	2.75

Table 3.2. The coding representing the different levels of clear sky days, diffuse fraction, and humidity.

Parameter/Colour	Humidity (H)	Clear Sky Days (CL)	Diffuse Fraction (DF)
Green	$H < 50\%$	$CL > 20\%$	$DF < 0.2$
Yellow	$50\% < H < 60\%$	$15\% < CL < 20\%$	$0.2 < DF < 0.25$
Blue	$60\% < H < 70\%$	$10\% < CL < 15\%$	$0.25 < DF < 0.32$
Orange	$H > 70\%$	$CL < 5\%$	$DF > 0.32$

In Table 3.1, the cells for clear sky days, diffuse fraction, and humidity are color coded. The coding represents different levels of clear sky days, diffuse fraction, and humidity. The meteorological information is used to explain the reason behind the performance of the satellite-based and the reanalysis-based datasets in different stations. The color-coding limits and distributions are summarized in Table 3.2.

ERA5 hourly cloud data from 2013–2019 were used to calculate the average clear sky days per year for each station, and this information was used to evaluate the performance of datasets based on the frequency of cloud occurrences. Diffuse fraction and relative humidity information for the study site were from Mabasa et al. (2018) [3].

3.2. Pre-Processing of Observation Data

The methodology used to pre-process and data quality control of GHI in this study is summarized in the flow chart in Figure 3.2. The methodology consists of the following steps: (1) Preprocessing and Baseline Solar radiation Network quality control of in-situ GHI data; (2) Removing outliers; (3) Averaging one minute to 15 minutes; (4) Averaging four slots of 15 minute to obtain hourly averages; (5) Calculate daily averages from hourly averages; (6) Daily average quality check using HelioClim model; (7) Removing outliers.

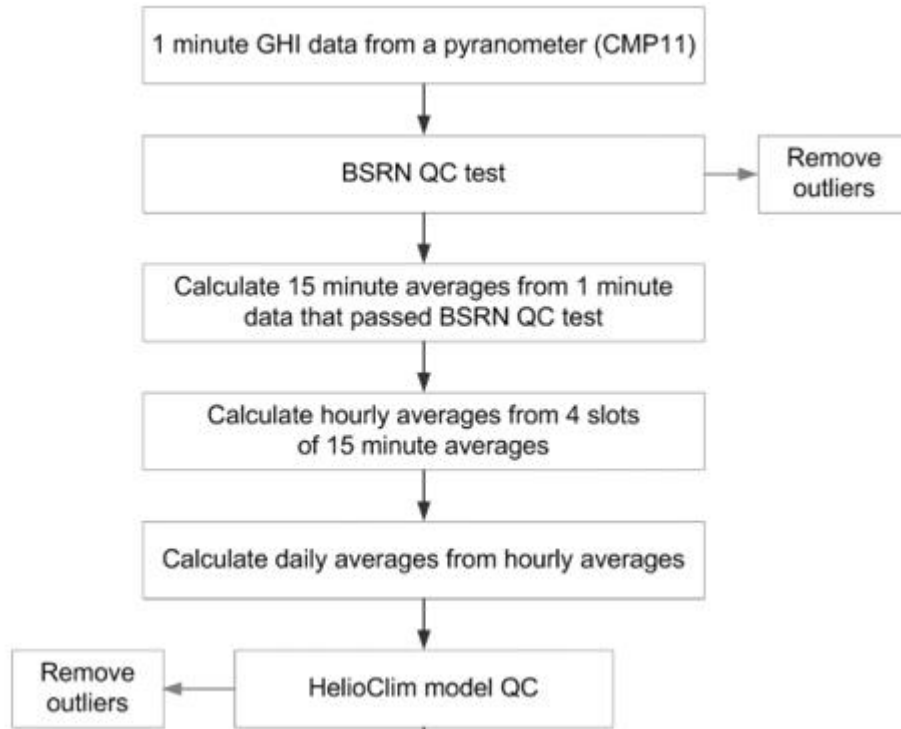


Figure 3.2: The flowchart of the approach used to process and quality control reference GHI data.

One-minute average GHI data recorded from each of the 13 stations using a CMP11 pyranometer from Kipp and Zonen were pre-processed, subjected to “physically possible” limit check, which is aimed at detecting extremely large errors in the radiation data and “extremely rare” values checks of the BSRN QC test [4-6], and the outliers were removed. The minute data that passed the BSRN QC test were converted to 15-minute averages; then hourly averages were calculated from four slots of 15-minute averages. This methodology was used in [3, 5, 7-11]. All the datasets were synchronized to South African standard time (SAST) to avoid the misalignment of time series datasets.

Hourly average GHI observation data were also further processed by removing all the data points recorded on hours when the solar zenith angle was greater than 90 °. Hourly averages were then averaged to daily average values. Daily averages were then subjected to HelioClim model QC, described by Geiger et al. in [12], outliers, which were daily average points coded

1, were discarded before further analysis. Hourly data on days that failed the HelioClim QC test were removed before any further analysis. The percentage of outliers removed from each station is given in the last column in Table 3.1. Solar zenith angles were calculated using the solar position algorithm (SPA) on Python PV_LIB [13,14].

Hourly sunshine duration data were obtained by determining the burn made by the sun on a coated card in a Campbell–Stokes sunshine recorder [15]. Hourly data were then summed to obtain total daily sunshine duration (n).

3.3. Performance Evaluation Criteria (Statistical Metrics)

The statistical metrics that were used to quantify the difference between estimated *GHI* data with the observed (reference) *GHI* data were derived from the literature [16-19] and these are:

1. Mean Bias Error (MBE), which estimates the average error in the prediction. A positive MBE indicates that the prediction is overestimated and vice versa; the lower values of MBE indicate a strong correlation between the prediction and observation. A relative Mean Bias Error (rMBE), which measures the size of the error in percentage terms, was also calculated. The metrics are expressed as:

$$MBE = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (3.1)$$

$$rMBE = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (3.2)$$

2. Mean Absolute Error (MAE), which measures the absolute value of the differences between the observed and the predicted values, gives a better idea of the prediction accuracy; relative Mean Absolute Error (rMAE), which measures the size of the error in percentage terms, was also calculated. The caution with MBE and rMBE is with the cancelling of positive and negative bias, which can lead to a false interpretation. The metrics are expressed as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (3.3)$$

$$rMAE = 100 * \frac{1}{n} \sum_{i=1}^n \frac{|P_i - O_i|}{\bar{O}_i} \quad (3.4)$$

3. Root Mean Square Error (RMSE), which compares the predicted and observed datasets, measures the statistical variability of the prediction accuracy, and is expressed as shown in Equation (3.5), while Equation (3.6) shows the relative Root Mean Square Error (rRMSE), which measures the size error in percentage terms. The RMSE and rRMSE are also indifferent to the direction of the error. They are considered in this study since these put extra weight on large errors. The metrics are expressed as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - P_o)^2} \quad (3.5)$$

$$rRMSE = \frac{100}{\bar{O}_i} * \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - P_o)^2} \quad (3.6)$$

4. Coefficient of Determination (R^2), which is a statistical measure of the strength of the relationship between the movement of predicted and observed. R^2 also measures how well the regression line represents the data. The value of R^2 is such that $0 \leq R^2 \leq 1$. The closer R^2 is to 1, the better the prediction. The metric is expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (P_i - \bar{O}_i)^2} \quad (3.7)$$

where O_i is the observation value, P_i is the estimated value, $\bar{O}_i$ is the average of the observation values, i is the time point, and n is the total number of points used.

References

1. SANS 204-2, The South African National Standard. Energy efficiency in buildings, part 2: The application of energy efficiency requirements for buildings with environmental control SABS Standards Division. 2008. https://zbook.org/read/cd12a_sans-204-energy-efficiency-in-buildings.html.
2. D.C.U.Conradie, (2012). South Africa's Climatic Zones: Today, Tomorrow, International Green Building Conference and Exhibition Future Trends and Issues Impacting on the Built Environment , July 25-26, 2012, Sandton, South Africa.
3. Mabasa, M.B.; Botai, J.; Ntsangwane, M.L. Update on the re-establishment of the south african weather services (SAWS) radiometric network in all six climatological regions and the quality of the data. In Proceedings of the South African Solar Energy Conference (SASEC), Blue Waters Hotel, KwaZulu-Natal, South Africa, 25–27 June 2018. Available online: https://www.sasec.org.za/full_papers/68.pdf (accessed on 20 January 2020).
4. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2. 2010. Available online: https://epic.awi.de/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 11 December 2019).
5. Roesch, A.; Wild, M.; Ohmura, A.; Dutton, E.G.; Long, C.N.; Zhang, T. Assessment of BSRN radiation records for the computation of monthly means. *Atmos. Meas. Tech.* 2011, 4, 339–354. <https://doi.org/10.5194/amt-4-339-2011>.
6. Driemel, A.; Augustine, J.; Behrens, K.; Colle, S.; Cox, C.; Cuevas-Agulló, E.; Denn, F.M.; Duprat, T.; Fukuda, M.; Grobe, H.; et al. Baseline surface radiation network (BSRN): Structure and data description (1992–2017). *Earth Syst. Sci. Data* 2018, 10, 1491–1501. <https://doi.org/10.5194/essd-10-1491-2018>.
7. Urraca, R.; Gracia-Amillo, A.M.; Huld, T.; Martinez-De-Pison, F.J.; Trentmann, J.; Lindfors, A.V.; Riihelä, A.; Sanz-Garcia, A. Quality control of global solar radiation data with satellite-based products. *Sol. Energy* 2017, 158, 49–62. <https://doi.org/10.1016/j.solener.2017.09.032>.
8. Mabasa, B.; Lysko, M.D.; Tazvinga, H.; Mulaudzi, S.T.; Zwane, N.; Moloi, S.J. The ångström–prescott regression coefficients for six climatic zones in South Africa. *Energies* 2020, 13, 5418. <https://doi.org/10.3390/en13205418>.
9. Thomas, C.; Wey, E.; Blanc, P.; Wald, L. Validation of three satellite-derived databases of surface solar radiation using measurements performed at 42 stations in Brazil. *Adv. Sci. Res.* 2016, 13, 81–86. <https://doi.org/10.5194/asr-13-81-2016>.

10. Mabasa, B.; Lysko, M.; Tazvinga, H.; Zwane, N.; Moloi, S. The performance assessment of six global horizontal irradiance clear sky models in six climatological regions in South Africa. *Energies* **2021**, *14*, 2583. <https://doi.org/10.3390/en14092583>.
11. Riihelä, A.; Kallio, V.; Devraj, S.; Sharma, A.; Lindfors, A.V. Validation of the SARAHE satellite-based surface solar radiation estimates over India. *Remote Sens.* **2018**, *10*, 392. <https://doi.org/10.3390/rs10030392>.
12. Geiger, M.; Diabaté, L.; Ménard, L.; Wald, L. A web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy* **2002**, *73*, 475–480. [https://doi.org/10.1016/s0038-092x\(02\)00121-4](https://doi.org/10.1016/s0038-092x(02)00121-4).
13. Reda, I.; Andreas, A. Solar position algorithm for solar radiation applications. *Sol. Energy* **2004**, *76*, 577–589. <https://doi.org/10.1016/j.solener.2003.12.003>.
14. Holmgren, W.F.; Hansen, C.W.; Mikofski, M.A. Pylib python: A python package for modeling solar energy systems. *J. Open Source Softw.* **2018**, *3*, 884. <https://doi.org/10.21105/joss.00884>.
15. Singh, J.; Kruger, A. Is the summer season losing potential for solar energy applications in South Africa? *J. Energy South. Afr.* **2017**, *28*, 52, doi:10.17159/2413-3051/2017/v28i2a1673.
16. Gueymard, C.A. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renew. Sustain. Energy Rev.* **2014**, *39*, 1024–1034. <https://doi.org/10.1016/j.rser.2014.07.117>.
17. Fernández-Peruchena, C.M.; Gastón, M. A simple and efficient procedure for increasing the temporal resolution of global horizontal solar irradiance series. *Renew. Energy* **2016**, *86*, 375–383. <https://doi.org/10.1016/j.renene.2015.08.004>.
18. Maupin, K.A.; Swiler, L.P.; Porter, N.W. Validation Metrics for Deterministic and Probabilistic Data. *J. Verificat. Valid. Uncertain. Quantif.* **2018**, *3*, 031002, doi:10.1115/1.4042443.
19. Paulescu, M.; Stefu, N.; Calinoiu, D.; Tulcan-Paulescu, E.; Pop, N.; Boata, R.; Mares, O. Ångström–Prescott equation: Physical basis, empirical models and sensitivity analysis. *Renew. Sustain. Energy Rev.* **2016**, *62*, 495–506, doi:10.1016/j.rser.2016.04.012.

Chapter 4: Calibration and Validation of Clear sky Models Six climatic zones in South Africa

4.1. Introduction

This chapter aims to answers a problem statement:

1. Which clear sky model performs better in estimating minute global horizontal irradiance (GHI) in each location/station under study and in each macro-climatological region of South Africa? Which GHI clear sky model is better overall in estimating minute clear sky GHI in South Africa?

The problem statement was addressed by setting the following objectives:

1. Calibrate (using meteorological input parameters) and validate six global horizontal irradiance (GHI) clear sky models namely: Bird, Simple Solis, McClear, Ineichen–Perez, Haurwitz and Berger–Duffie against GHI data recorded by a reference pyranometer only on clear sky days, to determine the overall the best performing clear sky model for sites in South Africa.

The work presented in this chapter has submitted on 22 March 2021 to the Energies Journal (ISSN 1996-1073; CODEN: ENERGA), a peer-reviewed open access journal of related scientific research, technology development, engineering, and the studies in policy and management. The Journal publishes scientific research work under the Multidisciplinary Digital Publishing Institute (MDPI) publisher with its headquarters based in Basel, Switzerland. The manuscript was reviewed by experts in the field and accepted on 14 April 2021 by the Journal and it was published on 30 April 2021. The work presented herein according to the submission guidelines suggested by the Journal editors. The Energies Journal

(ISSN 1996-10730) is accredited by the Department of Higher Education and Training (DHET).

Mabasa, B.; Lysko, M.D.; Tazvinga, H.; Zwane, N.; Moloi, S.J. The Performance Assessment of Six Global Horizontal Irradiance Clear Sky Models in Six Climatological Regions in South Africa. *Energies* **2021**, *14*, 2583. <https://doi.org/10.3390/en14092583>.

Abstract: This study assesses the performance of six global horizontal irradiance (GHI) clear sky models, namely: Bird, Simple Solis, McClear, Ineichen–Perez, Haurwitz and Berger–Duffie. The assessment is performed by comparing 1-minute model outputs to corresponding clear sky reference 1-minute Baseline Surface Radiation Network quality controlled GHI data from 13 South African Weather Services radiometric stations. The data used in the study range from 2013 to 2019. The 13 reference stations are across the six macro climatological regions of South Africa. The aim of the study is to identify the overall best performing clear sky model for estimating minute GHI in South Africa. Clear sky days are detected using ERA5 reanalysis hourly data and the application of an additional 1-minute automated detection algorithm. Metadata for the models' inputs were sourced from station measurements, satellite platform observations, reanalysis and some were modelled. Statistical metrics relative Mean Bias Error (rMBE), relative Root Mean Square Error (rRMSE) and the coefficient of determination (R^2) are used to categorize model performance. The results show that each of the models performed differently across the 13 stations and in different climatic regions. The Bird model was overall the best in all regions, with an rMBE of 1.87 %, rRMSE of 4.11 % and R^2 of 0.998. The Bird model can therefore be used with quantitative confidence as a basis for solar energy applications when all the required model inputs are available.

Author Contributions: The contribution of the authors is summarised below.

Conceptualization: Brighton Mabasa. **Methodology:** Brighton Mabasa and Meena D. Lysko. **Software:** Brighton Mabasa. **Validation:** Meena D. Lysko. **Formal analysis,** Brighton Mabasa and Meena D. Lysko. **Investigation:** Brighton Mabasa. **Data curation:** Brighton Mabasa. **Writing—original draft preparation:** Brighton Mabasa. **Supervision:** Meena D. Lysko and Sabata J. Moloi. **Writing—review and editing:** Brighton Mabasa; Meena D. Lysko; Sabata J. Moloi; Henerica Tazvinga and Nosipho Zwane. All authors have read and agreed to the published version of the manuscript.

Clear sky irradiance data are important for various applications. There are no sensors to measure clear sky irradiance nor a universal model; the only way to generate clear sky irradiance data is by filtering in situ readings taken from pyranometers for periods when there are no clouds or by trying different models at different locations. Clear sky models are used to estimate the amount of solar irradiance that reaches the Earth's surface under cloudless atmospheres. The global horizontal irradiation (GHI) clear sky radiation model is an algorithm that is solved by providing inputs that describe the state of the cloudiness atmosphere at a specific location and time to output theoretical GHI at that location and time [1]. Various models have been developed, improved and validated to estimate clear sky GHI data. Even though there are various models developed there are still challenges because the optical transparency of the atmosphere, physics of the model inputs and the effects thereof are still not fully understood [2], making it difficult to accurately model clear sky GHI. The dynamic nature of the atmosphere results in biases when the modelled data are compared to data recorded using a reference pyranometer under clear skies. To determine the most suitable model at an area validation of different models was emphasised by Badescu et al., 2013 [3] and Antonanzas-Torres et al., 2019 [4].

The chapter aims to validate six GHI clear sky models, namely: Bird (1981) [5,6], Simple Solis (2008) [7,8], McClear (2013) [9], Ineichen–Perez (2002) [10,11], Haurwitz (1945) [12,13] and Berger–Duffie (1979) [14], against in-situ data from 13 South African Weather Services (SAWS) reference stations and across the diverse macro-climates in South Africa. This validation study covers more stations and uses more clear sky days than previous studies in South Africa [15-18] this enables a fuller assessment of clear sky models. The models used in the study were chosen based on the availability of inputs, implementability with open source (Python, from Amsterdam in the Netherlands), and their presence in the literature to enable the comparison with other validation results. The end goal is to determine the most suitable model at each of the 13 sites, 6 climatological zones and overall, best performing in South Africa.

4.2. Description of the Six Models in this study

4.2.1. McClear Clear Sky Model

The Copernicus Atmosphere Monitoring Service (CAMS) McClear clear sky model, as reported by Lefevre et al., 2013 [9], is a full physical model that was developed by Mines Paris Tech. The model uses lookup tables produced by the libRadtran radiative transfer model (RTM) to predict irradiance. It is a “blackbox” model, with no algorithm that shows how inputs are integrated to be able to estimate irradiance. The McClear model uses solar zenith angle, ground albedo, ground elevation, aerosol optical depth (AOD_{550}), angstrom coefficients, aerosol type, total column of ozone, total column of water vapour, vertical profile of temperature, pressure, density and volume mixing ratio for gases. The model outputs GHI, direct horizontal irradiance (DHI), diffuse horizontal irradiance (DIF) and direct normal irradiance (DNI) all under clear skies in addition to the irradiance atmospheric parameters used are also provided [19,20]. The McClear model provides data freely from 2014 to date with 2

days delay. The temporal resolution of the data ranges from minute averages, 15-min averages, hourly averages, daily averages and monthly averages. The McClear clear sky model has global coverage and the spatial resolution interpolated to the user's point of interest. The model is monitored and improved regularly. Notable improvements were reported by Gschwind et al., 2019 [21], where the McClear-v2 was modified to McClear-v3.

4.2.2. Berger–Duffie Clear Sky Model

The Berger–Duffie clear sky model is an algorithm that is cited in [14,22]. It only requires direct normal extra-terrestrial irradiance (DNI_{TOA}) and zenith angle (θ_Z) of the site where GHI is estimated. All the transmittances have been unified to a constant 0.7. The estimation of GHI is given by Equation (4.1) below:

$$GHI = DNI_{TOA} 0.7 \cos \theta_Z \quad (4.1)$$

4.2.3. Haurwitz Clear Sky Model

The Haurwitz clear sky model was developed by using the relationship between cloudiness, cloud density and solar air mass data collected at Blue Hill Observatory (Harvard University) from 1933 to 1943 [12,13]. The coefficients a and b were derived at that location under clear sky conditions. GHI is calculated as:

$$GHI = \frac{a}{AM} e^{-b \cdot AM} = 1098 \cos \theta_Z e^{\frac{-0.057}{\cos \theta_Z}} \quad (4.2)$$

where AM is the relative air mass and a , b are constants which depend on the cloudiness and cloud density for cloudless conditions [22]. The model only requires the θ_Z of the site where

GHI is estimated. According to [22], the model was found to have the best performance among models which require only θ_Z . The Haurwitz clear sky model is included in this study to investigate how the model performs in different zenith angles and how it compares with complex models.

4.2.4. Ineichen–Perez Clear Sky Model

The Ineichen–Perez model [10,11] was developed based on the Ineichen model (1983) [23], where 12 clear sky days in Geneva were used to develop a model to estimate DNI under a clear sky. A fixed Linke Turbidity (TL) of 3 was used and the clear sky irradiance model was developed based exclusively on the air mass value. The model was then further developed to be able to estimate GHI under a clear sky and to be able to fit the observed data recorded from high altitude areas. The new development to the algorithm was achieved by adjusting the Kasten (1984) [24] model's altitude coefficients (a_1 and a_2). The original Kasten equation is given by Equation (4.3) below:

$$GHI = 0.84DNI_{TOA} \cos \theta_Z e^{[-0.027AM(f_{h1}+f_{h2}(T_L-1))]} \quad (4.3)$$

With the inclusion of the two altitude dependent coefficients, a_1 and a_2 , the model is given by Equation (4.4) below:

$$GHI = a_1DNI_{TOA} \cos \theta_Z e^{[-a_2AM_a(f_{h1}+f_{h2}(T_L-1))]} \quad (4.4)$$

With

$$a_1 = 5.09 \cdot 10^{-5}h + 0.868 \quad (4.5)$$

$$a_2 = 3.92 \cdot 10^{-5}h + 0.0387 \quad (4.6)$$

$$f_{h1} = e^{\left(-\frac{h}{8000}\right)} \quad (4.7)$$

$$f_{h2} = e^{\left(-\frac{h}{1250}\right)} \quad (4.8)$$

where AM_a =absolute atmospheric air mass, h =altitude of the site, and f_{h1} and f_{h2} are the coefficients that relate the site altitude with the altitude of the atmospheric interactions. The Ineichen–Perez clear sky model is commonly used in the retrieval of GHI from satellite data

4.2.5. Bird Clear Sky Model

The Bird clear sky model was developed by Bird and Hulstrom (1981) [5] and cited by Daryl (2013) [6]. The algorithm modified the DNI_{TOA} by adding correction terms for the constituents in the atmosphere that attenuate irradiance on its way from the Sun to the Earth's surface. In addition to these constituents, the variation in the day of the year (D) and the Sun–Earth distance was also regarded as a contributing factor and its effect was accounted for. The atmospheric constituents are the transmittance after aerosol absorptance and scattering (T_A), transmittance after ozone absorptance in the stratosphere (T_O), transmittance after water vapor absorptance (T_W), transmittance after absorptance by uniformly mixed gases (T_{UM}), and transmittance after Rayleigh scattering (T_R). Taking the transmittance factors and the Sun–Earth distance variation due to the day of the year into consideration, an algorithm was derived to estimate DNI on the Earth's surface under cloudless skies. An algorithm to estimate scattered irradiance on the Earth's horizontal surface (DIF) was also derived, and the sum of the two algorithms can be used to estimate GHI. The result is given in Equations (4.9)– (4.13):

$$GHI = \frac{(DNI \cos \theta_Z + DIF)}{(1 - R_g R_s)} \quad (4.9)$$

where R_g is the ground albedo, R_s is the sky albedo, and

$$DNI = 0.9662DNI_{TOA}T_RT_WT_AT_OT_{UM}, \quad (4.10)$$

$$DIF = 0.79DNI_{TOA} \cos \theta_Z T_OT_{UM}T_WT_{AA} \frac{0.5(1-T_R)+B_A(1-T_{AS})}{(1-AM+AM^{1.02})}, \quad (4.11)$$

$B_A = 0.5(1 + \cos \theta_Z)$ is the asymmetric factor or the forward scattering ratio
(4.12)

$$T_{AS} = \frac{T_A}{T_{AA}} \quad \text{is the transmittance of aerosol scattering,} \quad (4.13)$$

with T_{AA} as the transmittance of aerosol absorbance.

4.2.6. Simple Solis Clear Sky Model

The Simple Solis clear sky model is an algorithm that is based on RTM and the Lambert–Beer relation to estimating irradiance [7,8]. The model was simplified to reduce the computational requirements, the time needed to run the model, and to reduce the number of measured inputs required to run the model. The simplified model requires water vapor (P_w) and aerosol optical depth (AOD_{700}) as the main input parameters while retaining a measure of accuracy in estimating solar radiation at the Earth’s surface under clear sky conditions.

The Lambert–Beer relation according to Ineichen (2008) [8] is given by Equation (4.14) below:

$$DNI = DNI_{TOA}e^{-AM\tau} \quad (4.14)$$

where τ is the total atmospheric optical depth.

The above Equation (4.14) is for monochromatic radiation. For a real case with different spectral transmission in the path of irradiance to the Earth’s surface, the air mass was modified by adding a fitting parameter, b . The adapted algorithm is called the Modified Lambert–Beer relation according to Muller et al., 2004 [24]. It is given by Equation (4.15) below:

$$DNI = DNI_{TOA} e^{\left(\frac{-\tau}{\cos \theta_z^b}\right)}, \quad (4.15)$$

where b is the constant of adjustment for DNI.

As provided by Ineichen (2008) [8], when estimating GHI the Lambert–Beer relation is not applicable anymore because of back scattering effects. To consider different altitudes and atmospheric conditions, the DNI_{TOA} was modified to the enhanced direct normal extra-terrestrial solar radiation (DNI'_{TOA}), as shown by Equation (4.16), from Ineichen (2008) [8]:

$$DNI'_{TOA} = DNI_{TOA} \left[0.12P_w^{0.56} \cdot AOD_{700nm}^2 + 0.97P_w^{0.032} \cdot AOD_{700nm} + 1.08P_w^{0.0051} + 0.017 \ln \left(\frac{p}{1013} \right) \right], \quad (4.16)$$

where p is the surface pressure.

The GHI function is given by Equation (4.17) [7,8]:

$$GHI = DNI'_{TOA} e^{\left(\frac{-\tau_g}{\cos \theta_z^g}\right)} \cdot \cos \theta_z, \quad (4.17)$$

where the GHI total optical depth, with

$$\tau_g = t_{g1} + AOD_{700} + t_{g0} + t_{gp} + \ln \left(\frac{p}{1013} \right),$$

where,

$$t_{g1} = 1.24 + 0.047 \ln(P_w) + 0.0061 + \ln(P_w)^2, \quad (4.18)$$

$$t_{g0} = 0.27 + 0.043 \ln(P_w) + 0.0090 + \ln(P_w)^2, \quad (4.19)$$

$$t_{gp} = 0.0079P_w + 0.1, \quad (4.20)$$

and the GHI fitting parameter g is obtained from RTM computations as

$$g = -0.0147 \ln(w) - 0.3079AOD_{700nm}^2 + 0.2846AOD_{700nm} + 0.3798 \quad (4.21)$$

Since $\cos \theta_z = \sin \alpha_s$, where α_s is the solar elevation, the Equation (4.17) can be expressed as:

$$GHI = DNI'_{TOA} e^{\left(\frac{-\tau_g}{\sin \alpha_z^g}\right)} \cdot \sin \alpha_s \quad (4.22)$$

4.3. Study Area

The study area was determined by considering the locations of the South African Weather Services (SAWS) radiometric stations which measure the 1-minute GHI data. The area of study is thirteen sites located in all six macro-climate regions in South Africa, as described in Chapter 3 (Subsections 3.1 and 3.2).

4.3.1. Data

The various clear sky models require different input parameters. Some input parameters are common. All the input parameters required by the six different models are summarised in Table 4.1.

Table 4.1. Model and input parameters required.

Input/Model	Bird	Simple Solis	Ineichen–Perez	McClea	Haurwitz	Berger–Duffie
Zenith Angle	X	X	X	X	X	X
Albedo	X	-	-	X	-	-
AOD1240nm	X	X	-	-	-	-
AOD550nm	X	X	-	X	-	-
AOD380nm	X	-	-	-	-	-
AOD500nm	X	-	-	-	-	-
AOD700nm	-	X	-	-	-	-
Temperature	X	X	-	X	-	-
Humidity	X	X	-	X	-	-
DNI_{TOA}	X	X	X	X	-	X
D (Julian day)	X	X	X	X	X	X
1367(solar constant)	X	X	X	X (1362)	-	-
Pressure	X	X	X	X	-	-
Altitude	X	X	X	X	X	X
Linke Turbidity	-	-	X	-	-	-
Ozone	X	-	-	X	-	-
Absolute airmass	-	-	X	-	-	-
Relative airmass	X	-	X	-	-	-
Apparent Elevation	-	X	-	-	-	-
Asymmetry	X	-	-	-	-	-
Total inputs	16	13	9	12	3	4

4.3.1.1. Observation Data

The observed 1-minute GHI data used in this study were collected from 13 SAWS solar radiometric station, as described in Chapter 3 (Subsections 3.1 and 3.2). Hourly average values of temperature, humidity and pressure from areas corresponding to GHI measures were also collected.

4.3.1.2. Surface Albedo Data

Surface albedo (R_g) quantifies the fraction short wave of irradiance which is reflected away from the surface. R_g data were acquired from the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)'s Satellite Application Facility on Climate Monitoring (CM SAF). EUMETSAT's CM SAF is described by Karlsson et al., 2020 [25]. The surface albedo data were acquired as monthly average values at a spatial resolution of $0.25^\circ \times 0.25^\circ$. The data, according to Karlsson et al., 2020 [25], were derived from the Advanced Very High-Resolution Radiometer (AVHRR) sensor onboard polar orbiting National Oceanic and Atmospheric Administration (NOAA) and METOP satellites and they were validated using the BSRN ground stations as a reference.

4.3.1.3. Cloud Data

Total cloud cover, which is the fraction of a sky that is covered by liquid or ice, was acquired from ERA5. ERA5 is the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for the global climate and weather and is described by Hersbach et al., 2018 [26]. Hourly data for a grid box with a spatial resolution of $0.25^\circ \times 0.25^\circ$ were extracted for the relevant points and periods.

4.3.1.4. Atmospheric Ozone

Atmospheric ozone (O_3) absorbs incoming solar radiation. As a result of ozone attenuation, no irradiance reaches the Earth's surface at wavelengths shorter than 295–300 nm. The Earth's surface radiation at a wavelength between 300 and 315 nm is affected by the amount of O_3 in the atmosphere, according to Fredrick (2015) [27]. The atmospheric ozone data used in the study were acquired from Solar Radiation Data (SoDa) [28] as mean atmospheric ozone stratospheric monthly climatological averages interpolated per site of the study. The monthly averages were calculated using values from 1960 to 2000, and the values were recorded using a Global Ozone Monitoring Experiment (GOME) in Dobson units, SoDa [28].

4.3.1.5. Linke Turbidity

The Linke Turbidity factor (TL) is described by Diabaté et al. (2003) [29] as the optical thickness of the atmosphere due to the attenuation of atmospheric constituents relative to a dry and clean atmosphere. TL is used to estimate the atmospheric attenuation of solar radiation under clear skies. TL would be equal to 1 if the atmosphere was clear and dry. The value of TL increases with an increase in the attenuation of the radiation by the clear sky atmosphere. The TL values used in the study were monthly mean climatological values and the data were extracted from SoDa [30] and interpolated for each point of study; these were TL values for an air mass 2.

4.3.1.6. Aerosol Optical Depth

Aerosol optical depth (AOD) is the coefficient of attenuation of irradiance by the atmosphere. AOD data at 550 nm (AOD_{550}) and at 1240 nm (AOD_{1240}) were acquired from Copernicus

Atmosphere Monitoring Service (CAMs-AOD) [30], previously known as MACC-AOD. The data are provided by ECMWF from 2014 to date with 2 days delay. AOD data at 380 nm (AOD_{380}), 500 nm (AOD_{500}) and 700 nm (AOD_{700}) were not available, but they are required as an input to the models. The Angstrom turbidity law [31,32] was applied to derive AOD_{380} , AOD_{500} and AOD_{700} . This was performed by first calculating the Angstrom exponent (α) of the aerosol from the two AOD wavelengths and then using the calculated α from those two wavelengths to calculate the unknown AOD. For this, Equations (4.23)–(4.27) from [31,32] were used:

$$\frac{\tau_{\lambda}}{\tau_{\lambda_0}} = \left(\frac{\lambda}{\lambda_0}\right)^{-\alpha} \Rightarrow \tau_{\lambda} = \tau_{\lambda_0} \left(\frac{\lambda}{\lambda_0}\right)^{-\alpha} \quad (4.23)$$

$$\tau_{700nm} = \tau_{\lambda_0} \left(\frac{700nm}{\lambda_0}\right)^{-\alpha} \quad (4.24)$$

$$\tau_{500nm} = \tau_{\lambda_0} \left(\frac{500nm}{\lambda_0}\right)^{-\alpha} \quad (4.25)$$

$$\tau_{380nm} = \tau_{\lambda_0} \left(\frac{380nm}{\lambda_0}\right)^{-\alpha} \quad (4.26)$$

Where,

τ_{λ} is the optical thickness at an unknown wavelength (AOD_{380} , AOD_{500} and AOD_{700}).

τ_{λ_0} is the optical thickness at the known wavelength (AOD_{550} or AOD_{1240}).

λ is the wavelength of the unknown AOD (AOD_{380} , AOD_{500} and AOD_{700}).

λ_0 is the wavelength of the known AOD (AOD_{550} or AOD_{1240}).

α is the angstrom or alpha exponent of the aerosol and it is related to the size distribution of the particles.

The alpha exponent or Angstrom of the aerosol (α) is calculated in Equation (4.27):

$$\alpha = -\log \left[\left(\frac{\tau_{\lambda_1}}{\tau_{\lambda_2}} \right) / \left(\frac{\lambda_1}{\lambda_2} \right) \right] \quad (4.27)$$

where $\frac{\tau_{\lambda_1}}{\tau_{\lambda_2}}$ is the ratio between measurements of known optical thickness, $\frac{AOD_{550}}{AOD_{1240}}$, and $\frac{\lambda_1}{\lambda_2}$ is the ratio between the known wavelengths, $\frac{550nm}{1240nm}$.

4.3.1.7. Water Vapor

Daily water vapor or precipitable water (P_w) in (cm) was estimated with an empirical model validated and applied in the following literature [33–35] and which uses the relationship between air temperature in °C and relative humidity in %. The formulation is given by Equations (4.28)– (4.32) below:

$$P_w = 0.1 \cdot H_v \cdot P_v \quad (4.28)$$

where H_v is the apparent water vapor scale height (in km) and P_v is the surface water vapor density (in $\frac{g}{m^3}$). These are given by Equations (5.29) and (5.31), respectively.

$$H_v = 0.4976 + 1.5265 \cdot X + e^{(13.68 \cdot X - 14.9188 \cdot X^3)} \quad (5.29)$$

$$\text{With, } X = \frac{T}{273.15} \quad (4.30)$$

$$P_v = 216.7 \cdot RH \cdot \frac{e_s}{T}, \quad (4.31)$$

With

$$e_s = e^{\left(22.33 - \frac{4914}{T} - \frac{109220}{T^2} - 0.0039015 \cdot T\right)} \quad (4.32)$$

and with T the temperature (in °C), and RH the relative humidity (in %).

4.3.1.8. Solar Geometry

Solar geometry described in Chapter 2 (subsection 2.2 -2.4)

4.3.2. Methodology

The methodology used in this study is summarized in the flow chart in Figure 4.1 below. The methodology consists of eight steps as follows: 1. Detection of clear sky days using the ERA5 hourly reanalysis data set [26]; 2. Pre-processing of observation data and BSRN quality check (as described in section 3.2); 3. Identification of 1-min clear sky periods; 4. Configuration and running models on clear sky days; 5. Application of the additional clear sky periods detection method [36]; 6. Matching common time steps of observation and modelled data; 7. Division of the common time steps into 5 degree zenithal angles; and 8. Calculation of statistical metrics for each 5 degree zenith bin.

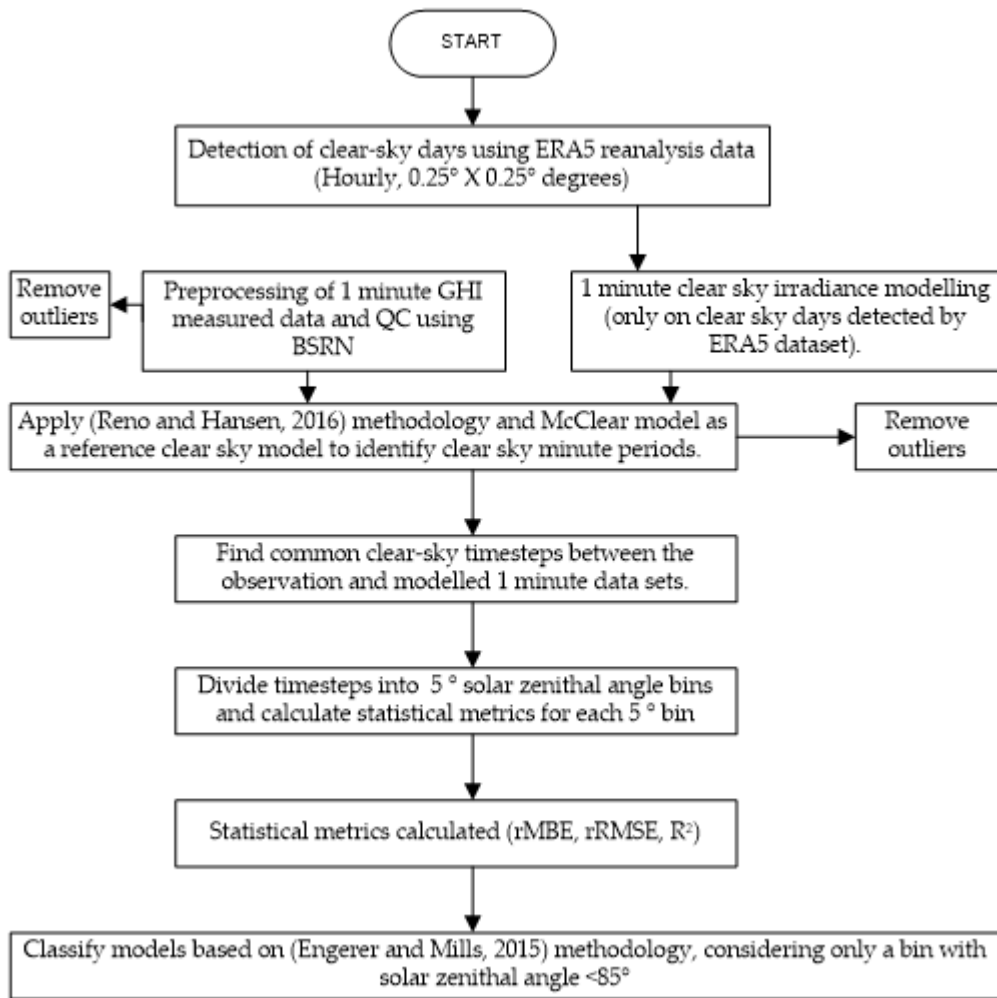


Figure 4.1. The flowchart of the approach summarising the methodology from clear sky detection to model classification based on their performance. {Fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5); Global horizontal irradiance (GHI); Quality control (QC); Baseline Surface Radiation Network (BSRN)}.

4.3.3. Statistical Metrics

The statistical metrics that were used to compare estimated 1-minute clear sky GHI data with the observed 1-minute GHI data were relative Mean Bias Error ($rMBE$), relative Root Mean Square Error ($rRMSE$) and Coefficient of Determination (R^2). Described in subsection 4.3. The model performance is gauged by using a classification methodology used by Engerer et al.

2015 [37], and which was also applied by [17,38]. The classification limits for a model's skill are summarised in Table 4.2.

Table 4.2. Four performance categories to classify clear sky GHI models. The colours in the table are used to visualize the model skill in addition to statistical thresholds with blue representing excellent performance, green good performance, yellow average performance and pink poor performance

Model Skill	rMBE	rRMSE	R²
Poor	$\geq \pm 10 \%$	$\geq 15\%$	≤ 0.97
Average	$\geq \pm 5 \%, < \pm 10 \%$	$\geq 10\%, < 15\%$	$> 0.97, \leq 0.98,$
Good	$\geq \pm 2 \%, < \pm 5 \%$	$0 \geq 5\%, < 10\%$	$> 0.98, \leq 0.99,$
Excellent	$< \pm 2 \%$	$< 5\%$	> 0.99

4.4 Results and Discussions

4.4.1. Clear Sky Detection

The ERA5 hourly reanalysis data set can also be used to detect cloud sky days—even though one of the clear sky model's applications is to determine the upper bound of GHI, it is worth knowing that in some cases when there are clouds the measured instantaneous GHI can surpass the clear sky boundary but it cannot surpass the GHI_{TOA} . According to Dazhi et al., 2015 [39], the reason behind this is the reflection from the clouds, and according to Long and Shi (2008) [40], it is because of the enhancement of the DIF and then the DNI not being blocked by clouds for short periods. This is demonstrated below by Figure 4.2, which shows the measured data collected from two stations (De Aar and Upington stations on 11 November 2014), even when GHI is calculated from the sum of $DNI \cdot \cos \theta_z$ and DIF, i.e., $(GHI = DNI \cdot \cos \theta_z + DIF)$.

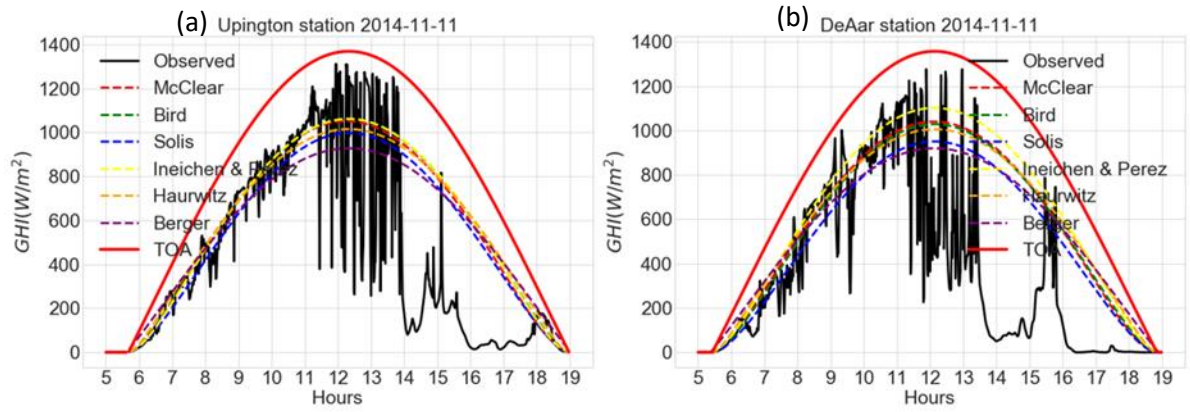


Figure 4.2. A plot of observed radiation data on 11 November 2014 at Uppington station (a), and De Aar station (b), 6 clear models and extra-terrestrial radiation on a horizontal surface on what ERA5 data classified as a cloudy day.

An additional clear sky detection methodology, such as that by Reno and Hansen 2016 [36], is needed to remove outliers other than clouds. The Reno and Hansen method has been applied in this study to detect observation errors that are not picked out automatically with the quality control procedures. As an illustration, Figure 4.3 a, b shows that the dent in the observation data plotted in Figure 4.3c, d was successfully removed with the application of the Reno Hansen method.

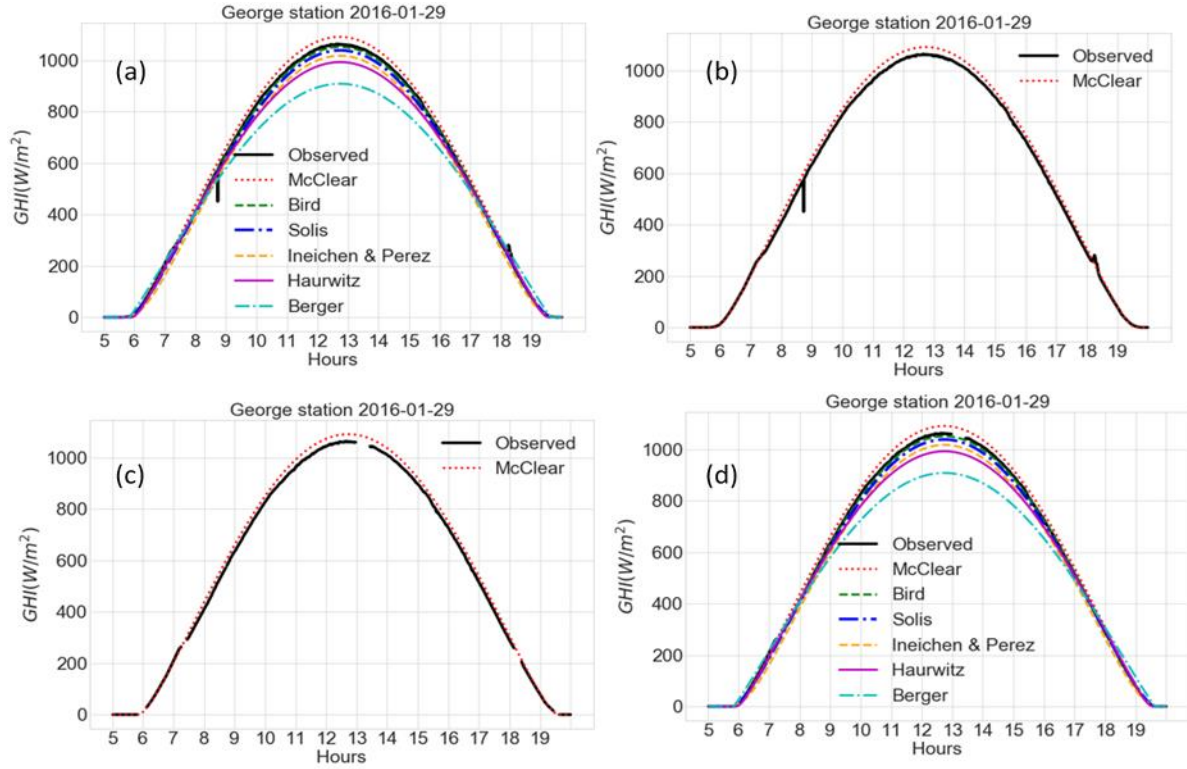


Figure 4.3: Illustration of Reno and Hansen methodology **(a)** Predicted irradiances with the 6 models and George station observed radiation on 29 January 2016 after only applying ERA5 reanalysis data set, **(b)** McClear and observed irradiance, **(c)** observed radiation after Reno and Hansen method is applied to detect not clear periods, **(d)** complete clear sky observed radiation and 6 clear sky models.

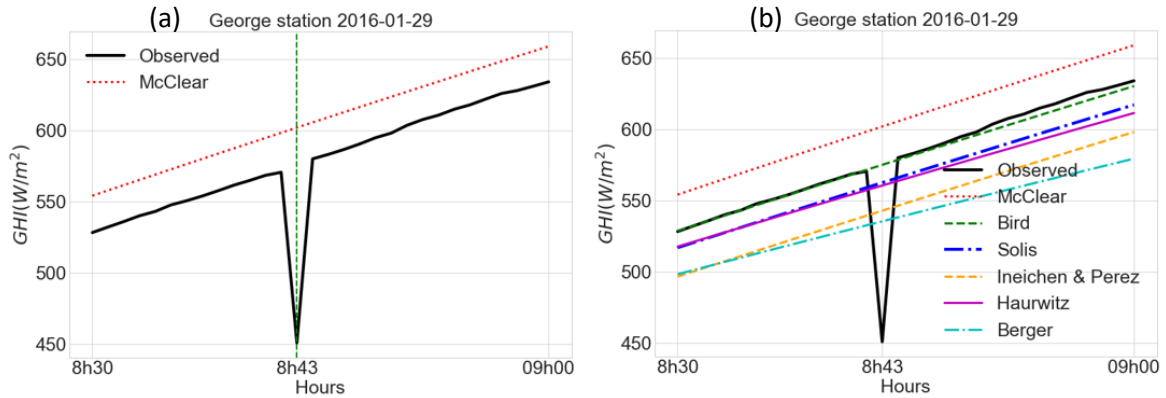


Figure 4.4: A plot of observed radiation data on 29 January 2016 from 08:30 to 09:00 at George station and McClear model on a corresponding period **(a)** and a plot of observed radiation data and 6 clear sky models on the corresponding period **(b)**.

From Figure 4.4 after zooming the period on 29 January 2016 from 08:30 to 09:00 at George station when the dent was noticed, it is clear that there was a disturbance or a dent, these types of dents and some that are not visible as dents with a naked eye are outliers that are detected by Reno and Hansen 2016 [36] and can be removed by the user.

4.4.2. Validation Results

After detecting clear sky times, a random set of days for each of the 13 stations is shown in Figures 4.5-4.7. All the models generated data that captured temporal variability but with varying accuracy. The full assessment of each model's performance per site is given by the statistics in Supplementary Tables S4.1–S4.6. Additionally, colormaps are provided in Figures 4.8 and 4.9 for the comparison of model performance per station and climate region. A summary of performance is given in Tables 4.3 and 4.4. Table 4.3 shows the optimum model w.r.t metrics for each station, whilst Table 4.4 shows the number of stations where a model shows an optimum metric. An optimum model per metric and station is found as follows: minimum $rMBE$, minimum $rRMSE$ and maximum R^2 . Further, scatter plots are presented in the study to show the correlation between measured and modelled clear sky GHI for each model and per station.

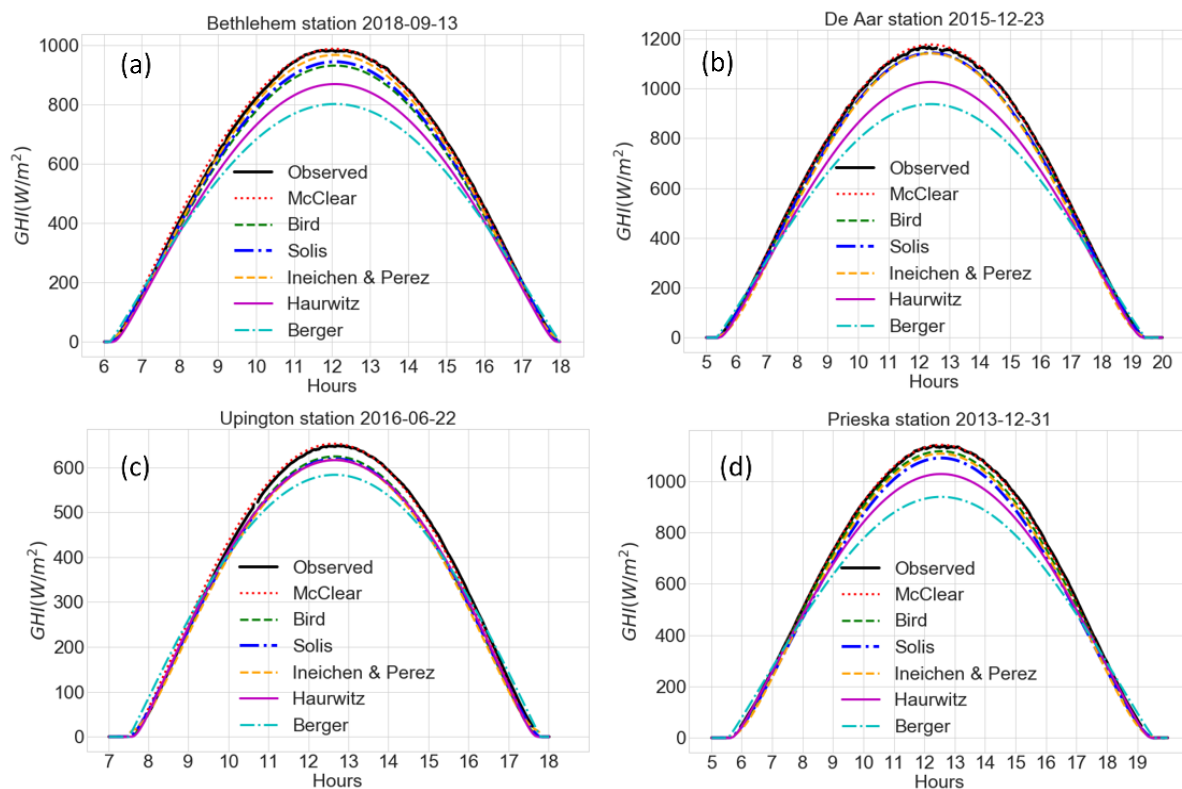


Figure 4.5 Comparison between observed and modelled minute GHI values in Bethlehem (a), De Aar (b), Upington (c), and Prieska (d).

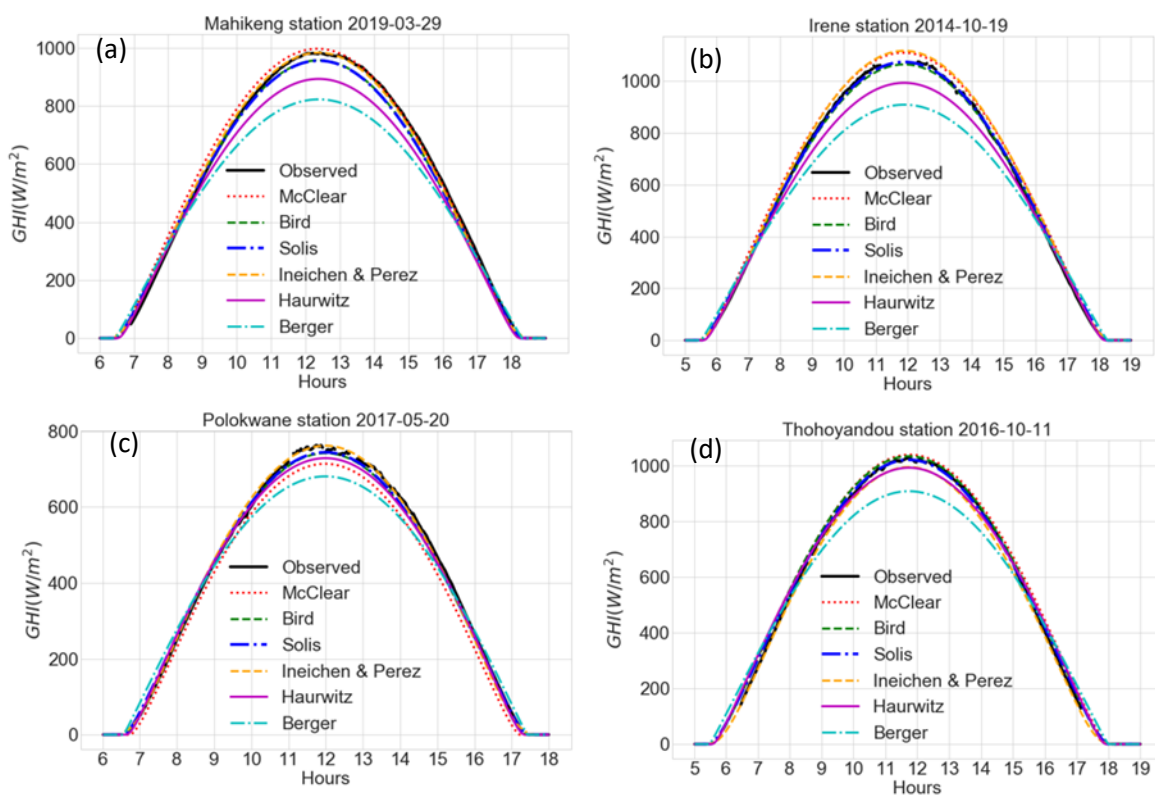


Figure 4.6. Comparison between observed and modelled minute GHI values in Mahikeng (a), Irene (b), Polokwane (c), and Thohoyandou (d).

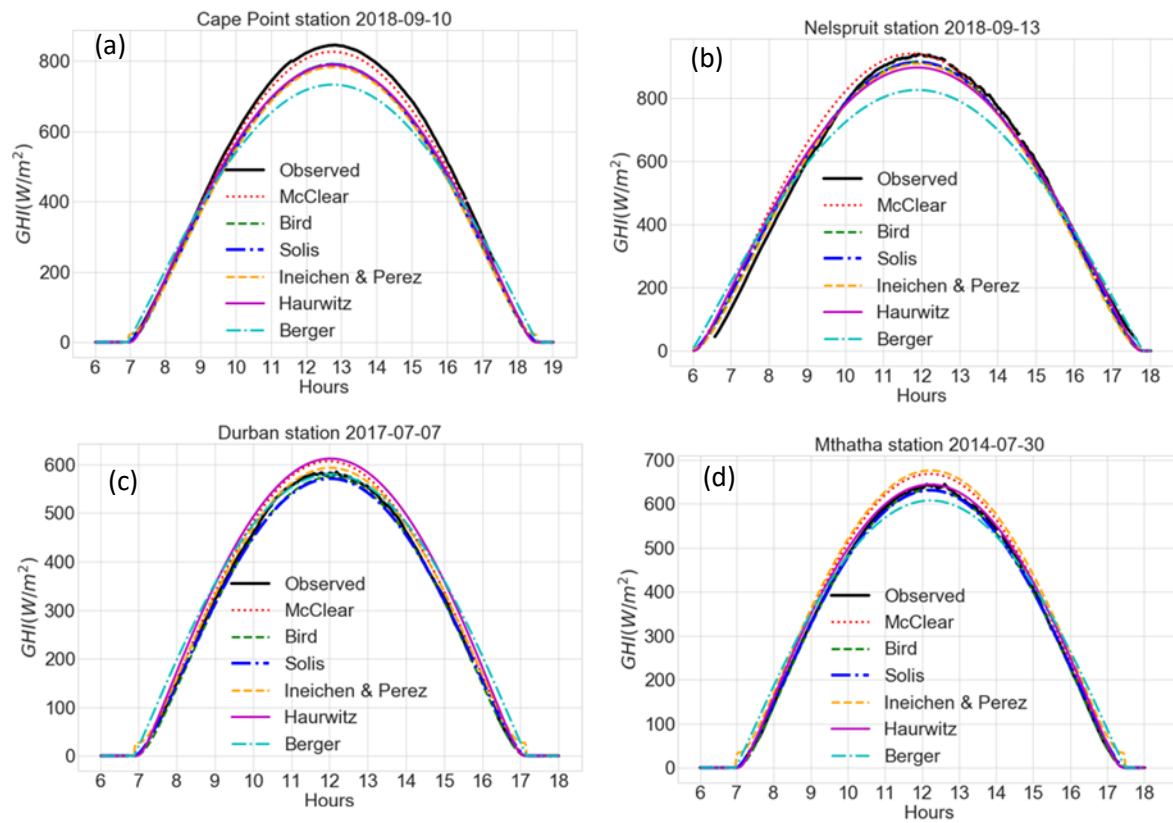


Figure 4.7. Comparison between observed and modelled minute GHI values in Cape Point (a), Nelspruit (b), Durban (c), and Mthatha (d).

Table 4.3. Optimum model w.r.t metrics for each station.

Station	Minimum rMBE	minimum rRMSE	Maximum R ²
Cape Point	Berger-Duffie	McCclear	McCclear
George	McCclear	McCclear	Bird
De Aar	Simple Solis	Bird	Bird
Bethlehem	Simple Solis	Simple Solis	Bird
Upington	McCclear	McCclear	McCclear
Prieska	Bird	Bird	Bird
Mthatha	Haurwitz	Bird	Bird
Durban	Bird	Bird	Bird
Mahikeng	Ineichen-Perez	Ineichen-Perez	McCclear

Irene	Simple Solis	Simple Solis	Bird
Polokwane	Simple Solis	Bird	Bird
Tohoyandou	Ineichen-Perez	Ineichen-Perez	McClea
Nelspruit	Simple Solis	Bird	Bird

Table 4.4. Indication of number of stations with an optimum metric.

Clear sky model	Number of Stations with Minimum rMBE	Number of Stations with Minimum rRMSE	Number of Stations with Maximum R^2
Berger-Duffie	1	0	0
Bird	2	6	9
McClea	2	3	4
Haurwitz	1	0	0
Ineichen-Perez	2	2	0
Simple Solis	5	2	0

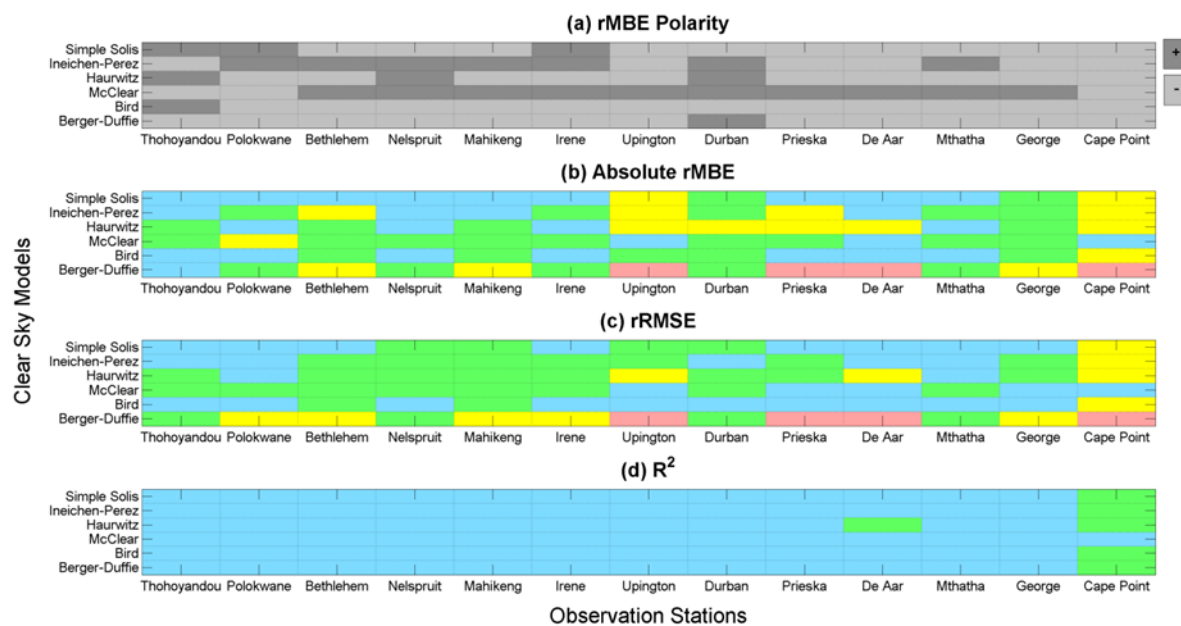


Figure 4.8. Overall performance of the six models at the 13 stations.

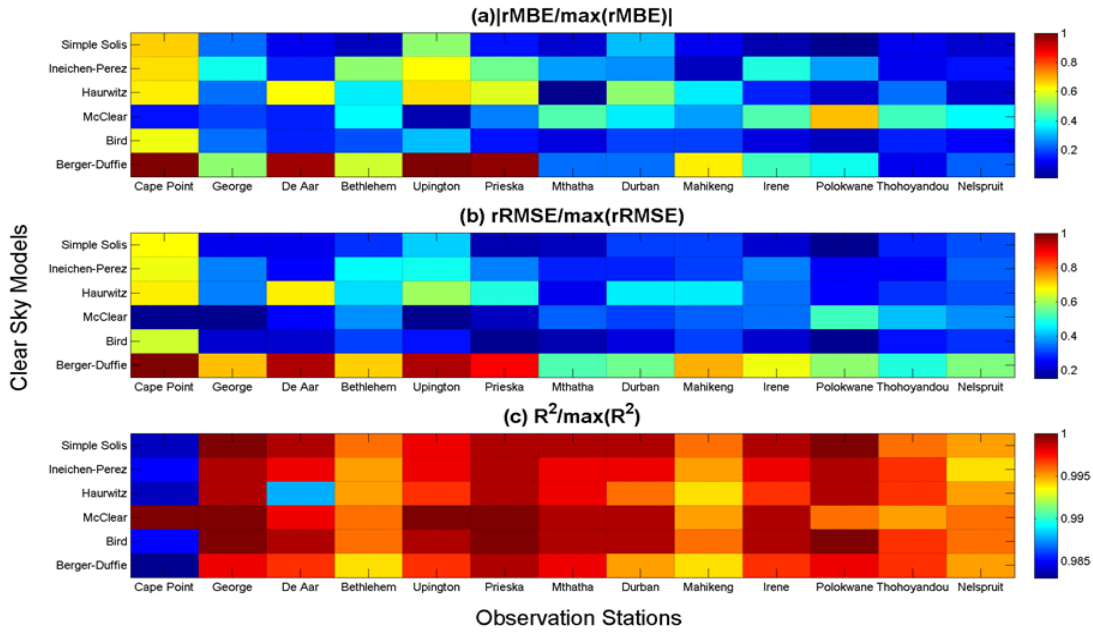


Figure 4.9. Stations grouped within climate zones with metrics normalised relative to the optimum value across all models and all stations, for each metric.

4.4.3. Validation and Comparison of the Results with other Literature

From Table S4.1 and Figure 4.8c, the Bird model had excellent performance, with an $rRMSE$ of less than 5 % at all the stations except for Bethlehem, Mahikeng, and Cape Point, and good performance in Bethlehem and Mahikeng (with $5 \% \leq rRMSE < 10 \%$). This clear sky model also has acceptable performance at the Cape Town station, with respect to $rRMSE$ (with $10 \% \leq rRMSE < 15 \%$). Overall, with just the $rRMSE$ in analysis, the Bird model did not show poor performance at any of the stations. The performance of the Bird model is also verified in Table 4.4, where the model is observed to have the minimum $rRMSE$ at six (46 %) of the stations when compared to the $rRMSE$ of the other five models. The McClear model performed optimally, w.r.t. a minimum $rRMSE$, at three stations, and the Ineichen–Perez and Simple Solis both have optimal performance at two stations. In considering the correlation coefficient, the Bird model has optimal performance, w.r.t. a maximum R^2 , at nine of the stations and the McClear model showed optimal performance at the remaining four stations.

Generally, the presented results for the Bird model at the six stations are comparable and even better than other literature studies which lie in the same latitudes as the South Africa stations. Some of the literature results are given in the shaded block in Table S4.1. The better results might be because of using monthly satellite albedo instead of a default value, calculating P_w from measured temperature and humidity instead of climatologies, and using actual station pressure. O_3 was the only input that was sourced as monthly climatology values from SoDa [28]. According to Bird et al., 1981 [5], O_3 has a minor effect on irradiance, and Polo et al. [1] reported that the variability of daily O_3 ozone is low. The good performance of the Bird model in the study might be because of using high spatio-temporal atmospheric input parameters. As shown in Figure 4.8a, in considering the $rMBE$ recorded with the Bird model, this model did underestimate GHI at all stations except for Thohoyandou, where the model overestimated the irradiance.

From Table S4.2 and Figure 4.8c, the McClear clear sky model had excellent performance in five stations: Upington, De Aar, Prieska, George and Cape Point with $rRMSE < 5 \%$, and good performance in the rest of the stations (with $5 \% \leq rRMSE < 10 \%$). The $rMBE$ polarity in Figure 4.8a shows that the McClear model overestimated GHI in 10 of the 13 stations, with underestimations at Thohoyandou, Polokwane, and Cape Point, which are locations in the “extreme” latitudes for the study region. Overall, the McClear model was outperformed by Simple Solis and Bird in South African locations. This may mean that the “black box” model needs further improvement to be able to estimate irradiance with better accuracy in South African locations. Still, the McClear model performs well, and it has an advantage over all the other models because the data are readily available and there is no need to configure and provide input parameters.

From Table S4.3 and Figure 4.8c, the Simple Solis clear sky model had excellent performance in Thohoyandou, Polokwane, Bethlehem, Irene, Prieska, De Aar, Mthatha and George with $rRMSE < 5\%$. The model had good performance in Nelspruit, Mahikeng, Upington, Durban and Cape Point with $5\% \leq rRMSE < 10\%$. As shown by the $rMBE$ polarity in Figure 5.8a, the Simple Solis model overestimated GHI in 3 of the 13 stations, namely Thohoyandou, Polokwane and Irene. The results for Simple Solis performance are comparable and even better than other literature studies which lie in the same latitudes as South Africa. The better results might be because of using high spatiotemporal resolution model inputs, using actual station pressure, daily AOD_{700} from SoDa [28] and P_w calculated from measured temperature and humidity.

From Table S4.4 and Figure 4.8c, the Ineichen–Perez clear sky model had excellent performance in Thohoyandou, Polokwane, Durban, De Aar and Mthatha with $rRMSE < 5\%$. The model had good performance for the rest of the stations with $5\% \leq rRMSE < 10\%$. The model underestimated GHI in 6 out of 13 stations, as shown in the $rMBE$ polarity in Figure 4.8a. The model performed less fairly, with an overall $rRMSE$ of 5.74 %, when compared against other complex models— $rRMSE$ of 4.11 % for Bird, 4.61% for Simple Solis and 5.06 % for McClear. This might be because Ineichen–Perez is the only model that depends on TL as an input. The TL data were sourced as monthly climatologies from SoDa [28]. According to Ineichen (2006) [41], TL has the highest influence on model accuracy. The model accuracy might improve with locally measured TL, though TL observations were not available in the area of study.

From Table S4.5 and Figure 4.8c, the Haurwitz clear sky model performed excellently in Polokwane and Mthatha with $rRMSE < 5\%$. The model had good performance at the other stations but average performance, i.e., $rRMSE > 10\%$, in Upington and De Aar, which are stations that receive very high irradiance. The model underestimated irradiance at all the

stations except in Thohoyandou, Nelspruit, and Durban, as shown by $rMBE$ polarity in Figure 4.8a. The performance of the model may be a result of the calibration of the Haurwitz model. The Blue Hill observatory where the coefficients a and b were calibrated has different atmospheric constituents to South African locations. There is a need to calibrate coefficients a and b using data measured from South African locations.

From Table S4.6 and Figure 4.8c, the Berger–Duffie clear sky model did not have excellent performance in any of the locations. No station had $rRMSE < 5\%$. The model had good performances in Thohoyandou, Nelspruit, Mthatha and Durban, with $5\% \leq rRMSE < 10\%$, but average performance in Polokwane, Bethlehem, Mahikeng, Irene, and George—with $10\% \leq rRMSE < 15\%$. The model had poor performance in areas that receive very high irradiance, namely Upington, De Aar and Prieska. The model underestimated irradiance in almost all the sites except in Durban, as shown by $rMBE$ polarity in Figure 4.8a. From the results, since the transmittance of the model was integrated to a constant of 0.7, there is a need to recalibrate the constant for each of the South African sites.

The plots in Figure 4.9a–c show the stations grouped within climate zones, and the data are presented with metrics normalised relative to the optimum value across all models and all stations, for each metric. Table 4.4 shows the resulting optimum model at each station and within the climate belts. Both Cape Point and George are in the temperate coastal belt, and Figure 4.9a–c and Table 4.4 show that the McClear model is the best performing for this temperate coastal belt. On the other hand, for the cold interior stations, namely De Aar and Bethlehem, the model with optimal results is either Bird or Simple Solis. For the arid interior region (with Upington and Prieska), either the McClear or Bird model is optimal. The Bird model is the most likely model for the subtropical coastal belt, with Mthatha and Durban. For the temperate interior—with Mahikeng, Irene and Polokwane—no clear optimal model can be picked out from Table 4.4, but visual inspection of Figure 4.9a–c suggests that the most feasible

model is either Simple Solis or Bird. There is no clear optimal model for the stations Thohoyandou and Nelspruit, which are both in the hot interior region.

From Tables S4.1–S4.13, the model performances based on the overall *rRMSE* were as follows: Bird 4.11 %, Simple Solis 4.61 %, McClear 5.06 %, Ineichen–Perez 5.7 4%, Haurwitz 7.04 % and Berger–Duffie 12.17 %. The overall *rMBE* was positive for McClear +1.39 %, meaning that the model overestimates GHI. The correlation between measured clear sky GHI and corresponding modelled GHI results, and a brief description are given in the supplementary materials (Figure S4.1–S4.13), and the 1:1 (black dotted) line explains the correlation—the closer the values to the line, the better the model, and vice versa. The correlation results agreed with the statistics from Tables S4.1–S4.6.

4.5. Supplementary Materials: Tables S4.1–S4.6 (Statistics of model’s performance per site) and Figures S4.1–S4.13, which show the correlation between measured clear sky GHI and corresponding modelled GHI per site.

4.5.1. Statistics of model’s performance per site

Table S4.1. Bird clear sky model predicted irradiance versus measured, with validation metrics

Station	Latitude	Altitude (m)	Measured (Wm^{-2})	Predicted (Wm^{-2})	rMBE (%)	rRMSE (%)	R ²
Thohoyandou	-23.08	619	496.94	505.81	1.79	4.69	0.996
Polokwane	-23.86	1233	498.90	494.99	-0.78	2.78	0.999
Bethlehem	-25.25	1688	471.68	460.74	-2.32	5.38	0.995
Nelspruit	-25.39	870	485.19	478.57	-1.36	4.96	0.995
Mahikeng	-25.81	1289	512.03	501.06	-2.14	5.30	0.995
Irene	-25.91	1524	493.23	488.26	-1.01	3.70	0.998
Uppington	-28.48	848	589.77	569.47	-3.44	4.59	0.998
Durban	-29.61	91	443.56	433.99	-2.16	3.82	0.998
Prieska	-29.68	989	606.53	597.22	-1.54	2.71	0.999
De Aar	-30.67	1284	570.57	560.25	-1.81	3.73	0.998
Mthatha	-31.55	744	453.34	448.13	-1.15	3.19	0.998
George	-34.01	192	491.98	479.45	-2.55	3.63	0.999
Cape Point	-35.35	86	585.88	546.49	-6.72	10.95	0.984

Adelaide	-34.95	2	284.8		5.22	5.35	0.992
Alice Springs	-23.8	546	538.3		3.66	6.26	0.987
Kalgoorlie	-30.78	365	465.2		1.85	3.45	0.996
Learmonth	-22.24	5	364.2		3.98	7.57	0.993
Mildura	-34.54	50	438.4		1.2	4.06	0.995
Mt Gambier	-37.75	63	284.9		4.44	6.99	0.991
Rockhampton	-23.38	10	387.4		5.23	5.3	0.991
Wagga	-35.16	212	398.3		2.94	4.47	0.994
Reunion Island	-20.89	4	821		1.3	3.4	0.99

Table S4.2. McClear clear sky model predicted irradiance versus measured, with validation metrics

Station	Latitude	Altitude (m)	Measured (Wm^{-2})	Predicted (Wm^{-2})	rMBE (%)	rRMSE (%)	R ²
Thohoyandou	-23.08	619	496.94	472.60	-4.90	7.07	0.994
Polokwane	-23.86	1233	498.90	460.80	-7.64	8.97	0.995
Bethlehem	-25.25	1688	471.68	491.29	4.16	6.34	0.995
Nelspruit	-25.39	870	485.19	505.78	4.24	6.52	0.995
Mahikeng	-25.81	1289	512.03	527.79	3.08	5.77	0.994
Irene	-25.91	1524	493.23	517.69	4.96	5.93	0.998
Upington	-28.48	848	589.77	593.44	0.62	2.63	0.999
Durban	-29.61	91	443.56	461.01	3.94	5.16	0.998
Prieska	-29.68	989	606.53	623.12	2.74	3.35	0.999
De Aar	-30.67	1284	570.57	580.76	1.79	4.24	0.997
Mthatha	-31.55	744	453.34	475.84	4.96	5.84	0.998
George	-34.01	192	491.98	502.21	2.08	2.74	0.999
Cape Point	-35.35	86	585.88	577.03	-1.51	2.76	0.999
Alice Springs	-23.8	547	715		2	3	0.99
Reunion Island	-20.89	4	821		0.8	3	0.99

Table S4.3. Simple Solis model predicted irradiance versus measured, with validation metrics

Station	Latitude	Altitude (m)	Measured (Wm^{-2})	Predicted (Wm^{-2})	rMBE (%)	rRMSE (%)	R ²
Thohoyandou	-23.08	619	496.94	503.31	1.28	4.78	0.995
Polokwane	-23.86	1233	498.90	499.53	0.13	2.79	0.999
Bethlehem	-25.25	1688	471.68	468.44	-0.69	4.93	0.995
Nelspruit	-25.39	870	485.19	481.19	-0.82	5.56	0.994
Mahikeng	-25.81	1289	512.03	505.52	-1.27	5.20	0.995
Irene	-25.91	1524	493.23	495.81	0.52	3.65	0.998
Upington	-28.48	848	589.77	555.87	-5.75	7.29	0.997
Durban	-29.61	91	443.56	428.03	-3.50	5.29	0.998
Prieska	-29.68	989	606.53	596.39	-1.67	3.27	0.998
De Aar	-30.67	1284	570.57	563.45	-1.25	4.04	0.998
Mthatha	-31.55	744	453.34	449.29	-0.89	3.40	0.998
George	-34.01	192	491.98	479.08	-2.62	4.12	0.999
Cape Point	-35.35	86	585.88	542.22	-7.45	11.77	0.983
Adelaide	-34.95	2	284.8		-1.65	4.97	0.993
Alice Springs	-23.8	546	538.3		-1.7	6.1	0.987
Kaloorlie	-30.78	365	465.2		-3.16	4.47	0.994
Learmonth	-22.24	5	364.2		-2.89	8.49	0.991
Mildura	-34.54	50	438.4		-5.01	6.35	0.987
Rockhampton	-23.38	10	387.4		-2.18	3.98	0.995
Wagga	-35.16	212	398.3		-1.94	4.9	0.993
Reunion Island	-20.89	4	821		1.2	4.8	0.97

Table S4.4. Ineichen –Perez model predicted irradiance versus measured, with validation metrics

Station	Latitude	Altitude (m)	Measured (Wm^{-2})	Predicted (Wm^{-2})	rMBE (%)	rRMSE (%)	R ²
Thohoyandou	-23.08	619	496.94	490.77	-1.24	4.28	0.996
Polokwane	-23.86	1233	498.90	514.39	3.11	4.46	0.998
Bethlehem	-25.25	1688	471.68	498.41	5.67	8.04	0.994
Nelspruit	-25.39	870	485.19	492.81	1.57	5.70	0.993
Mahikeng	-25.81	1289	512.03	515.81	0.74	5.17	0.994
Irene	-25.91	1524	493.23	515.75	4.57	6.10	0.997
Upington	-28.48	848	589.77	548.80	-6.95	8.26	0.997
Durban	-29.61	91	443.56	457.00	3.03	4.88	0.997
Prieska	-29.68	989	606.53	573.93	-5.38	6.15	0.998
De Aar	-30.67	1284	570.57	560.56	-1.75	4.34	0.997
Mthatha	-31.55	744	453.34	467.96	3.23	4.86	0.997
George	-34.01	192	491.98	470.90	-4.29	6.08	0.998

Cape Point	-35.35	86	585.88	543.41	-7.25	11.39	0.984
Adelaide	-34.95	2	284.8		-5.65	5.84	0.99
Alice Springs	-23.8	546	538.3		0.06	4.98	0.992
Kalgoorlie	-30.78	365	465.2		-3.68	4.26	0.994
Learmonth	-22.24	5	364.2		1.04	5.09	0.997
Mildura	-34.54	50	438.4		-3.13	4.74	0.993
Rockhampton	-23.38	10	387.4		-3.3	3.93	0.995
Wagga	-35.16	212	398.3		-4.88	5.8	0.99

Table S4.5. Haurwitz model predicted irradiance versus measured, with validation metrics

Station	Latitude	Altitude (m)	Measured (Wm^{-2})	Predicted (Wm^{-2})	rMBE (%)	rRMSE (%)	R ²
Thohoyandou	-23.08	619	496.94	510.00	2.63	5.10	0.996
Polokwane	-23.86	1233	498.90	494.33	-0.92	4.31	0.998
Bethlehem	-25.25	1688	471.68	452.98	-3.96	7.47	0.994
Nelspruit	-25.39	870	485.19	489.38	0.86	5.55	0.994
Mahikeng	-25.81	1289	512.03	491.88	-3.94	7.71	0.993
Irene	-25.91	1524	493.23	484.78	-1.71	5.89	0.996
Upington	-28.48	848	589.77	546.90	-7.27	10.29	0.996
Durban	-29.61	91	443.56	468.88	5.71	7.82	0.995
Prieska	-29.68	989	606.53	566.88	-6.54	8.42	0.998
De Aar	-30.67	1284	570.57	530.77	-6.98	11.92	0.987
Mthatha	-31.55	744	453.34	452.61	-0.16	4.20	0.997
George	-34.01	192	491.98	479.29	-2.58	6.17	0.998
Cape Point	-35.35	86	585.88	544.49	-7.07	12.04	0.983

Table S4.6. Berger-Duffie model predicted irradiance versus measured, with validation metrics

Station	Latitude	Altitude (m)	Measured (Wm^{-2})	Predicted (Wm^{-2})	rMBE (%)	rRMSE (%)	R ²
Thohoyandou	-23.08	619	496.94	490.76	-1.24	8.39	0.996
Polokwane	-23.86	1233	498.90	477.13	-4.36	10.00	0.997
Bethlehem	-25.25	1688	471.68	441.14	-6.47	12.30	0.993
Nelspruit	-25.39	870	485.19	472.93	-2.53	9.79	0.994
Mahikeng	-25.81	1289	512.03	475.14	-7.20	12.92	0.993
Irene	-25.91	1524	493.23	468.85	-4.94	11.53	0.996
Uppington	-28.48	848	589.77	523.92	-11.17	16.59	0.996
Durban	-29.61	91	443.56	455.23	2.63	9.72	0.994
Prieska	-29.68	989	606.53	539.94	-10.98	15.31	0.998
De Aar	-30.67	1284	570.57	508.96	-10.80	16.58	0.996
Mthatha	-31.55	744	453.34	441.08	-2.70	9.10	0.997
George	-34.01	192	491.98	464.12	-5.66	12.54	0.997
Cape Point	-35.35	86	585.88	520.53	-11.15	17.32	0.982

4.5.2. Scatter plots of measured and modelled clear sky global horizontal irradiance

From scatter plots in Figure S4.1 show the correlation between measured clear sky GHI and corresponding modelled GHI at De Aar station. Complex models show a very good correlation between measured and observed GHI (the data is close to the 1:1 (black dotted) line). For simple models, there is poor correlation between measured and modelled GHI data.

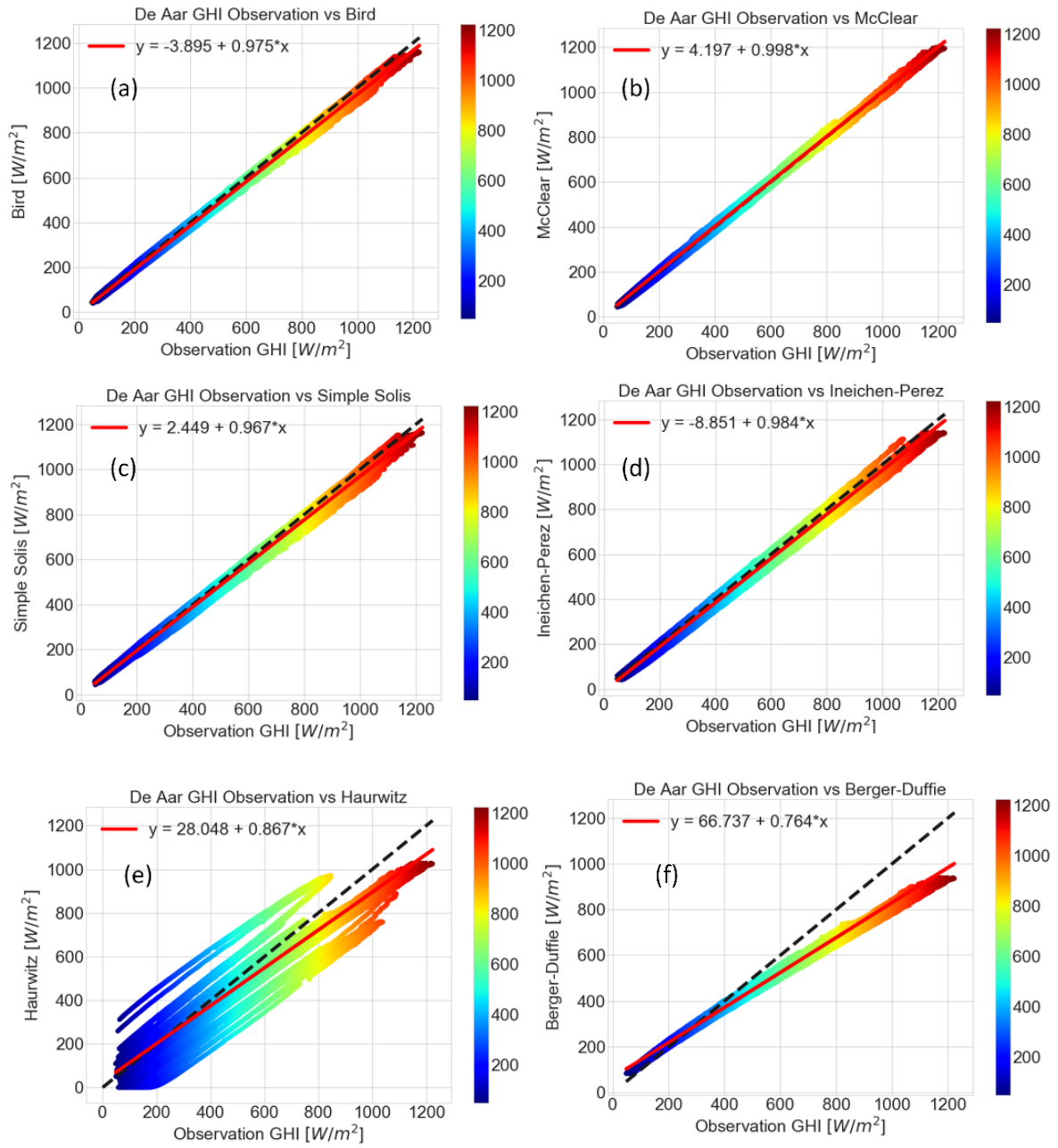


Figure S4.1. Scatter plot of measured and modelled clear sky global irradiance at De Aar for Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz model (e) and Berger-Duffie model (f).

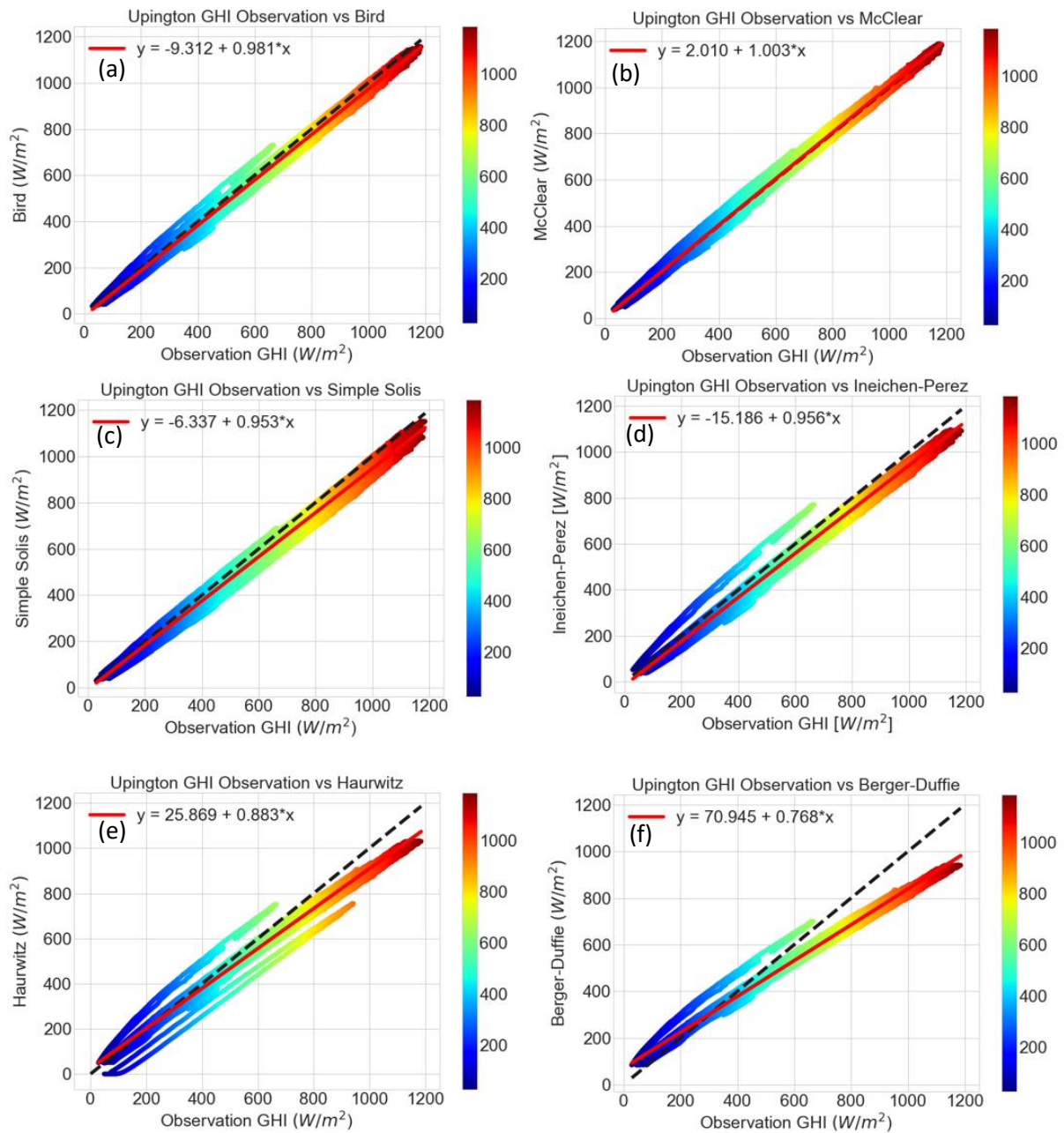


Figure S4.2. Scatter plot of measured and modelled clear sky global irradiance at Uppington station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.2, which show the correlation between measured clear sky GHI and corresponding modelled GHI at Uppington station. Bird, Simple Solis and McClear models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Ineichen-Perez, Haurwitz and Berger-Duffie models there is poor

correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

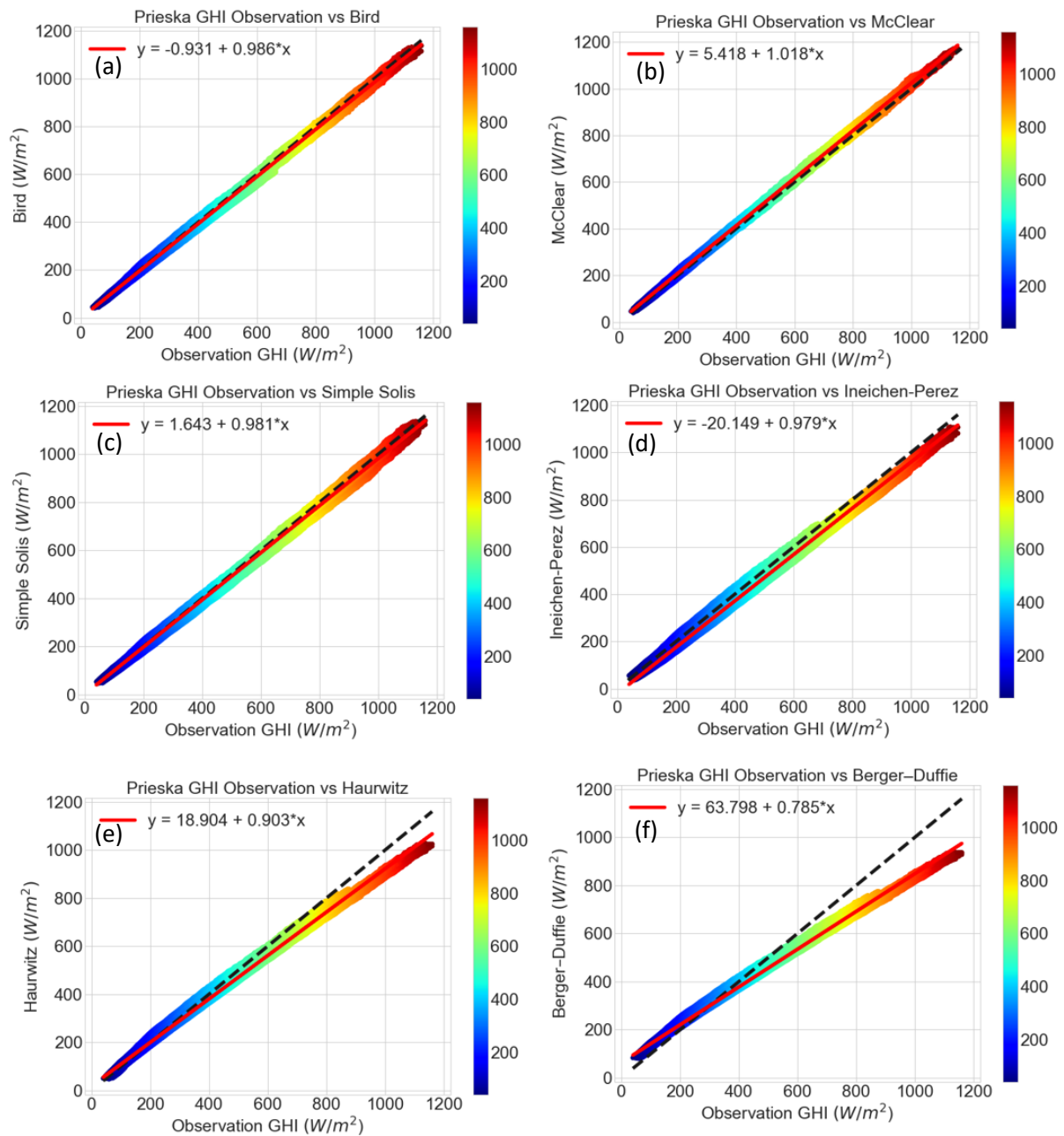


Figure S4.3. Scatter plot of measured and modelled clear sky global irradiance at Prieska station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.3, which show the correlation between measured clear sky GHI and corresponding modelled GHI at Prieska station. Bird, Simple Solis, McClear and Ineichen-Perez models show a very good correlation between measured and observed GHI the data is

close to the 1:1 (black dotted) line, while for Haurwitz and Berger-Duffie models there is poor correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

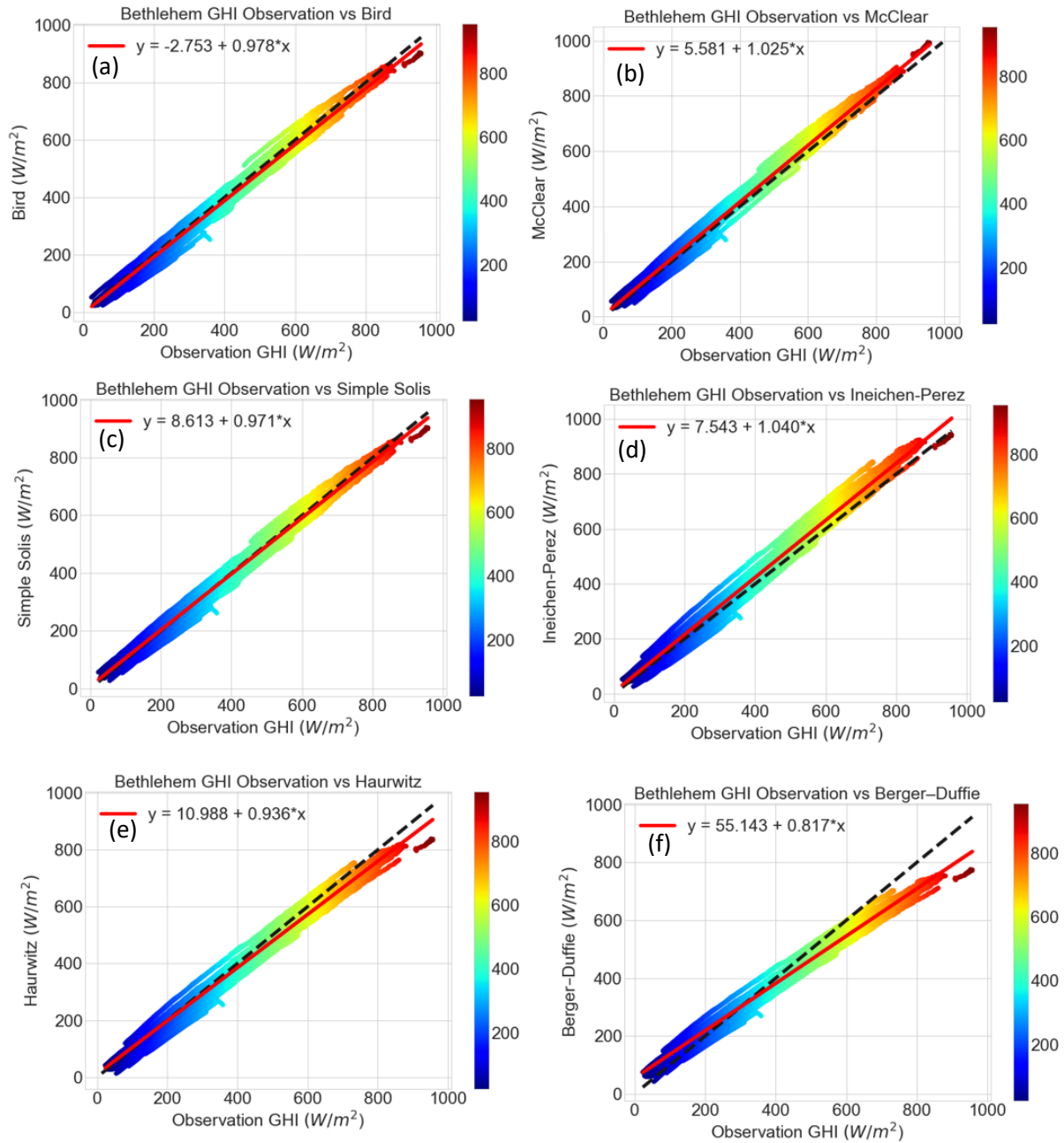


Figure S4.4. Scatter plot of measured and modelled clear sky global irradiance at Bethlehem station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.4 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Bethlehem station. Bird, Simple Solis, McClear and Haurwitz models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Ineichen-Perez and Berger-Duffie models there is poor correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

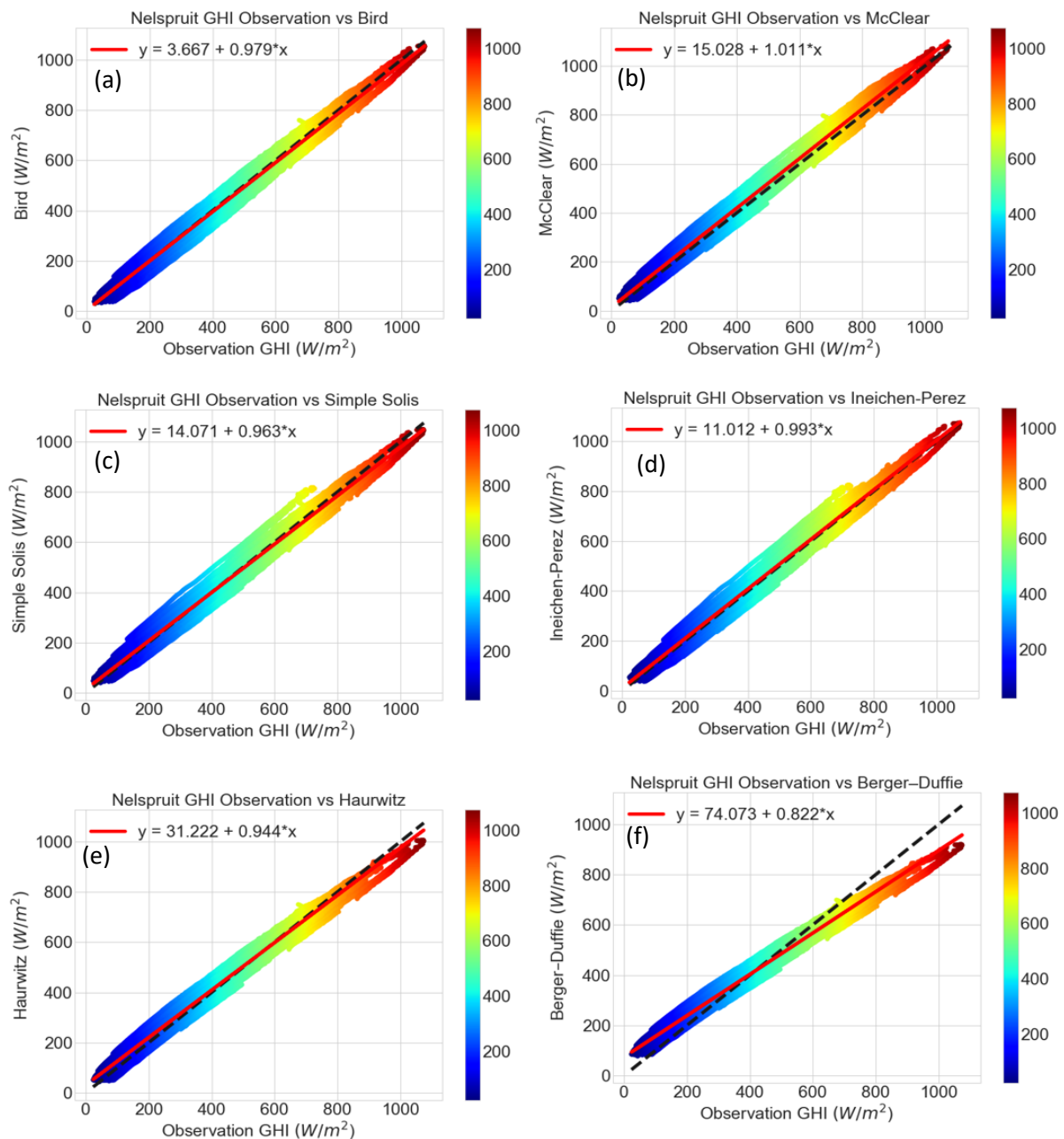


Figure S4.5 Scatter plot of measured and modelled clear sky global irradiance at Nelspruit station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.5 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Nelspruit station. Bird, Simple Solis, McClear, Ineichen-Perez and Haurwitz models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Berger-Duffie model there is poor correlation between measured and modelled GHI data when irradiance values are high, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

From scatter plot in Figure S4.6 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Irene station. Bird, Simple Solis, McClear, Ineichen-Perez and Haurwitz models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Berger-Duffie model there is poor correlation between measured and modelled GHI data when irradiance values are high, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

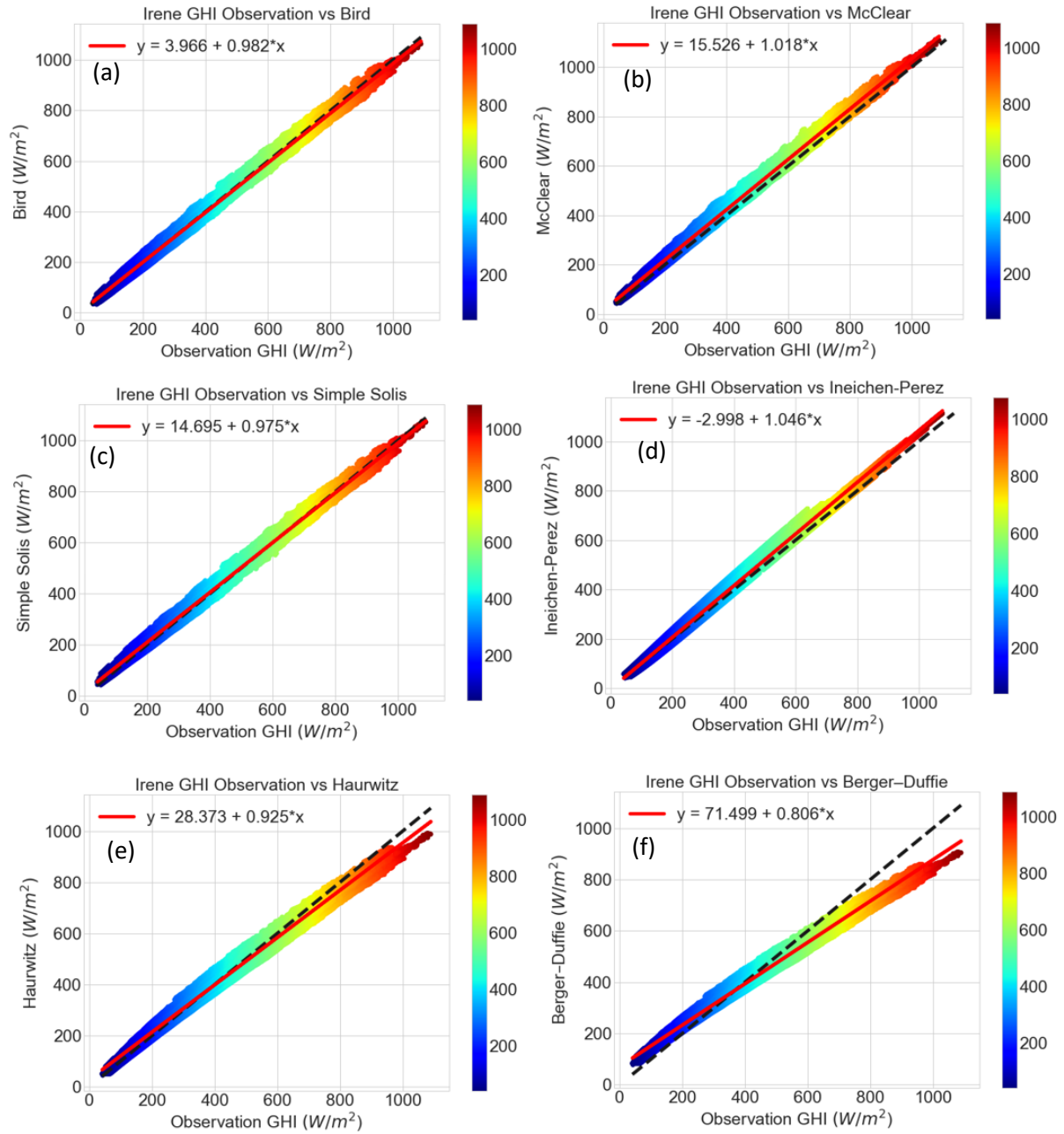


Figure S4.6. Scatter plot of measured and modelled clear sky global irradiance at Irene station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

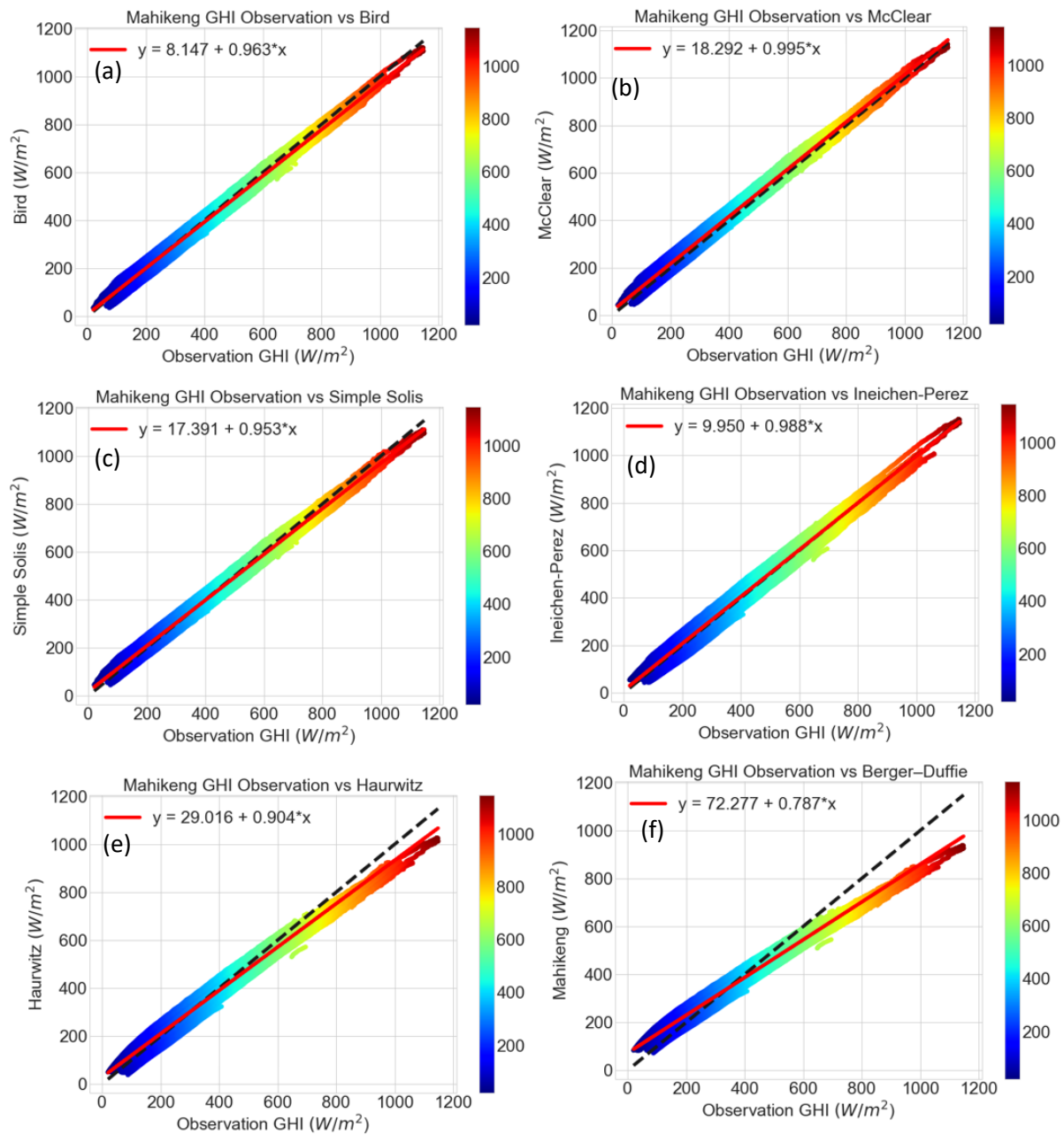


Figure S4.7. Scatter plot of measured and modelled clear sky global irradiance at Mahikeng station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.7 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Mahikeng station. Bird, Simple Solis, McClear and Ineichen-Perez models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Haurwitz and Berger-Duffie models there is poor correlation between measured and modelled GHI data when irradiance values are high,

the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

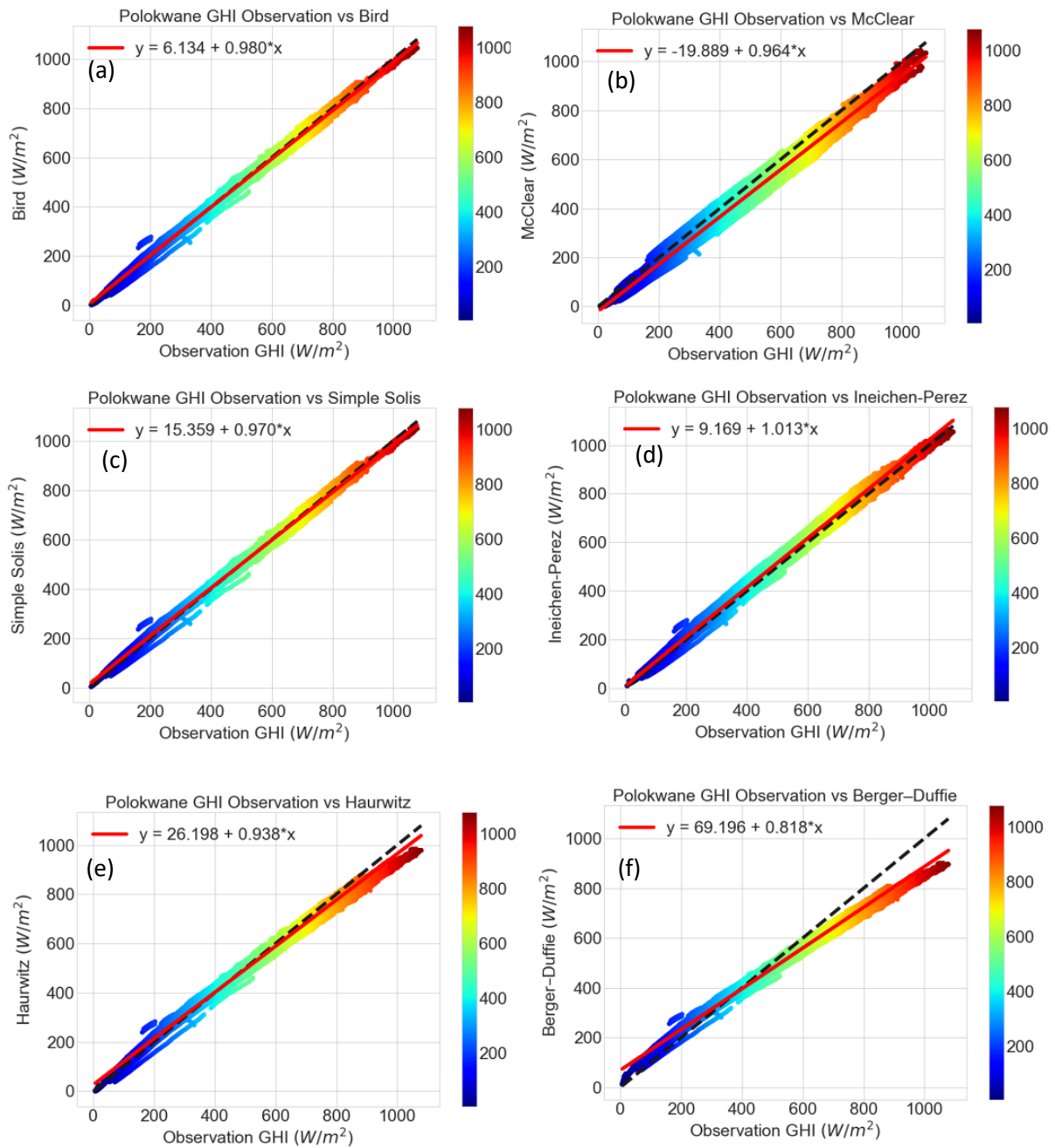


Figure S4.8. Scatter plot of measured and modelled clear sky global irradiance at Polokwane station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.8 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Polokwane station. Bird, Simple Solis, Ineichen-Perez and Haurwitz models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for McClear and Berger-Duffie models there is poor correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6

From scatter plot in Figure S4.9 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Thohoyandou station. Bird, Simple Solis, Ineichen-Perez and Haurwitz models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for McClear and Berger-Duffie models there is poor correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

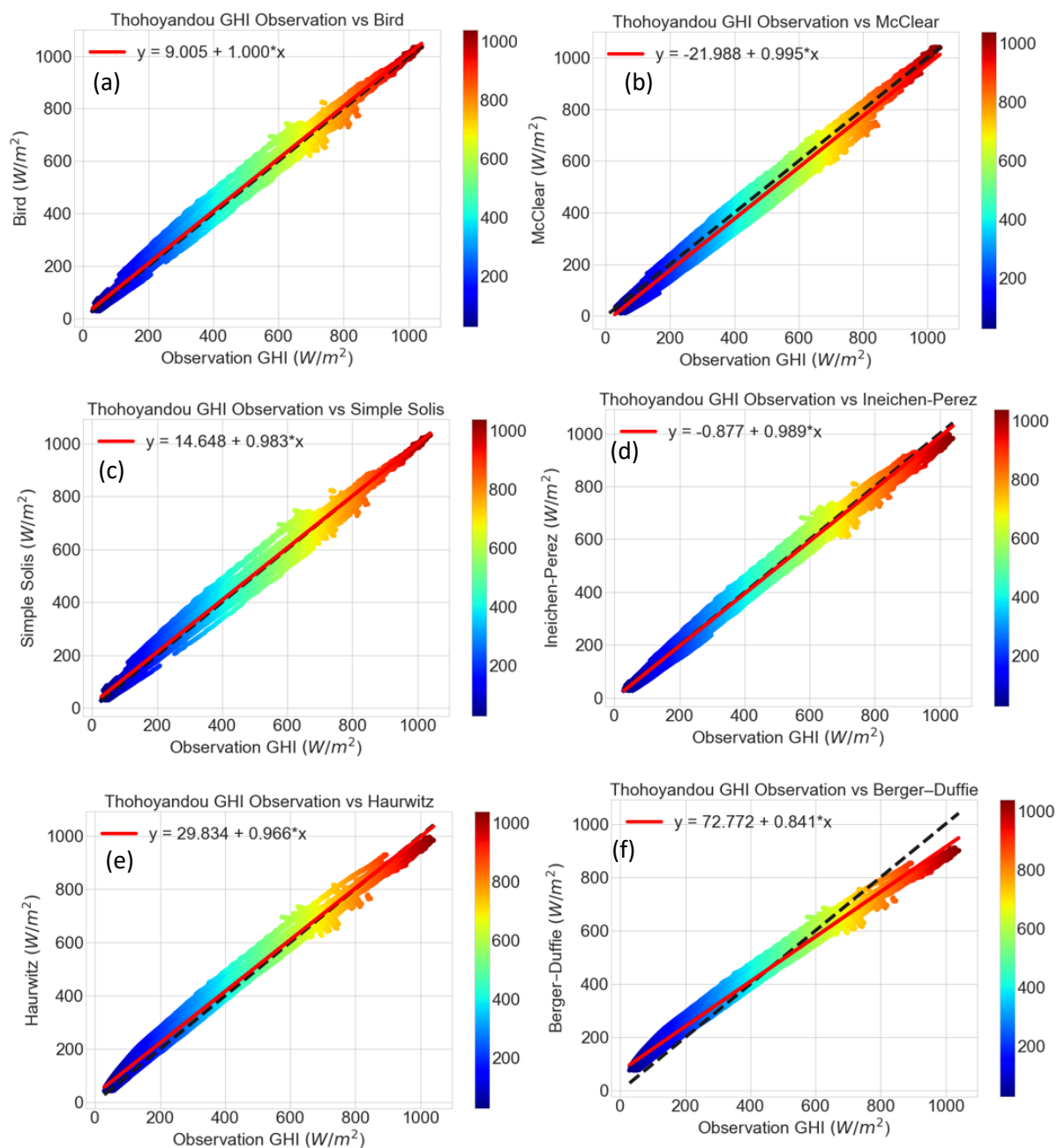


Figure S4.9. Scatter plot of measured and modelled clear sky global irradiance at Thohoyandou station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

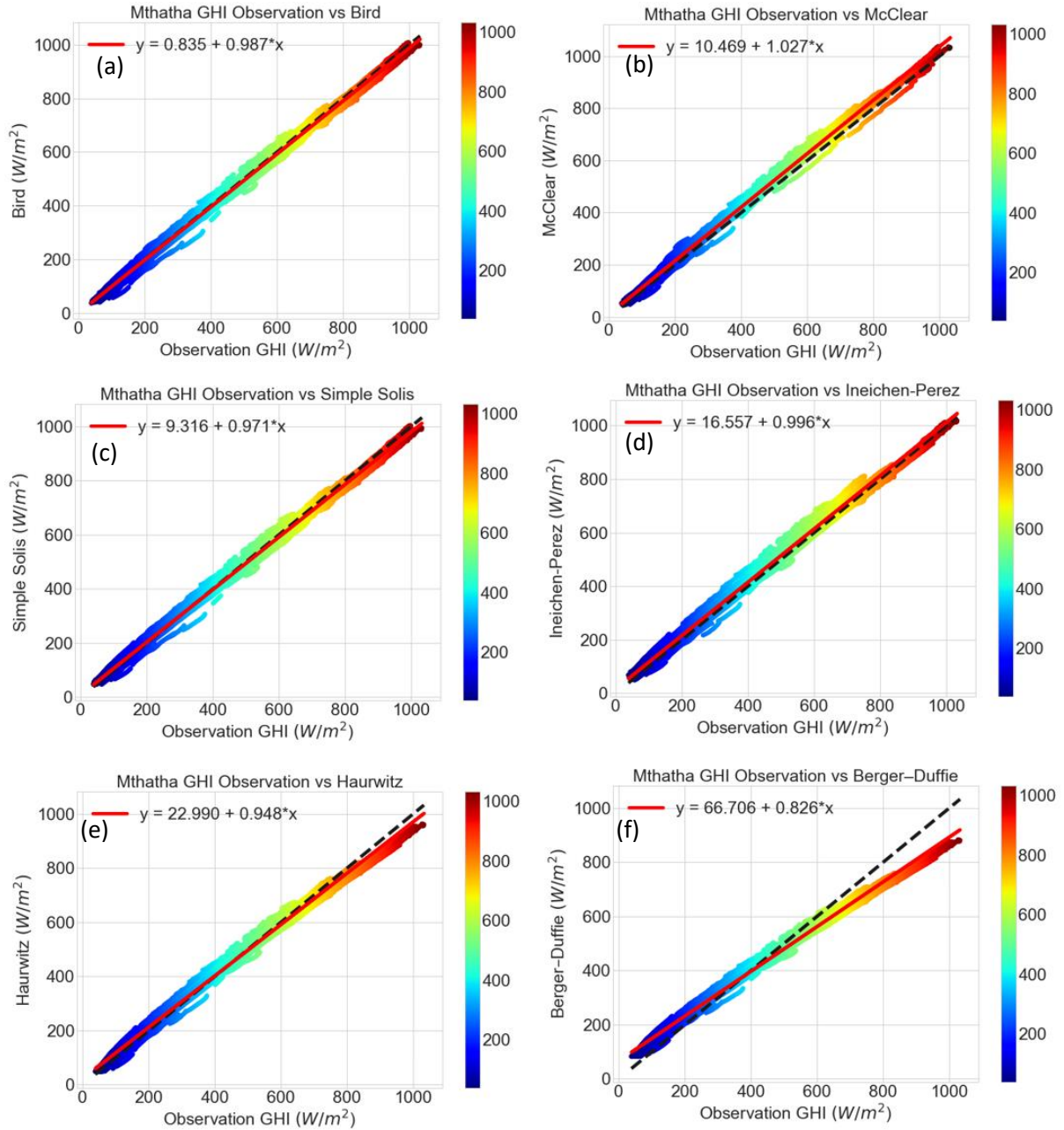


Figure S4.10. Scatter plot of measured and modelled clear sky global irradiance at Mthatha station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (c), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.10 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Mthatha station. Bird, Simple Solis, McClear, Ineichen-Perez and Haurwitz models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Berger-Duffie model there is poor

correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

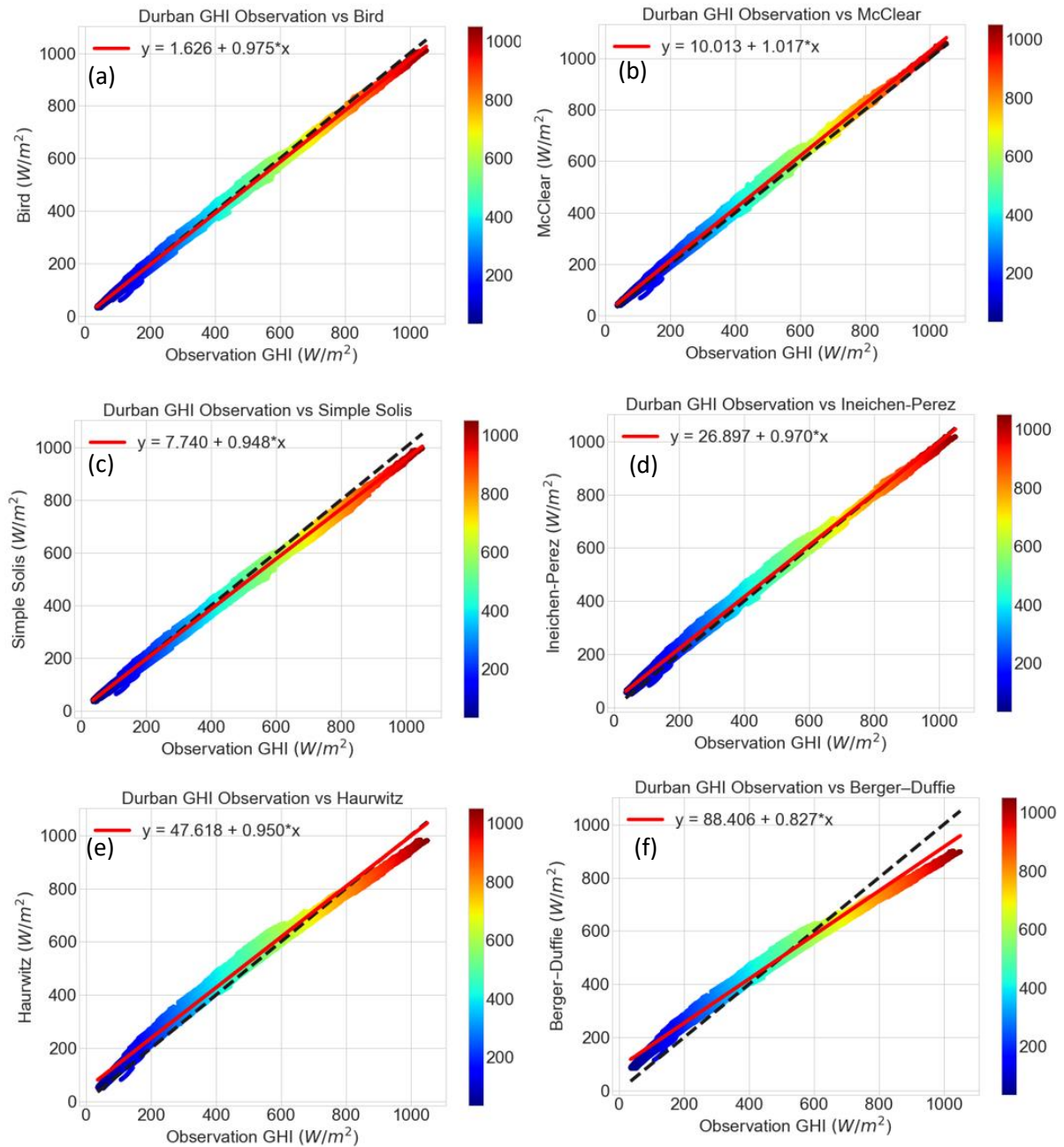


Figure S4.11. Scatter plot of measured and modelled clear sky global irradiance at Durban station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.11 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Durban station. Bird, Simple Solis, McClear, Ineichen-Perez and Haurwitz models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Berger-Duffie model there is poor correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

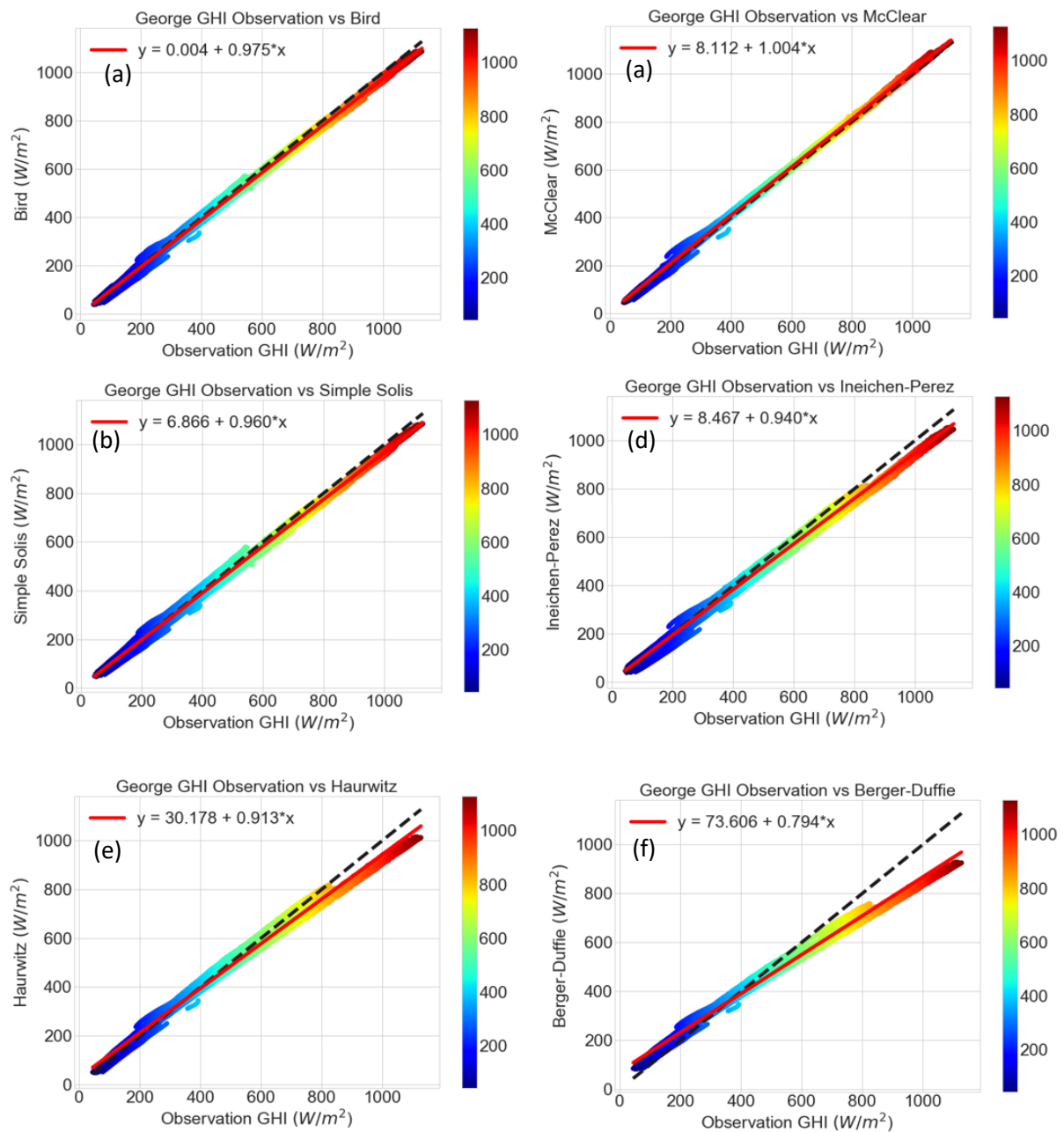


Figure S4.12. Scatter plot of measured and modelled clear sky global irradiance at George station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (d), Haurwitz (e) and Berger-Duffie (f).

From scatter plot in Figure S4.12 which show the correlation between measured clear sky GHI and corresponding modelled GHI at George station. Bird, Simple Solis and McClear models show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Ineichen-Perez, Haurwitz and Berger-Duffie models there is poor correlation between measured and modelled GHI data when irradiance values are high, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

From scatter plot in Figure S4.13 which show the correlation between measured clear sky GHI and corresponding modelled GHI at Cape Point station. and McClear model show a very good correlation between measured and observed GHI the data is close to the 1:1 (black dotted) line, while for Bird, Simple Solis, Ineichen-Perez, Haurwitz and Berger-Duffie models there is poor correlation between measured and modelled GHI data, the data diverged from the 1:1 line, the correlation shown here agrees with the statistics from table S4.1 to S4.6.

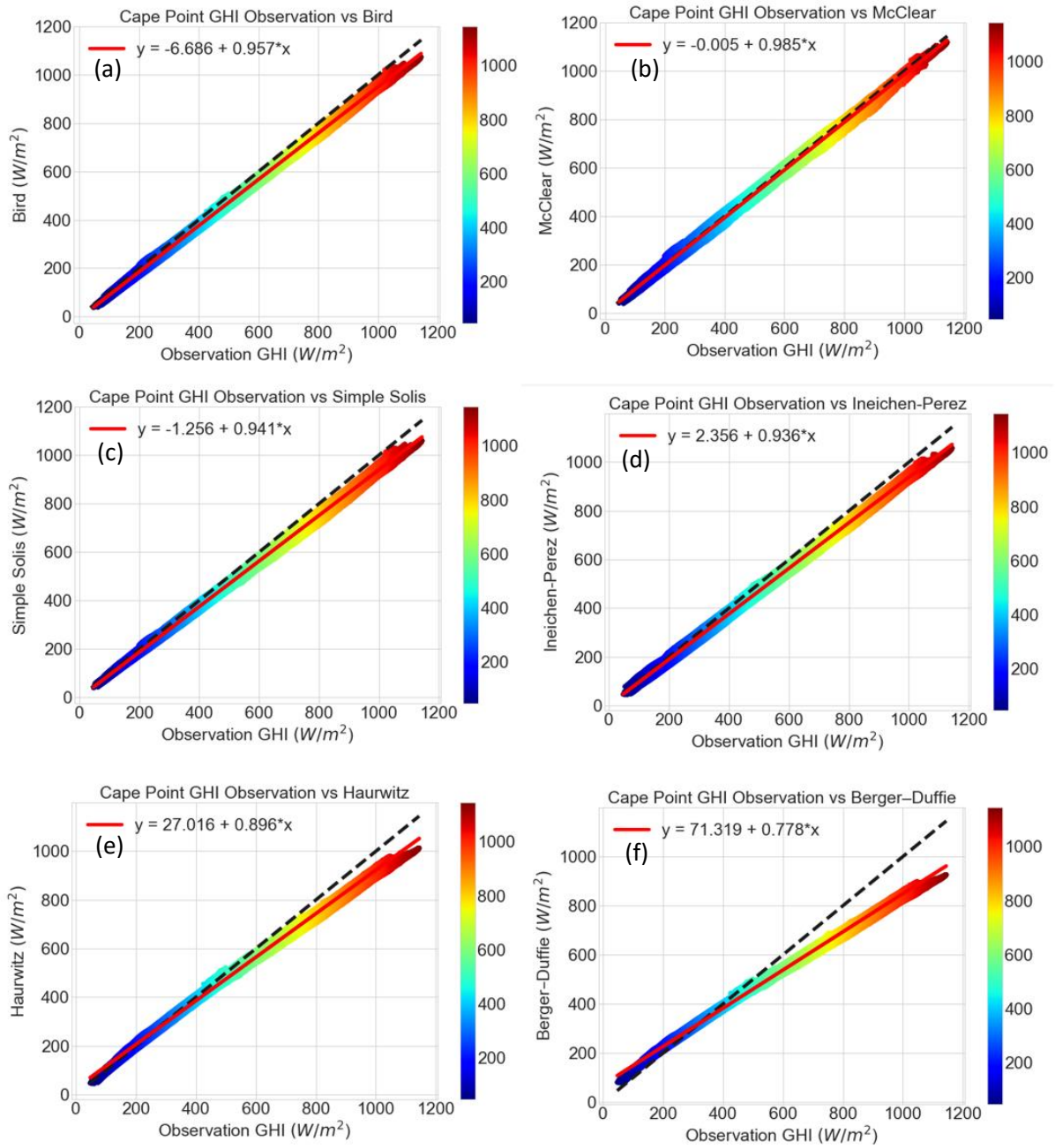


Figure S4.13. Scatter plot of measured and modelled clear sky global irradiance at Cape Point station, Bird model (a), McClear model (b), Simple Solis model (c), Ineichen-Perez model (c), Haurwitz (e) and Berger-Duffie (f).

4.6. Chapter Summary

In chapter 4, the performance of six GHI clear sky models, namely Bird, Simple Solis, McClear, Ineichen–Perez, Haurwitz and Berger–Duffie, have been evaluated against reference Baseline Solar Radiation Network quality guided 1-minute data for 13 stations in the South African Weather Services radiometric network. The stations are in six climatological regions. All the models assessed were capable of reproducing GHI irradiance, but with different accuracy in different stations. All four complex models, namely Bird, Simple Solis, McClear and Ineichen–Perez, performed well with rRMSE less than 10 % in all the stations. The simple models, Haurwitz and Berger–Duffie, did not perform well overall in South African locations. Bird was the overall best model, followed by Simple Solis, McClear, Ineichen–Perez, Haurwitz and Berger–Duffie. The McClear model overestimated GHI and the other five models underestimated GHI. Model inputs seem to influence the model performance. There is also an indication that each model behaves differently over particular climate regions. Overall, the McClear model overestimated GHI and the other five models underestimated GHI.

References

1. Polo, J.; Martín-Pomares, L.; Gueymard, C.A.; Balenzategui, J.L.; Fabero, F.; Silva, J.P. Fundamentals: Quantities, Definitions, and Units. In *Solar Resource Mapping—Fundamentals and Applications*; Green Energy and Technology; Polo, J., Martín-Pomares, L., Sanfilippo, A., Eds.; Springer: Zurich, Switzerland, 2019; pp. 1–14, doi:10.1007/978-3-319-97484-2_4.
2. Masoom, A.; Kashap, Y.; Bansal, A. *Solar Radiation Assessment and Forecasting Using Satellite Data*; Tyagi, H., Agarwal, A.K., Chakraborty, P.R., Powar, S., Eds.; Springer: Singapore, 2019; pp. 45–71, doi:10.1007/978-981-13-3302-6_3,
3. Badescu, V.; Gueymard, C.A.; Cheval, S.; Oprea, C.; Baci, M.; Dumitrescu, A.; Iacobescu, F.; Milos, I.; Rada, C. Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. *Renew. Energy* 2013, 55, 85–103, doi:10.1016/j.renene.2012.11.037.

4. Antonanzas-Torres, F.; Urraca, R.; Polo, J.; Perpiñán-Lamigueiro, O.; Escobar, R. Clear sky solar irradiance models: A review of seventy models. *Renew. Sus. Energ. Rev.* 2019, 107, 374–387, doi:10.1016/j.rser.2019.02.032
5. Bird, R.E.; Hulstrom, R.L. Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces; No. SERI/TR-642-761; Solar Energy Research Inst.: Golden, CO, USA, 1981.
6. Daryl, R.M. Solar radiation, Practical Modeling for Renewable Energy Applications. *Energy and the Environment*; CRC Press: Boca Raton, FL, USA, 2013.
7. Mueller, R.; Dagestad, K.; Ineichen, P.; Schroedter-Homscheidt, M.; Cros, S.; Dumortier, D.; Kuhlemann, R.; Olseth, J.; Piernavieja, G.; Resie, C.; et al. Rethinking satellite based solar irradiance modelling. The SOLIS clear-sky module. *Remote Sens. Environ.* 2004, 91, 160–174. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0034425704000690> (accessed on 27 September 2020).
8. Ineichen, P. A broadband simplified version of the Solis clear sky model. *Sol. Energy* 2008, 82, 758–762 doi:10.1016/j.solener.2008.02.009.
9. Lefèvre, M.; Oumbe, A.; Blanc, P.; Espinar, B.; Gschwind, B.; Qu, Z.; Wald, L.; Schroedter-Homscheidt, M.; Hoyer-Klick, C.; Arola, A.; et al. Mcclear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmos. Meas. Tech.* 2013, 6, 2403–2418. Available online: www.atmos-meas-tech.net/6/2403/2013/ doi:10.5194/amt-6-2403-2013 (accessed on 14 July 2020).
10. Ineichen, P.; Perez, R. A new airmass independent formulation for the Linke turbidity coefficient. *Sol. Energy* 2002, 73, 151–157, doi:10.1016/S0038-092X(02)00045-2.
11. Perez, R.; Ineichen, P.; Moore, K.; Kmiecik, M.; George, R.; Vignola, F. A new operational model for satellite-derived irradiances: Description and validation. *Sol. Energy* 2002, 73, 307–317, doi:10.1016/S0038-092X(02)00122-6.
12. Haurwitz, B. Insolation in relation to cloudiness and cloud density. *J. Meteorol.* 1945, 2, 154–166, doi:10.1175/1520-0469(1945)002<0154:IIRTCA>2.0.CO;2.
13. Haurwitz, B. Isolation in relation to cloud type. *J. Meteorol.* 1948, 5, 110–113, doi:10.1175/1520-0469(1948)005<0110:IIRTCT>2.0.CO;2.
14. Berger, X. Etude du Climat en Region Nicoise en vue d'Applications a l'Habitat Solaire; CNRS: Paris, France, 1979.
15. Zhandire, E. Predicting clear-sky global horizontal irradiance at eight locations in South Africa using four models. *J. Energy S. Afr.* 2017, 28, 77–86, doi:10.17159/2413-3051/2017/v28i4a2397.
16. Patel, S.S.; Rix, A.J. Water surface albedo modelling for floating photovoltaic plants. In *Proceedings of the South African Solar Energy Conference (SASEC)*, Mpekwini

Beach Resort, Eastern Cape Province, South Africa, 25–27 November 2019. Available online: <https://www.sasec.org.za/papers2019/8.pdf> (accessed on 27 February 2020).

17. Javu, L.; Winkler, H.; Roro, K. Validating clear-sky irradiance models in five South African locations. In Proceedings of the South African Solar Energy Conference (SASEC), Mpekweni Beach Resort, Eastern Cape Province, South Africa, 25–27 November 2019. Available online: <https://www.sasec.org.za/papers2019/6.pdf> (accessed on 15 February 2020).
18. Malaudzi, S.T.; Sankaran, V.; Lysko, M.D. Solar radiation analysis and regression coefficients for the Vhembe Region, Limpopo Province, South Africa. *J. Energy South. Afr.* 2017, 24, 2–7, doi:10.17159/2413-3051/2013/v24i3a3136.
19. Solar Radiation Data (SoDa) Service. Available online: <http://www.soda-pro.com/web-services> (accessed on 2 March 2020).
20. Copernicus Portal. Available online: <https://atmosphere.copernicus.eu/data> (accessed on 16 March 2020).
21. Gschwind, B.; Wald, L.; Blanc, P.; Lefèvre, M.; Schroedter-Homscheidt, M.; Arola, A. Improving the McClear model estimating the downwelling solar radiation at ground level in cloud-free conditions—McClear-v3. *Meteorol. Z.* 2019, 28, 147–163, doi:10.1127/metz/2019/0946.
22. Reno, M.J.; Hansen, C.W.; Stein, J.S. Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis; SAND2012-2389; Sandia National Laboratories: Albuquerque, NM, USA; Livermore, CA, USA, 2012. Available online: https://scholar.google.com/scholar?q=related:NRwgbk5tC6sJ:scholar.google.com/&scioq=&hl=en&as_sdt=0,5 (accessed on 23 June 2020).
23. Ineichen, P. Quatre Années de Mesures D’ensoleillement à Genève. Doctoral Dissertation, University of Geneva, Switzerland 1983, doi:10.13097/archive-ouverte/unige:17467.
24. Kasten, F. Parametrisierung der Globalstrahlung durch Bedeckungsgrad und Trübungsfaktor. In Proceedings of the Deutsche Meteorologen-Tagung 1983 vom 16. bis 19. Mai 1983, Bad Kissingen, Germany, 16–19 May 1983; pp. 49–50. Available online: <https://ci.nii.ac.jp/naid/10021899342/> (accessed on 11 June 2020).
25. Karlsson, K.; Anttila, K.; Trentmann, J.; Stengel, M.; Meirink, J. F.; Devasthale, A.; Hanschmann, T.; Kothe, S.; Jääskeläinen, E.; Sedlar, J., et al. CLARA-A2.1: CM SAF cLOUD, Albedo and surface RADIation dataset from AVHRR data—Edition 2.1, Satellite Application Facility on Climate Monitoring, CM-SAF 2020, doi:10.5676/EUM_SAF_CM/CLARA_AVHRR/V002_01.
26. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G.; Horányi, A.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I., et al. The ERA5 Global Reanalysis. *Q. J. Roy. Meteor. Soc.*, 2020, 146, 1999– 2049. <https://doi.org/10.1002/qj.3803>.

27. Frederick, J.E. Ozone Depletion and Related Topic, Ozone as a UV Filter; Elsevier, 2015, pp. 359–363.
28. SODA Portal. Available Online: <http://www.soda-pro.com/web-services/atmosphere/linke-turbidity-factor-ozone-water-vapor-and-angstroembeta> (accessed on 21 April 2020).
29. Diabate, L.; Remund, J.; Wald, L. Linke turbidity factors for several sites in Africa. *Sol. Energy* 2003, 75, 111–119, doi:10.1016/j.solener.2003.07.002.
30. SODA Portal. Available Online: <http://www.soda-pro.com/web-services/atmosphere/cams-aod> (accessed on 21 February 2020).
31. Ångström, A. On the Atmospheric Transmission of Sun Radiation and on Dust in the Air. *Geogr. Ann.* 1929, 11, 156–166, doi:10.2307/519399.
32. Ångström, A. Techniques of Determining the Turbidity of the Atmosphere. *Tellus* 1961, 13, 214–223, doi:10.3402/tellusa.v13i2.9493.
33. Keogh, W.M.; Blakers, A.W. Accurate measurement, using natural sunlight, of silicon solar cells. *Prog. Photovolt. Res. Appl.* 2004, 12, 1–19, doi:10.1002/pip.517.
34. Gueymard, C. Analysis of monthly average atmospheric precipitable water and turbidity in Canada and northern United States. *Sol. Energy* 1994, 53, 57–71, doi:10.1016/S0038-092X(94)90606-8.
35. Gueymard, C. Assessment of the accuracy and computing speed of simplified saturation vapour equations using a new ref-erence dataset. *J. Appl. Meteorol.* 1993, 32, 1294–1300, doi:10.1175/1520-0450(1993)032<1294:AOTAAC>2.0.CO;2.
36. Reno, M.J.; Hansen, C.W. Identification of periods of clear sky irradiance in time series of GHI measurements. *Renew. Energy* 2016, 90, 520–531, doi:10.1016/j.renene.2015.12.031.
37. Engerer, N.A.; Mills, F.P. Validating nine clear sky radiation models in Australia. *Sol. Energy* 2015, 120, 9–24, doi:10.1016/j.solener.2015.06.044.
38. Gairaa, K.; Benkacali, S.; Guermoui, M. Clear-sky models evaluation of two sites over Algeria for PV forecasting purpose. *Eur. Phys. J. Plus* 2019, 134, 534, doi:10.1140/epjp/i2019-12917-2.
39. Dazhia, Y.; Jirutitijaroen, P.; Walsh, W.M. The Estimation of Clear Sky Global Horizontal Irradiance at the Equator. *Energy Procedia* 2012, 25, 141–148, doi:10.1016/j.egypro.2012.07.019.
40. Long, C.; Shi, Y. An automated quality assessment and control algorithm for surface radiation measurements. *Open Atmos. Sci. J.* 2008, 2, 23–37. Available online: <https://openatmosphericsciencejournal.com/VOLUME/2/PAGE/23/FULLTEXT/> (accessed on 15 February 2020).

41. Ineichen, P. Comparison of eight clear sky broadband models against 16 independent data banks. *Sol. Energy* 2006, 80, 468–478, doi:10.1016/j.solener.2005.04.018.

Chapter 5: The Ångström-Prescott Regression Coefficients for Six Climatic Zones in South Africa

5.1. Introduction

This chapter aims to answers a problem statement:

1. Are the Ångström–Prescott (AP) coefficients which are used as inputs to the model that is used daily estimate GHI using daily sunshine duration (SD) and Top of the atmosphere (TOA) GHI as an input different from default values $a=0.25$ and $b=0.75$? What are the values of AP coefficients in in each location/station under study and in each macro-climatological region of South Africa? How accurate are the calibrated AP coefficients in estimating daily GHI in each location/station under study and in each macro-climatological region of South Africa?

The problem statement was addressed by setting the following objectives:

1. Calibrate the Ångström–Prescott (AP) model regression coefficients a and b that could be used to estimate GHI in different climatic zones of South Africa, thus increasing the density of available solar radiation data in the country.

The work presented in this chapter has been submitted on 31 July 2020 to the Energies Journal (ISSN 1996-1073; CODEN: ENERGA), a peer-reviewed open access journal of related scientific research, technology development, engineering, and the studies in policy and management. The Journal publishes scientific research work under the Multidisciplinary Digital Publishing Institute (MDPI) publisher with its headquarters based in Basel, Switzerland. The manuscript was reviewed by experts in the field and accepted on 29 September 2020 by the Journal and it was published on 16 October 2020. The work presented herein according to the submission guidelines suggested by the Journal editors. The Energies

Journal (ISSN 1996-10730) is accredited by the Department of Higher Education and Training (DHET).

Mabasa, B.; Lysko, M.D.; Tazvinga, H.; Mulaudzi, S.T.; Zwane, N.; Moloi, S.J. The Ångström–Prescott Regression Coefficients for Six Climatic Zones in South Africa. *Energies* **2020**, *13*, 5418. <https://doi.org/10.3390/en13205418>

Abstract: The South African Weather Service (SAWS) manages an in situ solar irradiance radiometric network of 13 stations and a very dense sunshine recording network, located in all six macroclimate zones of South Africa. A sparsely distributed radiometric network over a landscape with dynamic climate and weather shifts is inadequate for solar energy studies and applications. Therefore, there is a need to develop mathematical models to estimate solar irradiation for a multitude of diverse climates. In this study, the annual regression coefficients, a and b , of the Ångström–Prescott (AP) model, which can be used to estimate global horizontal irradiance (GHI) from observed sunshine hours, were calibrated and validated with observed station data. The AP regression coefficients were calibrated and validated for each of the six macroclimate zones of South Africa using the observation data that span 2013 to 2019. The predictive effectiveness of the calibrated AP model coefficients was evaluated by comparing estimated and observed daily GHI. The maximum annual relative Mean Bias Error (rMBE) was 0.371 %, relative Mean Absolute Error (rMAE) was 0.745 %, relative Root Mean Square Error (rRMSE) was 0.910 %, and the worst-case correlation coefficient (R^2) was 0.910. The statistical validation metrics results show that there is a strong correlation and linear relation between observed and estimated GHI values. The AP model coefficients calculated in this study can be used with quantitative confidence in estimating daily GHI data at locations in South Africa where daily observation sunshine duration data are available.

Author Contributions: The contribution of the authors is summarised below.

Conceptualization: Brighton Mabasa.; Meena D. Lysko and Sophie T. Mulaudzi.

Methodology: Brighton Mabasa and Meena D. Lysko. **Software:** Brighton Mabasa.

Validation: Meena D. Lysko and Sophie T. Mulaudzi. **Formal analysis:** Brighton Mabasa.

Investigation: Brighton Mabasa. **Data curation:** Brighton Mabasa. **Writing—original draft**

preparation: Brighton Mabasa. **Supervision:** Meena D. Lysko and Sabata J. Moloi.

Writing—review and editing: Brighton Mabasa; Henerica Tazvinga.; Meena D. Lysko.;

Sabata J. Moloi.; Nosipho Zwane and Sophie T. Mulaudzi. All authors have read and agreed to the published version of the manuscript.

Solar radiation data are important for various applications, solar radiation data recording stations are sparsely distributed in South Africa this because solar radiation recording stations are expensive to install and maintain. As a result, data are only available for a limited number of locations, and it is inadequate for solar energy studies and applications. To compensate for the inadequate solar radiation stations, there is a need to develop mathematical models to estimate solar irradiation for a multitude of diverse climates. In this chapter, the annual regression coefficients, a and b , of the Ångström–Prescott (AP) model, which can be used to estimate daily global horizontal irradiance (GHI) from observed sunshine hours (SD), were calibrated and validated with observed station data. The AP regression coefficients were calibrated and validated for each of the six macroclimate zones of South Africa using the observation data that, described in Chapter 3.

The AP model estimates daily GHI using daily extraterrestrial (top of the atmosphere) GHI radiation (GHI_{TOA}), daily astronomical day length (N), daily measured sunshine duration (n), and Ångström model coefficients a and b . The model was first proposed by Ångström [1] in

1924 before Prescott [1] modified it in 1940 by adding GHI_{TOA} to replace GHI on a clear sky day. The original AP coefficients were $a = 0.25$ and $b = 0.75$; these were calculated using data from Stockholm [3]. The regression coefficients a and b are site-dependent; therefore, there is a need to calibrate them using a linear relationship in Equation (1) at regions where they will be used to estimate GHI.

The coefficients a and b of the AP model were calculated by using the linear regression analysis between the irradiance fraction or clearness index, $\frac{GHI}{GHI_{TOA}}$ and daily sunshine fraction, $\frac{n}{N}$ for each day, based on a linear relationship shown by Equation (5.1) proposed by Ångström [1] and then, modified by Prescott [2].

$$\frac{GHI}{GHI_{TOA}} = a + b(n/N), \quad (5.1)$$

where GHI is the daily Global Horizontal Irradiance in W/m^2 ,

GHI_{TOA} is an approximation of the top of the atmosphere GHI, as given by equation (2.6) in Chapter 2.

N is the Astronomical sunshine duration or the daylength as given by equation (2.18) in Chapter 2.

n = daily recorded sunshine duration (in hours),

where a and b represent Ångström–Prescott regression coefficients.

5.2. Study Area

The study area was determined by considering the locations of the South African Weather Services (SAWS) radiometric stations which measure the 1-minute GHI data. The area of study is eight sites located in all six macro-climate regions in South Africa, as described in Chapter 3 (Subsections 3.1 and 3.2). Four stations Bethlehem, Nelspruit, Mahikeng, Cape Point and Prieska were excluded in the study because there were no sunshine observation data to be used as an input to Ångström–Prescott (AP) model.

5.3. Data

Observation GHI daily data and daily sunshine duration data were used, the datasets were described in Chapter 3 (subsection 3.2). Daily top-of-atmosphere (TOA) irradiance (GHI_{TOA}) and theoretical sunshine duration (N) were calculated using Equations (2.6) and (2.18) respectively from Chapter 2.

5.4. Methodology

The methodology used in this study is summarized in the flow chart in Figure 5.1 and 5.2.

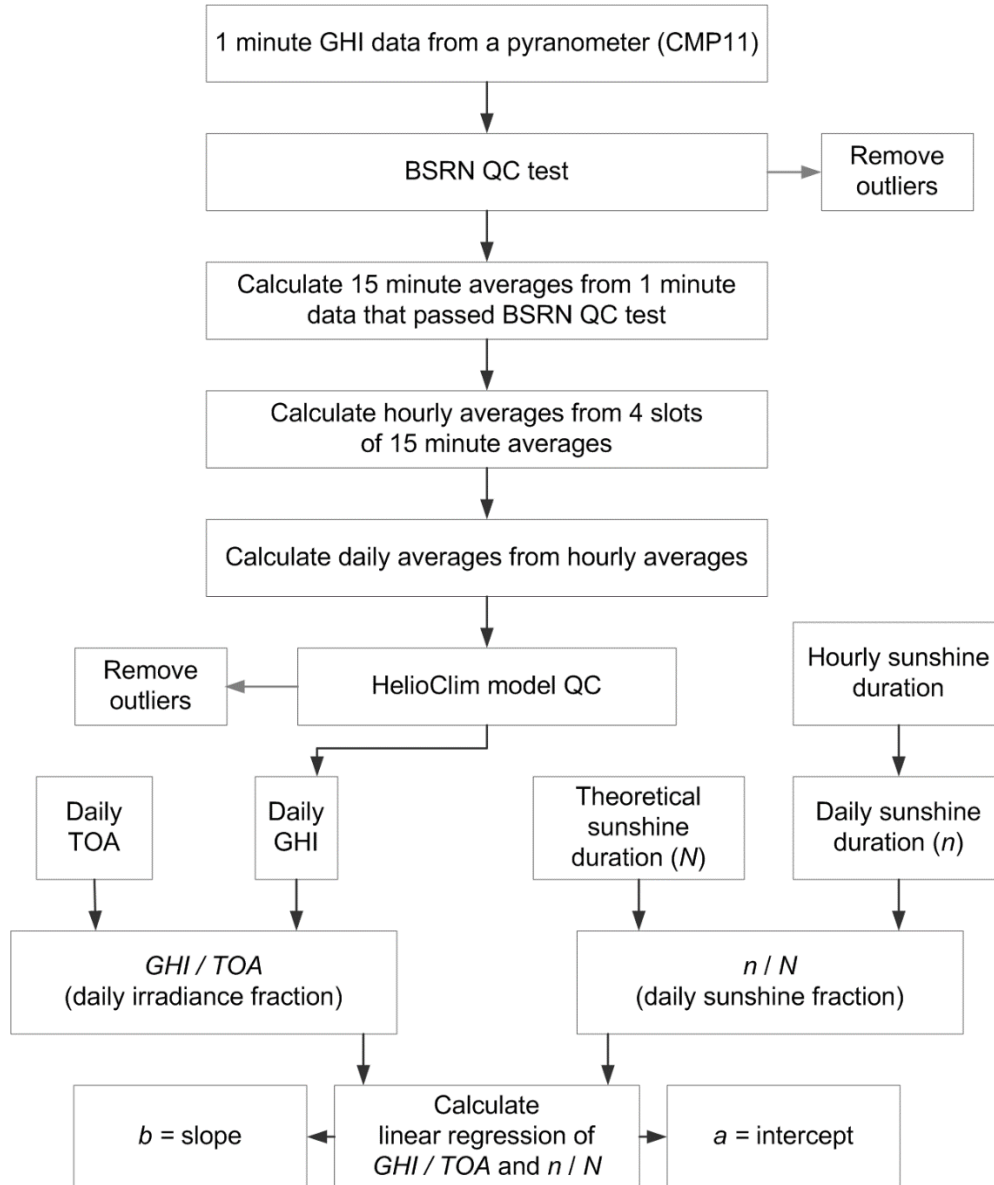


Figure 5.1. The flowchart of the approach used to calculate Ångström–Prescott regression coefficients a and b .

The Annual AP coefficients were calculated for eight stations. The observation periods for concurrent GHI and sunshine duration data for these stations are given in Table 3.2. Datasets up to the end of 2018 were used for determination of the AP coefficients, and the daily

observation data for 2019 were used to validate the corresponding estimated daily GHI data.

For Thohoyandou, the 2017 data were used to validate the coefficients.

```
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib import style
from statistics import mean
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn import metrics
from scipy import stats
from scipy.stats import linregress
import matplotlib.pylab as pylab
matplotlib.rcParams['figure.figsize'] = (12,8)
sns.set_palette(sns.color_palette("hls", 20))
sns.set_style("whitegrid")

obs_file = 'Station_Name.csv'
columns = ['DATE','X','Y']
obs = pd.read_csv(obs_file, header = None, skiprows = 1, names = columns)

obs['DATE'] = pd.to_datetime(obs['DATE'], format='%Y-%m-%d')
obs = obs.replace(np.inf, np.nan).replace(-np.inf, np.nan).dropna()

x = obs['X']
y = obs['Y']

def best_fit_slope(x,y):
    b = ((mean(x)*mean(y)) - mean(x*y)) /
        ((mean(x)**2) - mean(x**2))
    return b
b = best_fit_slope(x,y)
a = mean(y) - b*mean(x)

def xyline(x, y):
    xypts = min(x.min(), y.min()), max(x.max(), y.max())
    return xypts, xypts

np.polyfit(obs['X'], obs['Y'], 1)
regression_line=np.polyfit(obs['X'], obs['Y'], 1)
plt.plot(*xyline(x, y), 'k--')
plt.plot(x,regression_line[1]+regression_line[0]*x)
plt.title('Station name', fontsize=25)
plt.xlabel('n/N',fontsize=25)
plt.ylabel('GHI/TOA',fontsize=25)
plt.plot(x, (a + b*x), label = 'y = {:.3f} + {:.3f}*x'.format(a, b))
plt.legend(loc='lower left',fontsize=25)
plt.scatter(x, y, c=x, cmap='jet')
plt.colorbar()
plt.savefig("Station_name")
```

Figure 5.2. A Python code or script that was written and used to calculate and plot regression coefficients a and b from $X = \frac{n}{N}$ and $Y = \frac{GHI}{GHI_{TOA}}$.

The statistical metrics that were used to compare estimated daily GHI data with the observed daily GHI data are: relative mean bias error (rMBE), relative root mean square error (rRMSE), relative mean absolute error (rMAE), and the coefficient of determination (R^2). Described in subsection 4.3.

The results were converted from W/m^2 to $MJ\ m^{-2}d^{-1}$ by dividing by 11.57415, [i.e., $1\ W\ m^{-2} = 0.0864\ MJ\ m^{-2}\ day^{-1}$; $1\ MJ\ m^{-2}\ day^{-1} = (1\ W\ m^{-2})/0.0864 = 11.57415$], a methodology also used by Almorox et al. [4] to allow for easy comparison with other literature studies. Monthly averages of each metric were calculated and then aggregated to annual averages, and where observation data were not available, data were replaced by NaN. The annual AP coefficients a and b were calculated.

5.5. Results

5.5.1. Annual Ångström–Prescott (AP) Results

In this study, the annual AP regression coefficients a and b were calculated using Equation (5.1) and the following variables daily n , daily mean GHI, daily mean GHI_{TOA} , and daily N were used as inputs. The calculated a and b were then used together with daily n and N to estimate daily GHI, which was then compared to corresponding observed daily GHI. Statistical metrics in Equations 3.1 to 3.7 were used to quantify the errors between the two datasets; the results are shown in Table 5.1 and Figures 5.3-5.6.

In Figures 5.3 and 5.4, the annual AP coefficients and the data points that were used to derive them are displayed. The values of the AP coefficients ranged from 0.188 to 0.243 for a , while those for b ranged from 0.515 to 0.6. Values of $a = 0.25$ and $b = 0.5$ were recommended by

Allen et al. [5] to be used when there is no local observation *GHI* data to calibrate the coefficients. The minimum value of a in this study was less than 0.25 and the maximum value was greater than 0.25; the minimum and maximum values of b were greater than 0.5. The difference in default AP coefficients and calibrated AP coefficients proved that calibrating the coefficients locally is a necessity.

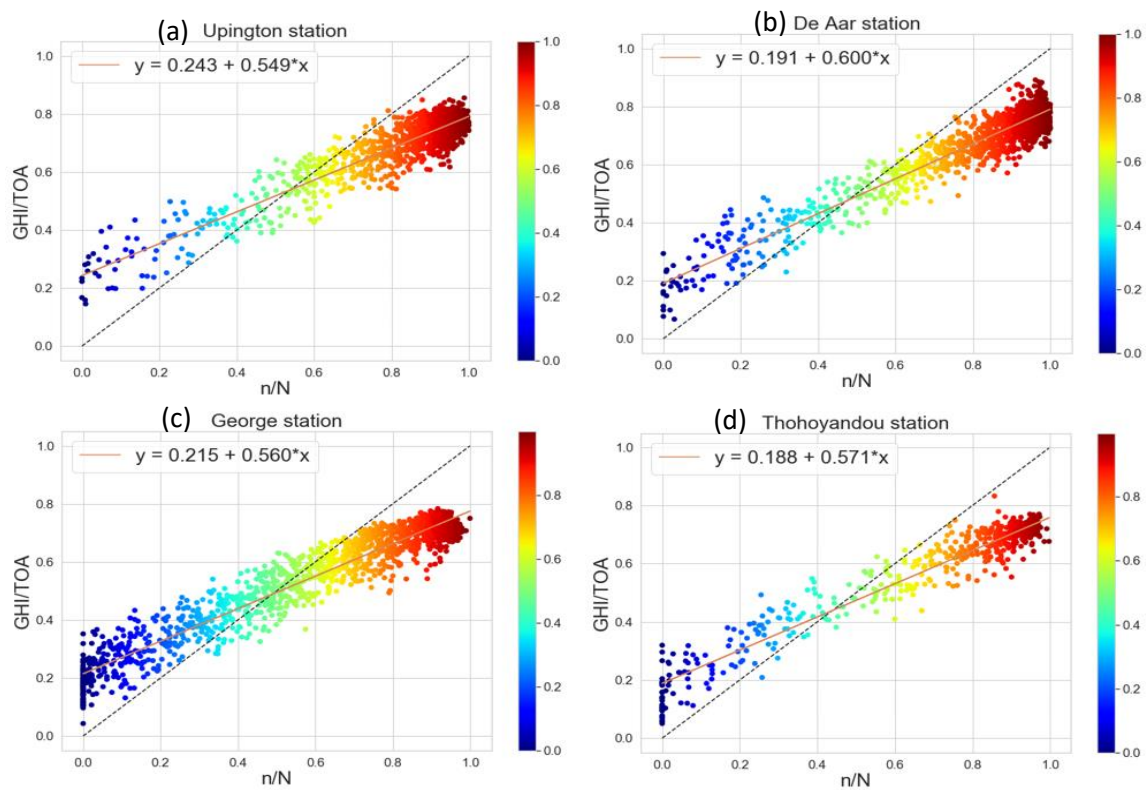


Figure 5.2 Regression lines and Ångström–Prescott coefficients for Upington station (a), De Aar (b), George (c), and Thohoyandou (d).

Studies done by Zhang et al., De Medeiros et al., Almorox et al., and Tsung et al. in [4,6-8] also found different results to Allen et al. [5] when they did a local calibration. The AP coefficients from this study are in line with the coefficients from similar studies done elsewhere in the world.

When comparing the factors for stations that are in the same climatological zone such as Irene and Polokwane located in the Temperature Interior climatic zone, and Mthatha and Durban in

the Tropical Coastal climatic zone (Table 1.1), the difference was less than 0.05 for both a and b , which is a very small difference. This means that the AP coefficients a and b calibrated for a climatic zone could be used as a representative for an entire climatic zone to estimate GHI when observed sunshine duration data for the location are available.

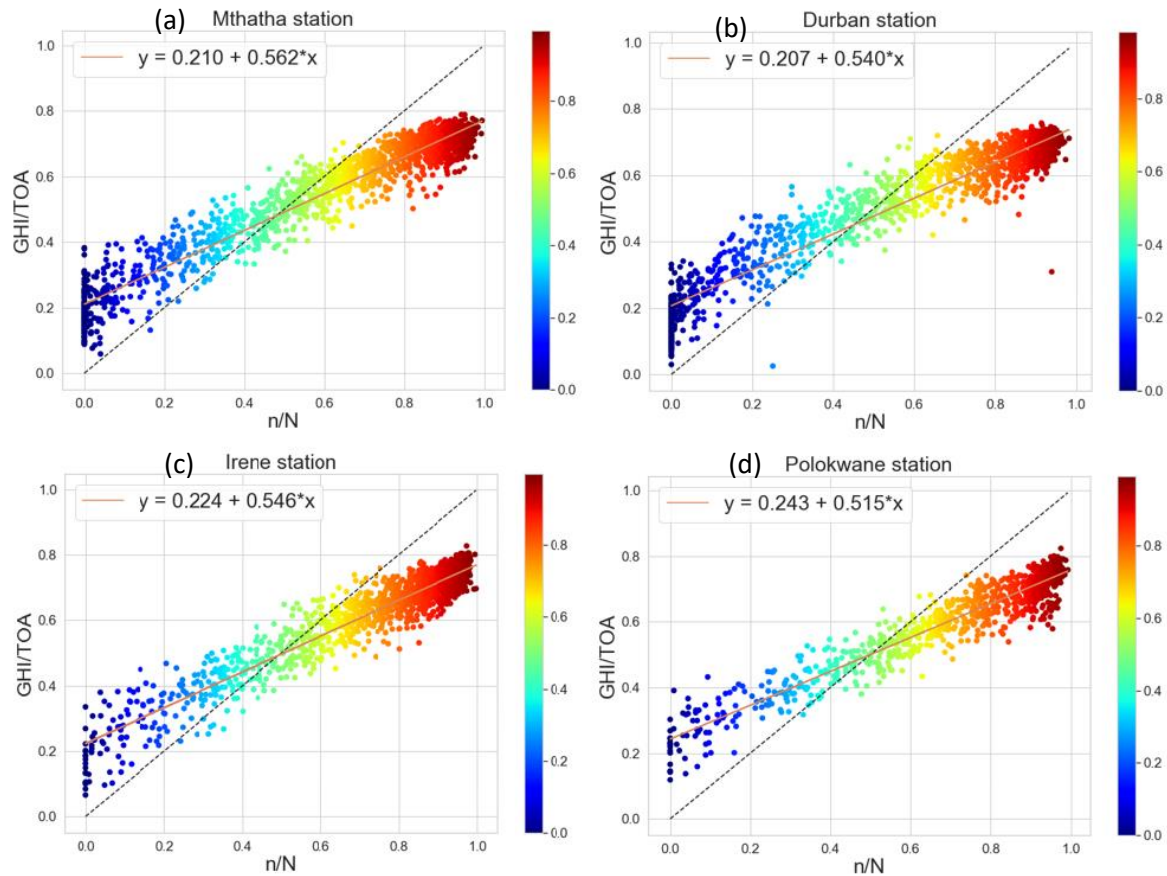


Figure 5.3 Regression lines and Ångström–Prescott coefficients for Mthatha station (a), Durban station (b), Irene station (c), and Polokwane (d).

5.5.2. Validation Results

Estimated GHI values were compared to the measured GHI values; errors were quantified by validation metrics in Equations 3.1 to 3.7 and the results were tabulated in Table 5.1. In Table 5.1, the rMBE ranged from -1.20 to 0.371 %, rMAE from 0.311 to 0.745 %, rRMSE

from 0.393 to 0.910 %, and R^2 from 0.910 to 0.948. De Aar, Irene, and Thohoyandou had a positive MBE, meaning that the model overestimated GHI, while Uppington, Durban, Mthatha, George, and Polokwane had a negative MBE, meaning that the model underestimated GHI values at these locations. The values of MBE and rMBE for all the stations were less than 1, indicating that there was a strong correlation between the predicted and observed GHI values. The worst case R^2 value was 0.910, suggesting that there is a very strong linear relation between observed and predicted values.

The maximum RMSE of $1.741 \text{ MJ m}^{-2}\text{d}^{-1}$ was less than 1.94 and $1.9 \text{ MJ m}^{-2}\text{d}^{-1}$ that De Medeiros et al. [6] and Tsung et al. [8], respectively, determined. The maximum MAE of $1.425 \text{ MJ m}^{-2}\text{d}^{-1}$ was less than $1.8 \text{ MJ m}^{-2}\text{d}^{-1}$ that Tsung et al. [8] determined. The maximum MBE of $0.733 \text{ MJ m}^{-2}\text{d}^{-1}$ was less than 1.040 and $0.85 \text{ MJ m}^{-2}\text{d}^{-1}$ that De Medeiros et al. [6] and Tsung et al. [8], respectively, determined, and the worst case R^2 of 0.910 was greater than 0.875, 0.74, and 0.58 that Zhang et al. [3], Adamala et al. [9], and De Medeiros et al. [6], respectively, determined. As shown in Table 5.1, the results across the range of climate zones (including the ones from the literature), all differ significantly.

Table 5.1: Calibration coefficients a and b and validation metrics results in ($\text{MJ m}^{-2}\text{d}^{-1}$).

Station	a	b	RMBE	rMBE (%)	MAE	rMAE (%)	RMSE	rRMSE (%)	R^2
Uppington	0.243	0.549	-0.360	-0.120	0.841	0.311	1.061	0.393	0.930
De Aar	0.191	0.600	0.733	0.371	1.136	0.506	1.375	0.598	0.930
Irene	0.224	0.546	0.689	0.353	1.328	0.608	1.618	0.729	0.912
Mthatha	0.210	0.562	-0.104	-0.013	1.168	0.582	1.474	0.735	0.951
George	0.215	0.560	-0.270	-0.036	1.261	0.636	1.520	0.769	0.948
Durban	0.207	0.540	-0.322	-0.106	1.425	0.745	1.741	0.910	0.915
Polokwane	0.243	0.515	-0.286	-0.085	1.272	0.488	1.572	0.606	0.910
Thohoyandou	0.188	0.571	0.286	0.168	1.071	0.550	1.433	0.746	0.937
Almorox et al.	0.287	0.452	-0.002	-	-	-	1.260	-	-

De Medeiros et al.	0.39	0.29	1.040	6.29	-	-	1.94	-	0.58
Tsung et al.	0.5	0.11	0.85	3.4	1.8	-	1.9	-	-
Zhang et al.	0.214	0.552	-	-	2.249	-	0.214	-	0.875
Adamala et al.	0.28	0.52	-	-	-	-	7.04	-	0.74

The overall validation results from this study are comparable and even better than what was found in similar studies like [4,6–9], which concluded that the AP coefficients could be used to estimate GHI with confidence based on those validation results. The data used in the study were collected using secondary standard pyranometers (CMP11), which, according to Urraca et al. [10], generate high quality records of GHI. Urraca et al. [23] found large and unstable validation errors from stations that uses second class pyranometers and silicon-based photodiodes compared to the ones using CMP11. This might be because CMP11 gives an integrated measurement of the total GHI available under all conditions due to its capability to measure the total solar spectrum from 0.3 to 3 micrometres wavelength. CMP11 pyranometers' use very high-quality quartz double domes, which improve the stability of a calibration factor over time; it also improves directional response and reduces thermal offsets. The CMP11 can give correct integrated values over a day, with the use of smaller sampling intervals. At SAWS, a sampling interval of 5 s is used, and this enables sudden changes such as passing small clouds, birds sitting on top of the pyranometers, and some other factors that result in shading of pyranometers to be identified and factored in the data. It is noted that smaller sampling intervals are not possible with CMP11 predecessors. The CMP11 has a specified expected daily uncertainty of less than 2 % [11,12] and Table 2.1.

GHI data were subjected to robust quality control methodologies BSRN QC [13] and HelioClim model QC [14] before any analysis and outliers which might affect the results were discarded, as described in Subsection 3.2. The use of Python codes in data analysis enabled big data to be handled much more efficiently and execution of a code in data analysis resulted in correct and consistent outputs; these are some of the reasons why the results in this study are

better. This means that the AP coefficients results from this study could also be used with confidence to estimate GHI in different climatological zones of South Africa.

In Figure 5.4, 2019 monthly GHI observed data were compared to corresponding estimated 2019 monthly GHI data. Thohoyandou is the only station where validation was done using 2017 monthly datasets and the observation data were only available from February to October (January, November, and December 2017 datasets were not available). The need to fill in missing data further motivates this study, i.e., development and validation of models, and the results of this study can be used to fill any missing monthly mean GHI values for South African locations.

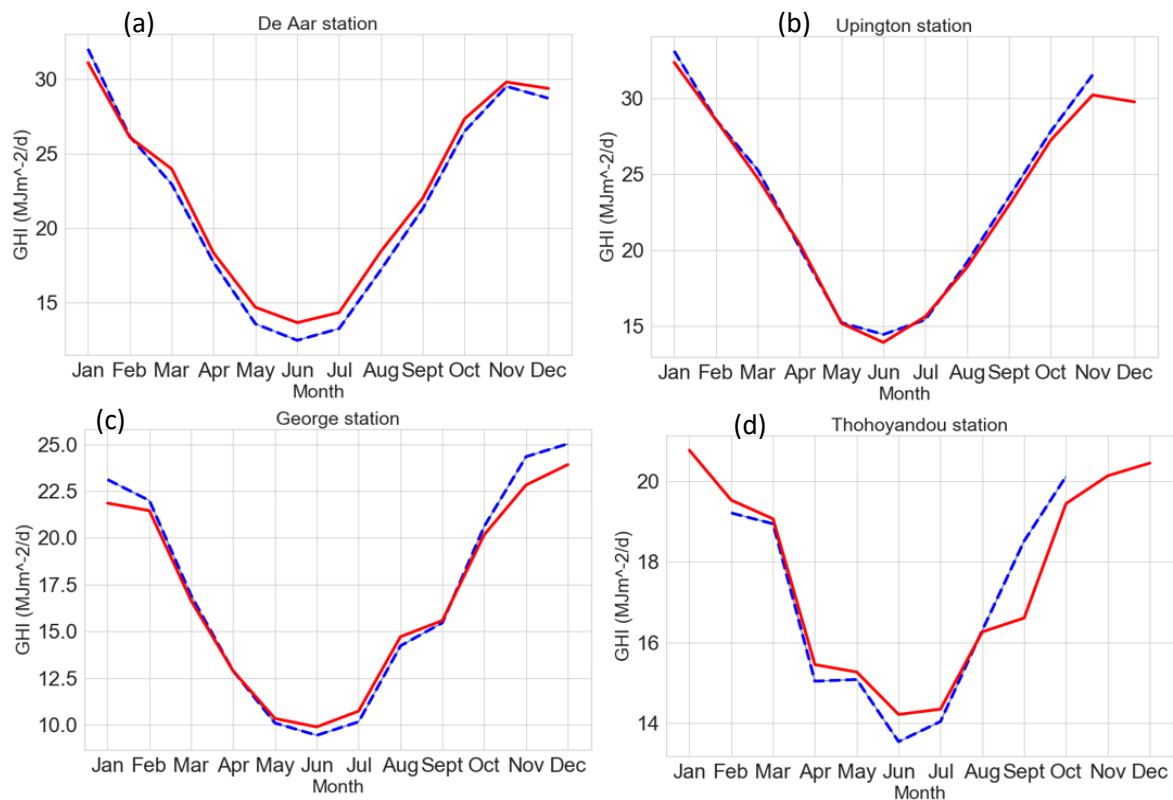


Figure 5.4. Comparison between measured and predicted monthly GHI values in De Aar (a), Upington (b), George (c), and Thohoyandou (d). Blue dotted line represents observed, GHI red solid line represents estimated GHI.

Similarly, in Figure 5.5, the 2019 monthly *GHI* observed data were compared to corresponding estimated 2019 monthly *GHI* data. In Durban, *GHI* observation data for September were not available. In Polokwane, the *GHI* observation data for March, April, May, and June were not available. The results from the study can be used to fill those missing monthly mean *GHI* values.

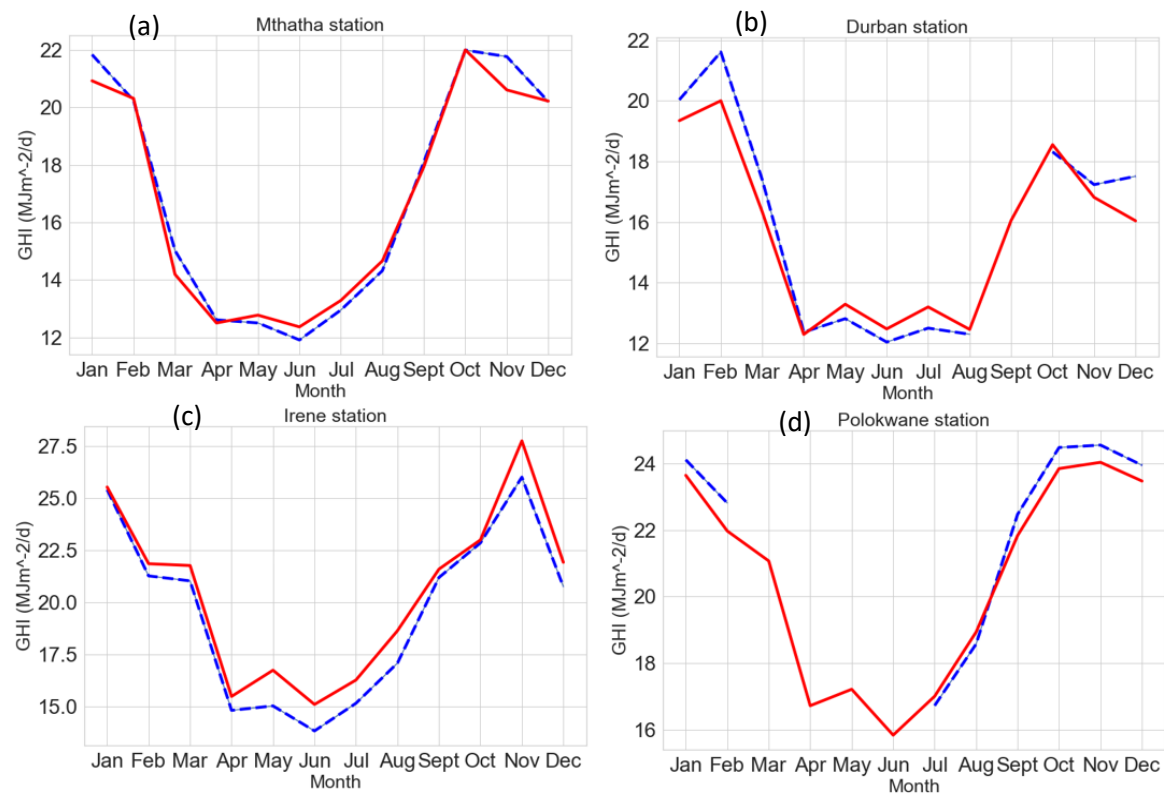


Figure 5.5. Comparison between measured and predicted *GHI* values in Mthatha (top left), Durban (top right), Irene (bottom left), and Polokwane (bottom right).

5.6. Chapter Summary

In chapter 5, The annual Ångström–Prescott coefficients a and b were calculated using the linear relationship between ratio of daily radiation on a horizontal surface to the daily extra-terrestrial radiation on that surface and the ratio of daily sunshine duration to the theoretical sunshine duration. They were used to estimate global horizontal irradiance and there was a very

close agreement with the corresponding observation global horizontal irradiance. The agreement was quantified by statistical metrics in Equations 3.1 to 3.7, i.e., relative Mean Bias Error, relative Mean Absolute Error, relative Root Mean Square Error, and correlation coefficient (R^2). The validation results were as follows: the rMBE ranged from -1.20 to 0.371 %, rMAE from 0.311 to 0.745 %, rRMSE from 0.393 to 0.910 %, and R^2 from 0.910 to 0.948 . The Ångström–Prescott coefficients calibrated for each station can be used as a representative for the climatic zone where that station is located. The Ångström–Prescott coefficients calculated in this study could enable the estimation of daily global horizontal irradiance data at any location in South Africa, where daily observation sunshine duration data are available, and the climate is correctly classified

References

1. Angstrom, A. Solar and terrestrial radiation: Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quart. J. Roy. Meteorol. Soc.* 1924, 50, 121–126, doi:10.1002/qj.49705021008.
2. Prescott, J.A. Evaporation from a water surface in relation to solar radiation. *Trans. Roy. Soc. S. Aust.* 1940, 46, 114–118. [<https://ci.nii.ac.jp/naid/10025613338/>].
3. Iqbal, M. *An Introduction to Solar Radiation*; Academic Press: Toronto, ON, Canada, 1983.
4. Almorox, J.; Benito, M.; Hontoria, C. Estimation of monthly Angström-Prescott equation coefficients from measured daily data in Toledo, Spain. *Renew. Energy* **2005**, 30, 931–936, doi:10.1016/j.renene.2004.08.002.
5. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration-Guidelines for Computing Crop Requirements-Irrigation and Drainage Paper 56*; FAO: Rome, Italy, 1998; Volume 300, p. D05109. Available online: https://appgeodb.nancy.inra.fr/biljou/pdf/Allen_FAO1998.pdf (accessed on 15 February 2020)
6. De Medeiros, F.J.; Silva, C.M.; Bezerra, B.G. Calibration of Ångström-Prescott Equation to Estimate Daily Solar Radiation on Rio Grande do Norte State, Brazil. *Rev. Bras. Meteorol.* 2017, 32, 409–416, doi:10.1590/0102-77863230008.

7. Zhang, Q.; Cui, N.; Feng, Y.; Jia, Y.; Li, Z.; Gong, D. Comparative Analysis of Global Solar Radiation Models in Different Regions of China. *Adv. Meteorol.* **2018**, 1–21, doi:10.1155/2018/3894831.
8. Tsung, K.Y.; Tan, R.; Li, G.Y. Estimating the global solar radiation in Putrajaya using the Angstrom-Prescott model. *IOP Conf. Series: Earth Environ. Sci.* **2019**, 268, 012056. [<https://iopscience.iop.org/article/10.1088/1755-1315/268/1/012056/meta>]
9. Adamala, S.; Reddy, Y.V.K. Evaluation of Different Solar Radiation Estimation Methods for Indian Locations. *Water Resour. Environ. Eng. II.* **2018**, 47–56, doi:10.1007/978-981-13-2038-5_5.
10. Urraca, R.; Gracia-Amillo, A.M.; Huld, T.; Martinez-De-Pison, F.J.; Trentmann, J.; Lindfors, A.V.; Riihelä, A.; Sanz-Garcia, A. Quality control of global solar radiation data with satellite-based products. *Sol. Energy* 2017, 158, 49–62, doi:10.1016/j.solener.2017.09.032.
11. CMP11 Pyranometer. Available online: <https://www.campbellsci.eu/cmp11> (accessed on 17 March 2020).
12. CMP11 Pyranometer. Available online: <https://www.kippzonen.com/Product/13/CMP11-Pyranometer#.X1UGh-fivIV> (accessed on 12 May 2020).
13. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2. 2010. Available online: https://epic.awi.de/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 11 December 2019).
14. Geiger, M.; Diabaté, L.; Ménard, L.; Wald, L. A web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy* **2002**, 73, 475–480, doi:10.1016/S0038-092X(02)00121-4.

CHAPTER 6: Validating Hourly Satellite Based and Reanalysis Based Global Horizontal Irradiance Datasets over South Africa

6.1. Introduction

This chapter aims to answers a problem statement:

1. Which is the best freely available (for research purposes) GHI satellite-based datasets and GHI reanalysis-based dataset in each location/station under study and overall, in South Africa? What are the benchmarks for hourly (rMBE, rRMSE, rMAE and R^2) in study based on the distribution of the results from the study?

The problem statement was addressed by setting the following objectives:

1. Validate and rank (based on the performance) hourly satellite-based global horizontal irradiance (GHI) datasets namely: SOLCAST, Copernicus Atmosphere Monitoring Service (CAMS), and Satellite Application Facility on Climate Monitoring (CMSAF SARAH) and two reanalysis-based, namely, Fifth generation European Center for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) and Modern-Era Retrospective Analysis for Research and Applications (MERRA2) for sites in South Africa.

The work presented in this chapter has been submitted on 9 September 2021 to Geomatics (ISSN 2673-7418), an international and interdisciplinary peer-reviewed open access journal of related scientific research , reviews, technical notes and communications covering all aspects surveying and mapping, positioning and navigation, satellite positioning, geodesy, photogrammetry, cartography and cadastral, hydrography, remote sensing, spatial data analysis, spatial information science, and geographic information system. The Journal publishes scientific research work under the Multidisciplinary Digital Publishing Institute

(MDPI) publisher with its headquarters based in Basel, Switzerland. The manuscript was reviewed by experts in the field and accepted on 3 November 2021 by the Journal and it was published on 5 November 2021. The work presented herein according to the submission guidelines suggested by the Journal editors. The Geomatics journal (ISSN 2673-7418) is accredited by the Department of Higher Education and Training (DHET).

Mabasa, B.; Lysko, M.D.; Moloi, S.J. Validating Hourly Satellite Based and Reanalysis Based Global Horizontal Irradiance Datasets over South Africa. *Geomatics* **2021**, *1*, 429–449. <https://doi.org/10.3390/geomatics1040025>.

Abstract: This study validates the hourly satellite based and reanalysis based global horizontal irradiance (GHI) for sites in South Africa. Hourly GHI satellite based namely: SOLCAST, Copernicus Atmosphere Monitoring Service (CAMS), and Satellite Application Facility on Climate Monitoring (CMSAF SARA) and two reanalysis based, namely, Fifth generation European Center for Medium Range Weather Forecasts atmospheric reanalysis (ERA5) and Modern-Era Retrospective Analysis for Research and Applications (MERRA2) were assessed by comparing in-situ measured data from 13 South African Weather Service radiometric stations, located in the country's six macro climatological regions, for the period 2013–2019. The in-situ data were first quality controlled using the Baseline Surface Radiation Network methodology. Data visualization and statistical metrics relative mean bias error (rMBE), relative root mean square error (rRMSE), relative mean absolute error (rMAE), and the coefficient of determination (R^2) were used to evaluate the performance of the datasets. There was very good correlation against in situ GHI for the satellite based GHI, all with R^2 above 0.95. The R^2 correlations for the reanalysis based GHI were less than 0.95 (0.931 for ERA5 and 0.888 for MERRA2). The satellite and reanalysis based GHI showed a positive rMBE (SOLCAST 0.81 %, CAMS 2.14 %, CMSAF 2.13 %, ERA5 1.7 %, and MERRA2 11 %), suggesting consistent overestimation over the country. SOLCAST satellite based GHI showed

the best rRMSE (14 %) and rMAE (9 %) combinations. MERRA2 reanalysis based GHI showed the weakest rRMSE (37 %) and rMAE (22 %) combinations. SOLCAST satellite based GHI showed the best overall performance. When considering only the freely available datasets, CAMS and CMSAF performed better with the same overall rMBE (2 %), however, CAMS showed slightly better rRMSE (16 %), rMAE (10 %), and R^2 (0.98) combinations than CMSAF rRMSE (17 %), rMAE (11 %), and R^2 (0.97). CAMS and CMSAF are viable freely available data sources for South African locations.

Author Contributions: The contribution of the authors is summarised below.

Conceptualization: Brighton Mabasa. **Methodology:** Brighton Mabasa and Meena D. Lysko. **Software:** Brighton Mabasa. **Validation:** Meena D. Lysko. **Formal analysis:** Brighton Mabasa and Meena D. Lysko. **Investigation:** Brighton Mabasa. **Data curation:** Brighton Mabasa. **Writing—original draft preparation:** Brighton Mabasa. **Supervision:** Meena D. Lysko. and Sabata J. Moloi. **Writing—review and editing:** Brighton Mabasa.; Meena D. Lysko and Sabata J. Moloi. All authors have read and agreed to the published version of the manuscript.

Validation of satellite-based, and reanalysis-based GHI datasets is important as they provide a wider coverage of solar radiation. Validated satellite and reanalysis-based GHI datasets can be used to provide reliable alternative GHI data and compensate for the scarcity of monitoring stations by increasing the density of GHI data. This study aims to quantify the errors between the in situ measured GHI and estimated gridded datasets such as to validate satellite-based datasets (SOLCAST, CAMS, and CMSAF SARA) and reanalysis-based datasets (ERA5 and MERRA2) relative to quality controlled in situ data from 13 reference stations managed by the

South African Weather Services (SAWS). The validation was conducted on an hourly temporal scale over all six macro climatic zones of South Africa. This can be applied in studies for overall local accuracy of the datasets to be evaluated.

Validated datasets could be used to estimate GHI in the long-term and over a wide spatial resolution in South Africa. This will enable climate studies, which is generally not possible with ground observed data because there are no continuous long-term records covering decades and covering all areas of the country. The validated datasets could also be used as an additional quality control parameter of the measured data. The bias information from different sources of GHI in different areas of the country can be used as a basis for bias correction. The bias corrected data sources could also be merged with measured in situ data to produce GHI maps.

6.2. Study Area

The study area was determined by considering the locations of the South African Weather Services (SAWS) radiometric stations which measure the 1-minute GHI data. The area of study is thirteen sites located in all six macro-climate regions in South Africa, as described in Chapter 4 (Subsections 4.1 and 4.2).

6.3.1. Reanalysis-based Datasets

The choice of reanalysis-derived datasets was based on the general free availability. The spatial and temporal characteristics as well as the level of accessibility and applicable regions of the reanalysis datasets are given in Table 6.1. Reanalysis-derived datasets are generated by assimilating historical observation data from various platforms (ground observation, satellites, ships, and aircrafts) and numerical weather prediction models using a consistent algorithm.

Reanalysis-derived datasets have an advantage of global spatial coverage and long-time series [1].

ERA5, as described by Hersbach et al. (2020) [2], is the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF, Reading, United Kingdom) atmospheric reanalysis of the global climate. ERA5 has a spatial resolution of $0.25^\circ \times 0.25^\circ$ and has an hourly time resolution. It uses a new advanced model cycle for Integrated Forecasting System (IFS Cycle 41r2) for data assimilation, which increased the computational efficiency and forecast accuracy. ERA5 uses climatological aerosol information, and the Global Ozone Chemistry Aerosol Radiation and Transport (GEOCART) stratospheric sulfate aerosol from volcanic eruptions is also included [2]. The dataset covers a period from 1979 to the present, hourly data are available in Joule per square meter (J/m^2), so to obtain the hourly values in watt per square meter (W/m^2), values should be divided by 3600 (number of seconds in one hour).

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA2), which was introduced by Gelaro et al. (2017) in [3], it assimilates space-based observations of aerosols and represents their interactions with other physical processes in the climate system. MERRA2 products are generated using the GEOS 5.12.4 model, which uses real time aerosols as inputs. MERRA2 data are available from 1980 to the present with two months delay, has a spatial resolution of $0.5^\circ \times 0.625^\circ$, and has an hourly resolution for surface irradiance variables [3].

6.3.2. Satellite-based Datasets

The choice of satellite-based datasets was based on the general free availability and availability for research purposes. Information on the satellite-based datasets for this study is given in Table 6.1. The relationship between the satellite images and the actual ground GHI was established with an algorithm that combines cloud information, which is generated from a clear sky model

and cloud index. Aerosol, water vapor, linke turbidity, and ozone information were also added as inputs to an algorithm, which will then generate estimated GHI at the surface. The technique is also described in the literature including Ameen et al. (2018) [4].

The Surface Solar Radiation Data Record—Heliosat Edition 2 (SARAH) dataset, as described by Pfeifroth et al. (2019) [5] is provided by the EUMETSAT Satellite Application Facility on Climate Monitoring (CMSAF) and covers the time period from 1983 to present, with the temporal resolution ranging from 30-minute instantaneous values. Amongst the other products, the dataset provides GHI. CMSAF SARAH datasets are derived from the geostationary METEOSAT satellite service of the first and second generation, which are geostationary over Europe, Africa, and a small part of South America. In this case, the data were retrieved by using the Heliosat method to estimate the cloud index, clear sky radiative transfer model, and several climatological parameters (precipitable water vapor, monthly AOD climatology, monthly ozone climatology, and ground albedo). CMSAF SARAH data have a higher stability in early years due to the removal of erroneous satellite images during the transition from the first to the second generation METEOSAT satellite. CMSAF, as also given in Table 6.1, has a spatial resolution of $0.05^\circ \times 0.05^\circ$ [5].

CMSAF SARAH was validated using twelve BSRN stations from three continents, eight stations from Europe, one station from South America, and three from Africa (South Africa, Algeria, and Namibia). The validating irradiance datasets were first quality controlled using BSRN methodology and the outliers were discarded [6].

The Copernicus Atmosphere Monitoring Service (CAMS) radiation service is part of the Copernicus Program, an Earth observation program coordinated and managed by the European Commission in partnership with the European Space Agency. The CAMS radiation service is available for free via the CAMS [7] and solar radiation data (SoDa) [8]. It makes use of the

Heliosat-4 method [9], which models the radiative transfer in the atmosphere to compute the solar radiation parameters. The McClear model [10] is used to estimate clear sky irradiance. The AVHRR Processing scheme Over cLOUDs, Land and Ocean (APPOLLO) method is used to process the satellite images from the German Aerospace Center database to yield information (cloud coverage, cloud level, and cloud type) for each pixel (3 km at nadir) every 15 min [9]. Ground albedo from Moderate Resolution Imaging Spectroradiometer (MODIS) is used, as given in Qu et al. (2017) [9]. The CAMS radiation service provides a time series of GHI with a spatial coverage of -66° to 66° in both latitudes and longitudes (i.e., North to South and East to West directions). The data are interpolated to the point of the user's interest. The time coverage of data was from 1 February 2004 to date with a two day delay and with a temporal resolution that ranges from 1-minute average, 15-minute averages, hourly averages, daily averages, and monthly averages [7-9].

The CAMS satellite dataset is regularly validated using global BSRN and non BSRN stations. The validating in-situ irradiation datasets were first quality controlled using the BSRN methodology and outliers were discarded. The recent validation process in 2020 used 32 in-situ stations and only one station from Southern Africa, Namibia (Gobabeb), was used in the recent validation process [10].

SOLCAST is a commercial company [11-13]. The SOLCAST method estimates solar irradiance from satellites by detecting cloud cover and characterizing cloud cover in terms of its impact on solar radiance, modelling the available solar irradiance under clear skies and then combining the estimate of the amount of solar irradiance reaching the Earth's surface after it passes through the clouds. In Africa, the EUMETSAT Meteosat satellite and REST2v5 clear sky model [14] with MERRA2 reanalysis inputs (for metadata) were used in SOLCAST GHI data estimation. The temporal resolution ranged from basic time series (hourly averages) and alternative time series (5 minutes, 10 minutes, 15 minutes, and 30 minutes). One-minute data

are also available on request. The data are available from January 2007 to date with a seven day delay, through the SOLCAST website [11]. SOLCAST provides global coverage of the data, except for ocean and polar regions, with a spatial resolution of 1–2 km [11-13]. According to the website on SOLCAST validation and accuracy [15], SOLCAST has considered 46 of the BSRN sites for validation of GHI and reports a maximum bias deviation of 2.01 %. The BSRN sites for steep mountain areas, oceanic, and polar sites were excluded in the SOLCAST validation.

Table 6.1. Summary of the satellite-based and reanalysis-based datasets.

Data	Data derived from	Time Period	Spatial Resolution	Temporal Resolution	Data Availability	Region Available
SOLCAST [11]	satellite	2007 to present	1–2 km	1 hour	Not free	Almost Global (except Polar regions and oceans)
CMSAF SARAH [5]	satellite	1983 to present	0.05° x 0.05° (5 km)	½ hour, 1 day	Free	Europe, Africa, and a small part of South America
CAMS [7]	satellite	2004 to (current day - 2 days)	*Interpolated to a point of interest	1 minute, 15 minutes, 1 hour, 1 day, 1 month	Free	Europe, Africa, Middle East, Eastern part of South America and Atlantic Ocean
ERA5 [2]	reanalysis	1979 to present	0.25° x 0.25° (31 km)	1 hour	Free	Global
MERRA2 [3]	reanalysis	1980 to (present—2 months)	0.625° x 0.5° (50 km)	1 hour	Free	Global

6.4. Methodology

The methodology used in this study is summarized in the flow chart in Figure 6.1. The methodology consists of the following steps: (1) Pre-processing and quality control of in-situ GHI data; (2) Averaging one minute to 15 minutes; (3) Averaging four slots of 15 minute to obtain hourly averages; (4) Gathering and pre-processing of satellite-based and reanalysis datasets; (5) Matching common time steps of in situ GHI and satellite/reanalysis-based datasets; (6) Calculation of hourly zenith angles and removing datasets that fall on points where the zenith angle is greater than 90 degrees; and (7) Calculation of statistical metrics.

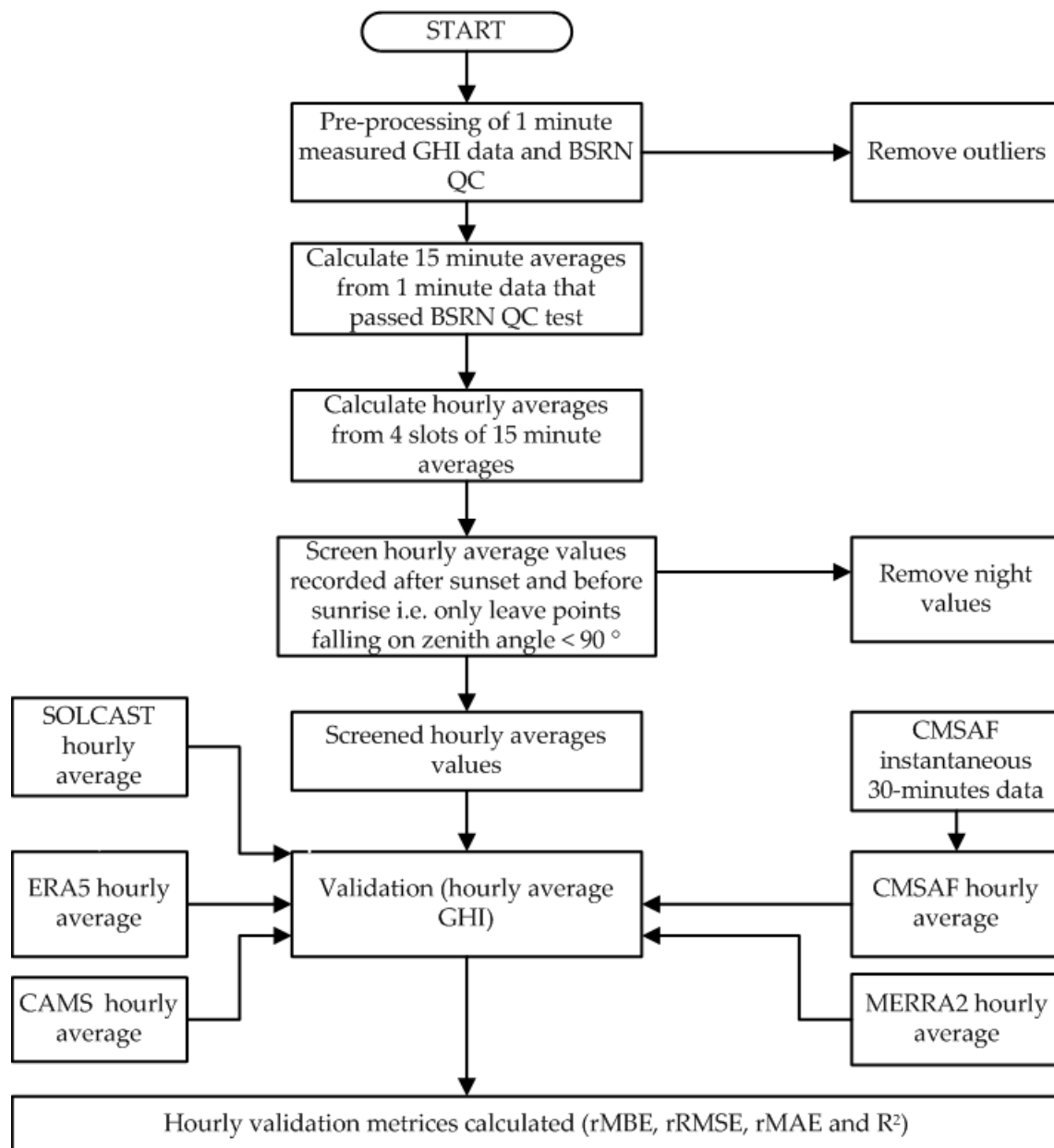


Figure 6.1. The flowchart summarizes the approach used from data preprocessing to data validation.

6.4.1. Pre-Processing of Observation Data and Validation Process

The pre-processing and validation of observation data was described in subsections 3.1 and 3.2.

Hourly average values were then compared to corresponding hourly average CAMS, CMSAF SARAH, SOLCAST, ERA5, and MERRA2. SOLCAST, ERA5, MERRA2, and CAMS were sourced as hourly averages while CMSAF SARAH was sourced as instantaneous 30-minute averages and then 2-time steps were averaged to obtain hourly averages.

6.4.2. Statistical Metrics

The statistical metrics that were used to quantify the difference between hourly estimated and hourly measured GHI are relative mean bias error (rMBE), relative root mean square error (rRMSE), relative mean absolute error (rMAE), and the coefficient of determination (R^2). Described in subsection 3.3.

6.4.3. Most Feasible Gridded Dataset

The most feasible option from the satellite-based and reanalysis datasets at each of the thirteen stations is determined by first considering the best performing model for each metric. The most feasible option at a station is then the model with the maximum count across all the metrics. A numerical relative rating of the optimal metric is calculated as the maximum count relative to the four metrics (rMBE, rRMSE, rMAE, and R^2).

6.5. Results

6.5.1. CAMS Results

From Figure 6.2, CAMS underestimated GHI in Upington (−3 %), Prieska (−0.1 %), De Aar (−2 %), and Cape Point (−0.1 %); and overestimated GHI at the remaining nine stations, rMBE

ranged from 2 %–6 %. From Figure 6.3, hourly rRMSE was less than and slightly above 20 % for all stations. When considering the correlation R^2 results in Table 6.2, CAMS had $R^2 > 0.96$ in all stations, suggesting that measured and estimated data correlate well. Good correlation was also demonstrated by scatterplots in Figures 6.9 to 6.21.

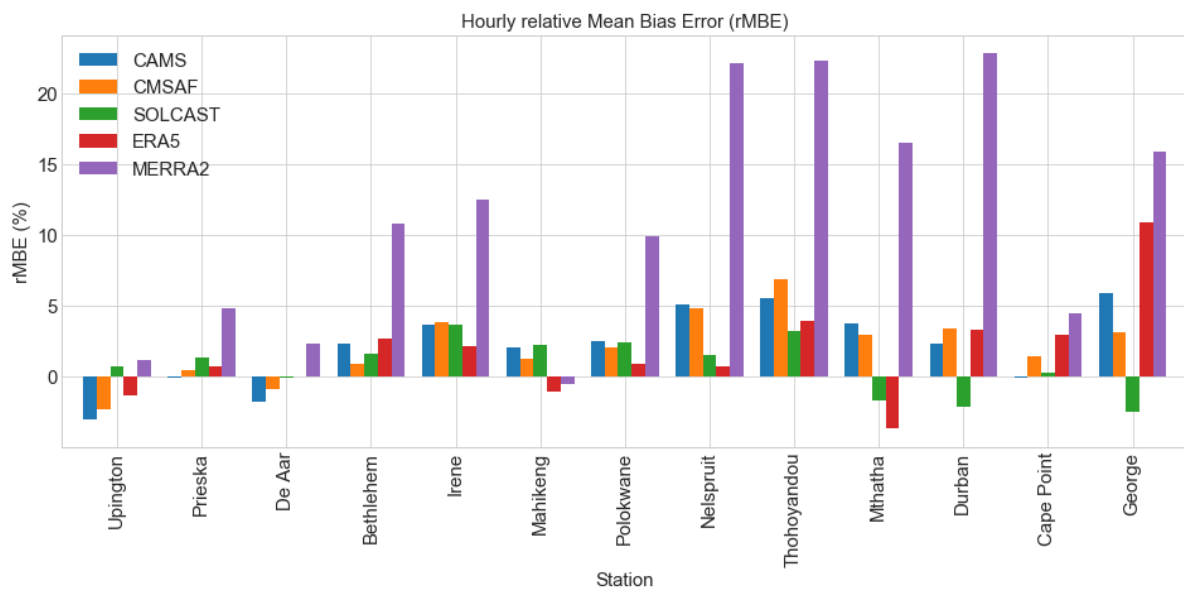


Figure 6.2. Hourly relative mean bias error of gridded datasets against measured in situ GHI.

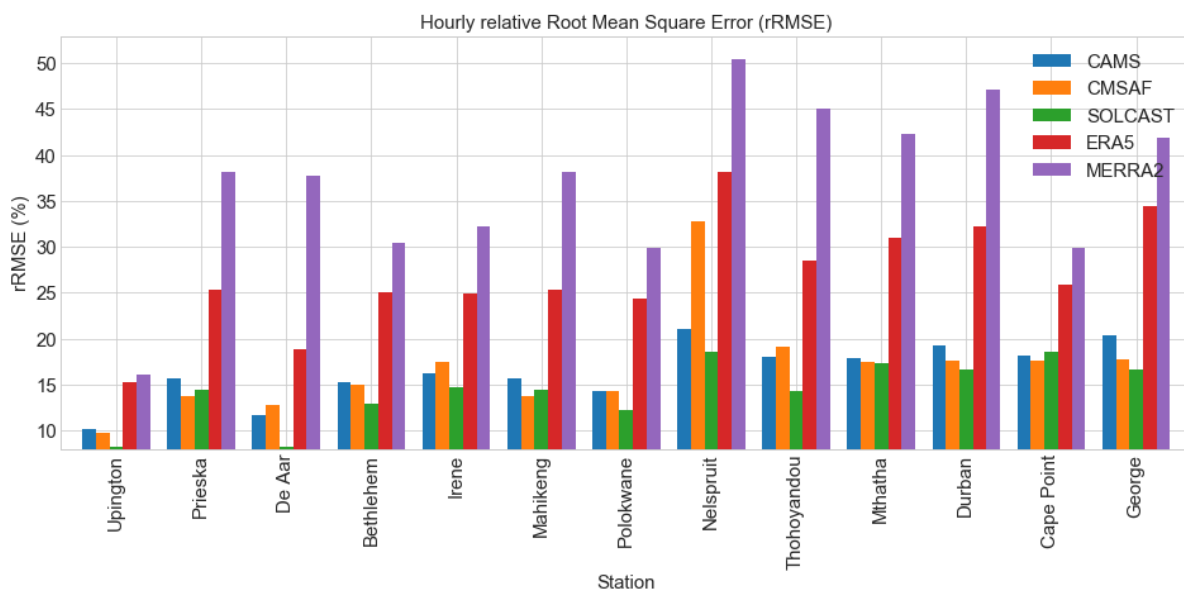


Figure 6.3: Hourly relative root mean square error of gridded datasets against measured in situ GHI.

From Figures 6.2-6.4 and Table 6.2, the hourly metrics for CAMS varied as given below:

- $-3 \% \leq \text{hourly } rMBE \leq 6 \%$;
- $10 \% \leq \text{hourly } rRMSE \leq 21 \%$;
- $6 \% \leq \text{hourly } rMAE \leq 13 \%$; and
- $0.962 \leq \text{hourly } R^2 \leq 0.995$.

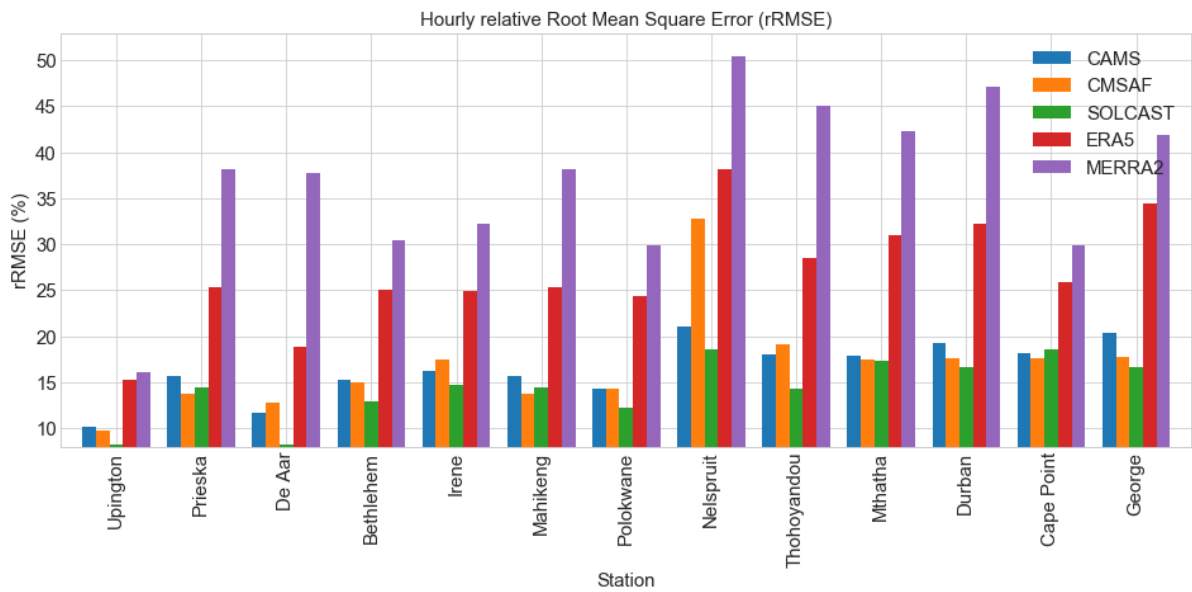


Figure 6.4. Hourly relative mean absolute error of gridded datasets against measured in situ GHI.

Table 6.2. Hourly mean measured GHI (W/m^2) and correlation (R^2) of gridded datasets against measured in situ GHI. Green color represents the best ($R^2 > 0.95$), blue color represents the intermediate ($0.9 < R^2 < 0.95$), and orange color represents the poor correlation ($R^2 < 0.9$).

Station	Mean GHI (W/m^2)	CAMS	CMSAF	SOLCAST	ERA5	MERRA2
Upington	522.86	0.995	0.995	0.996	0.987	0.985
Prieska	490.95	0.986	0.983	0.991	0.962	0.955
DeAar	497.67	0.984	0.981	0.992	0.958	0.833
Bethlehem	459.50	0.976	0.976	0.983	0.932	0.914
Irene	458.15	0.973	0.969	0.979	0.932	0.907
Mahikeng	492.13	0.972	0.978	0.977	0.923	0.826

Polokwane	466.63	0.978	0.977	0.985	0.932	0.910
Nelspruit	404.24	0.962	0.908	0.970	0.868	0.823
Thohoyandou	408.46	0.972	0.970	0.983	0.922	0.856
Mthatha	382.77	0.975	0.976	0.975	0.915	0.873
Durban	364.64	0.972	0.977	0.979	0.919	0.871
Cape Point	415.41	0.970	0.972	0.969	0.938	0.920
George	387.46	0.968	0.974	0.977	0.909	0.875

6.5.2. CMSAF Results

From Figure 6.2, CMSAF underestimated GHI in Upington (−2 %) and De Aar (−1 %) and overestimated GHI at the remaining 11 stations; rMBE ranged from 0 %–7 %. In Figure 6.3, when considering rRMSE, CMSAF showed a very good performance at Upington (10 %), the worst performance at Nelspruit (33 %), and the rest of the stations had a rRMSE between 10 % and 20 %. From Table 6.2, when considering the hourly correlation R^2 results, CMSAF had $R^2 > 0.96$ at all stations except in Nelspruit (0.908), which shows a very good correlation between the in situ observed and estimated data. This is also demonstrated by scatterplots in Figures 6.9 to 6.21.

From Figures 6.2–6.4 and Table 6.2, the hourly metrics for CMSAF varied as given below:

- $-2 \% \leq \text{hourly } rMBE \leq 7 \%$;
- $10 \% \leq \text{hourly } rRMSE \leq 33 \%$;
- $7 \% \leq \text{hourly } rMAE \leq 19 \%$; and
- $0.908 \leq \text{hourly } R^2 \leq 0.995$.

Table 6.3 Best performing gridded dataset per hourly metric, most feasible model, and level of performance (rating). The colors were used to show the best performing gridded dataset based on the hourly statical metric per station and the most feasible dataset per station-based level of rating out of 4 (the number of statistical metrics used). Green represents SOLCAST, yellow CAMS, blue CMSAF, orange ERA5, and red MERRA2.

Hourly	Minimum rMBE	Minimum rRMSE	Minimum rMAE	Maximum R^2	Most Feasible	Rating
Upington	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
Prieska	CAMS	CMSAF	SOLCAST	SOLCAST	SOLCAST	2/4
De Aar	ERA5	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Bethlehem	CMSAF	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Irene	ERA5	SOLCAST	CAMS	SOLCAST	SOLCAST	2/4
Mahikeng	MERRA2	CMSAF	CMSAF	CMSAF	CMSAF	3/4
Polokwane	ERA5	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Nelspruit	ERA5	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Thohoyandou	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
Mthatha	SOLCAST	SOLCAST	CAMS	CMSAF	SOLCAST	2/4
Durban	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
Cape Point	CAMS	CMSAF	SOLCAST	CMSAF	CMSAF	2/4
George	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4

6.5.3. SOLCAST Results

From Table 6.3, which gives the most feasible hourly dataset per station by combining all the hourly metrics, SOLCAST was most prominent in 10 out of 13 stations for the hourly aggregates. From Figure 6.2, the SOLCAST dataset slightly underestimated hourly GHI at De Aar (−0.1 %), Mthatha (−2%), George (−3 %), and Durban (−2 %) and it overestimated hourly GHI at the rest of the stations; rMBE ranged from 1 % to 4 %. From Table 6.2, SOLCAST had the correlation $R^2 > 0.96$ in all 13 stations, showing a very good agreement between the measured and estimated data. This was also demonstrated by the scatterplots in Figures 6.9 to 6.21. From Figure 6.3, SOLCAST hourly rRMSE was less than 15 % at eight stations and more than 15 % in five stations George (17 %), Cape Point (19 %), Durban (17 %), Mthatha (17 %), and Nelspruit (19 %).

From Figure 6.2-6.4 and Table 6.2, the hourly metrics for SOLCAST varied as given below:

- $-3 \% \leq \text{hourly rMBE} \leq 4 \%$;
- $8 \% \leq \text{hourly rRMSE} \leq 19 \%$;
- $5 \% \leq \text{hourly rMAE} \leq 12 \%$ and
- $0.969 \leq \text{hourly } R^2 \leq 0.996$.

6.5.4. ERA5 Results

From Figure 6.3, two stations Upington (15 %) and De Aar (19 %) had a rRMSE less than 20 %. Seven stations had rRMSE between 20% and 30%, and four stations, namely Mthatha (31 %), Durban (32%), George (34 %), and Nelspruit (38 %) had rRMSEs greater than 30 %. When considering hourly R^2 results, three stations namely Upington (0.987), Prieska (0.962), and De Aar (0.958) had $R^2 > 0.95$, and nine stations had $0.9 < R^2 < 0.95$. Nelspruit (0.868) was the only station with hourly $R^2 < 0.9$. The poor correlation of ERA5 reanalysis hourly data was also demonstrated by scatterplots in Figures 6.9 to 6.21, given in the Supplementary Materials, with data points not elongated along the 1:1 line. From Figure 6.2, ERA5 reanalysis data underestimated hourly GHI in Upington (−1 %), Mahikeng (−1 %), and Mthatha (−4 %) and overestimated in 10 stations with rMBE ranging from 0 % to 11 %.

From Figure 6.2-6.4 and Table 6.2, the hourly metrics for ERA5 varied as given below:

- $-4 \% \leq \text{hourly rMBE} \leq 11 \%$;
- $15 \% \leq \text{hourly rRMSE} \leq 38 \%$;
- $9 \% \leq \text{hourly rMAE} \leq 25 \%$ and
- $0.868 \leq \text{hourly } R^2 \leq 0.987$.

6.5.5. MERRA2 Results

From Figure 6.2, the MERRA2 reanalysis hourly dataset overestimated GHI in 12 stations, $rMBE$ ranged from 1 % to 23 %, and slightly underestimated GHI in Mahikeng (-0.54 %). From the $rRMSE$ s given in Figure 6.3, Upington (16 %), Cape Point (29.86 %), and Polokwane (29.89 %) were the only stations with an hourly $rRMSE$ less than 30 %. Ten other stations had hourly $rRMSE$ ranging from 30 % to 50 %. When considering hourly R^2 results, two stations, namely Upington (0.985) and Prieska (0.955) had $R^2 > 0.95$, three stations had $0.9 < R^2 < 0.95$. Seven stations had $R^2 < 0.9$. The poor correlation of hourly MERRA2 reanalysis data was also demonstrated by scatterplots in Figures 6.9 to 6.21, given in the Supplementary Materials, where the data points were not elongated along the 1:1 line and were mostly above the 1:1 line, indicating overestimation.

From 6.2-6.4 and Table 6.2, the hourly metrics for MERRA2 varied as given below:

- $-1\% \leq \text{hourly } rMBE \leq 23\%$;
- $16\% \leq \text{hourly } rRMSE \leq 50\%$;
- $9\% \leq \text{hourly } rMAE \leq 32\%$; and
- $0.823 \leq \text{hourly } R^2 \leq 0.985$.

Figures 6.5–6.8, which shows the aggregated hourly averages, demonstrates that all the gridded datasets could capture the temporal variability of GHI in different sites nonetheless, with varying accuracy. The MERRA2 reanalysis dataset overestimated GHI and it did not perform well as its line diverged from the reference measured dataset as well as the other datasets, in almost all the stations. ERA5 reanalysis also overestimated, and it diverged from observation and satellite datasets lines in most stations. SOLCAST, CMSAF, and CAMS satellite-based datasets lines were adjacent to the observation in most stations, which showed a good performance.

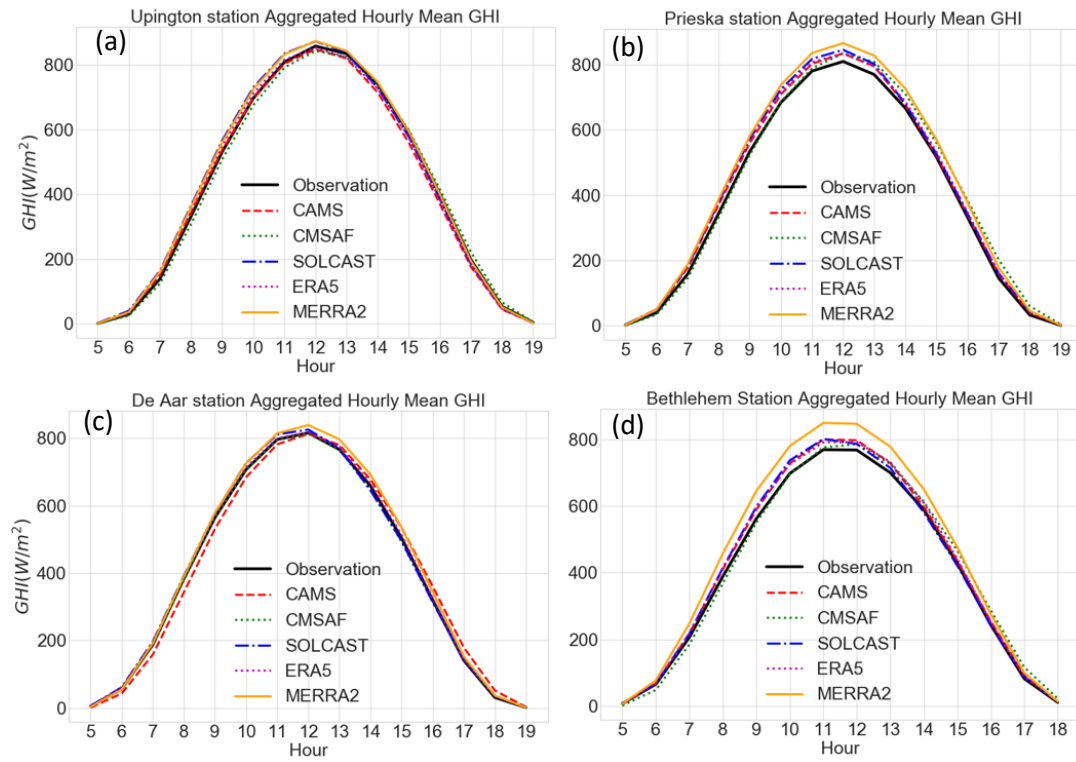


Figure 6.5: Aggregated measured and estimated hourly GHI values in De Aar (a), Bethlehem (b), Prieska (c), and Uppington (d). The aggregated GHI Observation period for each station is given in Table 3.2.

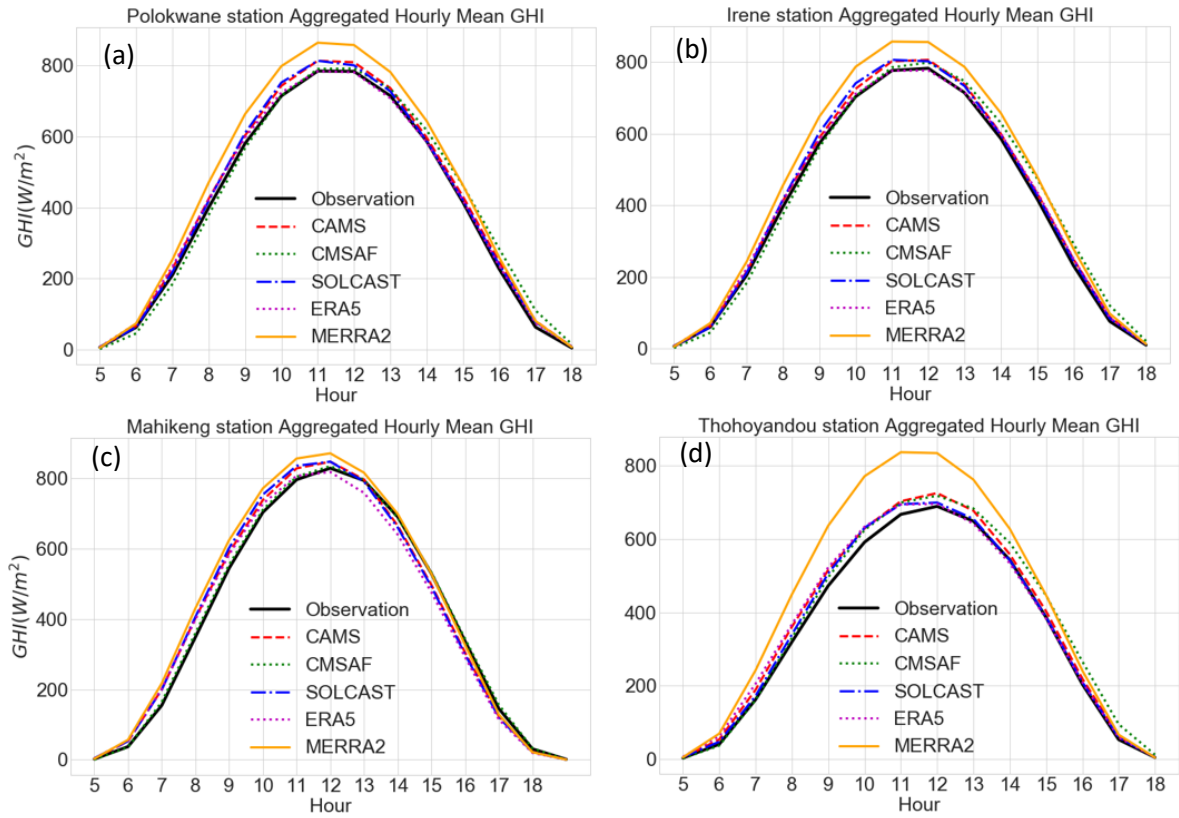


Figure 6.6: Aggregated measured and estimated hourly GHI values in Polokwane (a), Irene (b), Mahikeng (c), and Thohoyandou (d). The aggregated GHI Observation period for each station is given in Table 3.2.

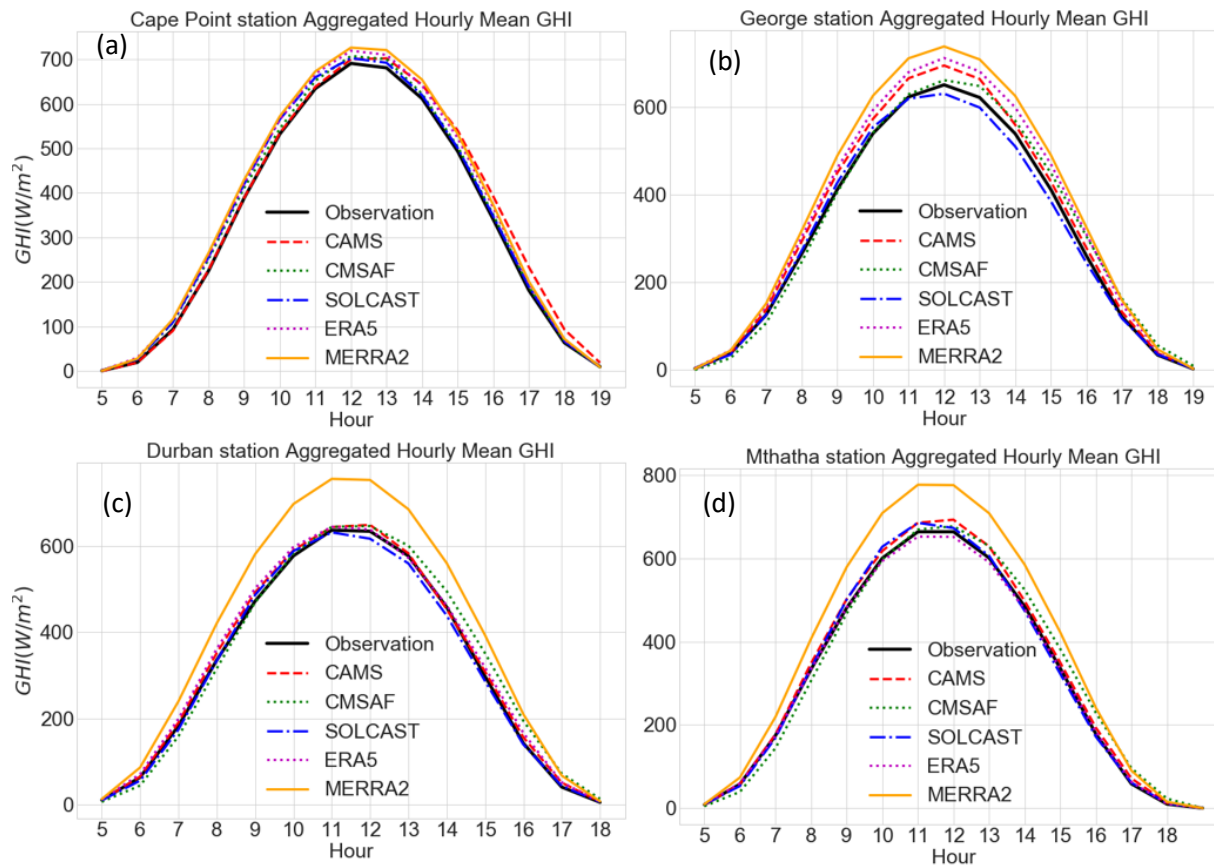


Figure 6.7. Aggregated measured and estimated hourly GHI values in values in Cape Point (a), George (b), Durban (c), and Mthatha (d). The aggregated GHI Observation period for each station is given in Table 3.2.

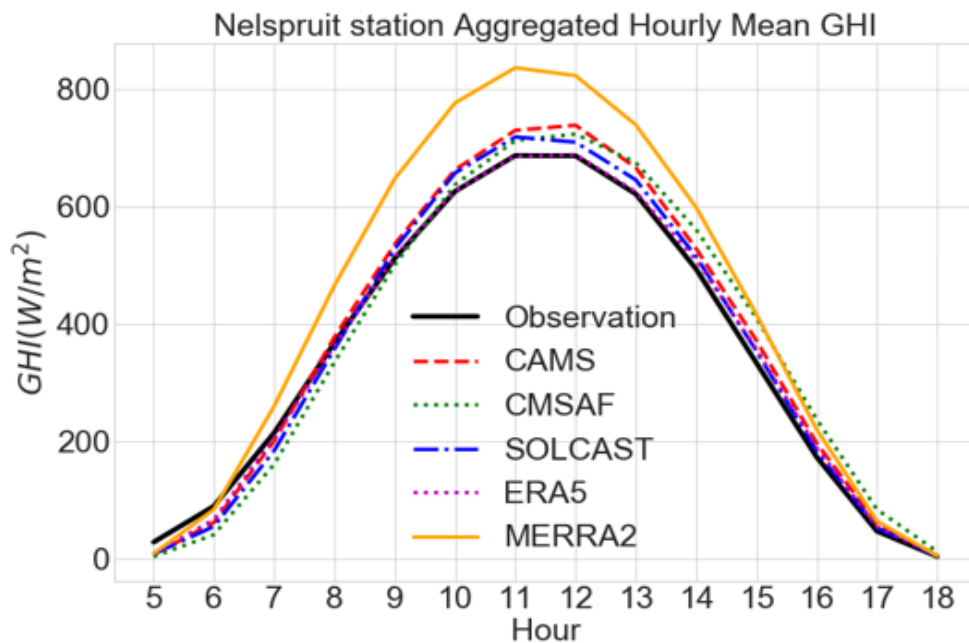


Figure 6.8 Aggregated measured and estimated hourly GHI values in values in Nelspruit. The aggregated GHI Observation period for each station is given in Table 3.2.

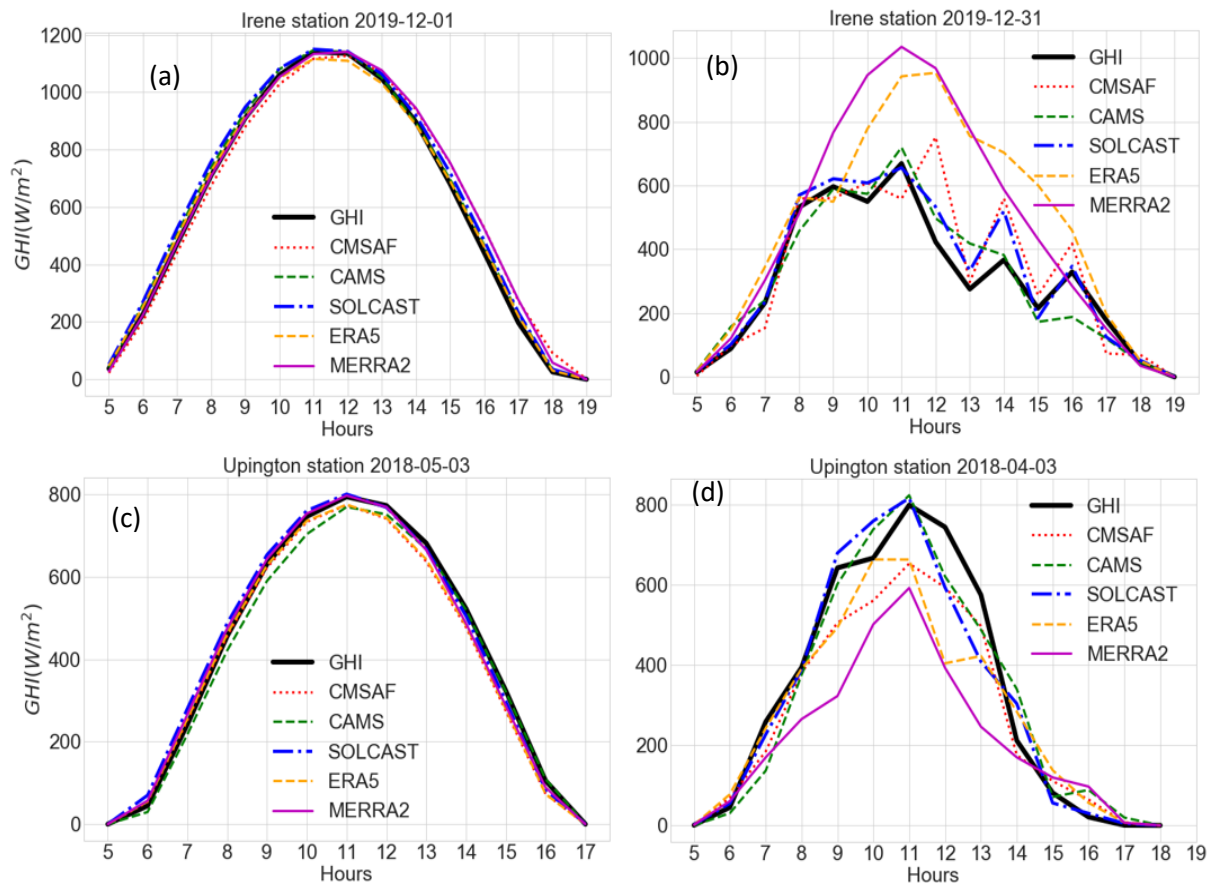


Figure 6.9 : Assessment of the hourly satellite-based and reanalysis-based datasets under clear sky conditions [(a)Irene on 2019-12-01 and (c) Uppington on 2018-05-03]) and cloudy sky conditions [(b) Irene on 2019-12-31 and (d) Uppington on 2018-04-03].

From Figure 6.9 the reanalysis-based datasets (ERA5 and MERRA2) struggle to estimates solar radiation under cloudy conditions 6.9 (a) and (c) but under clear sky conditions 6.9 (b) and (d) it estimates solar radiation accurately. Satellites based datasets (CAMS, CMSAF and SOLCAST) can estimates GHI well even under cloudy conditions. The overall poor performance of reanalysis-based datasets might be due to the struggle in accurately estimating GHI under cloudy conditions.

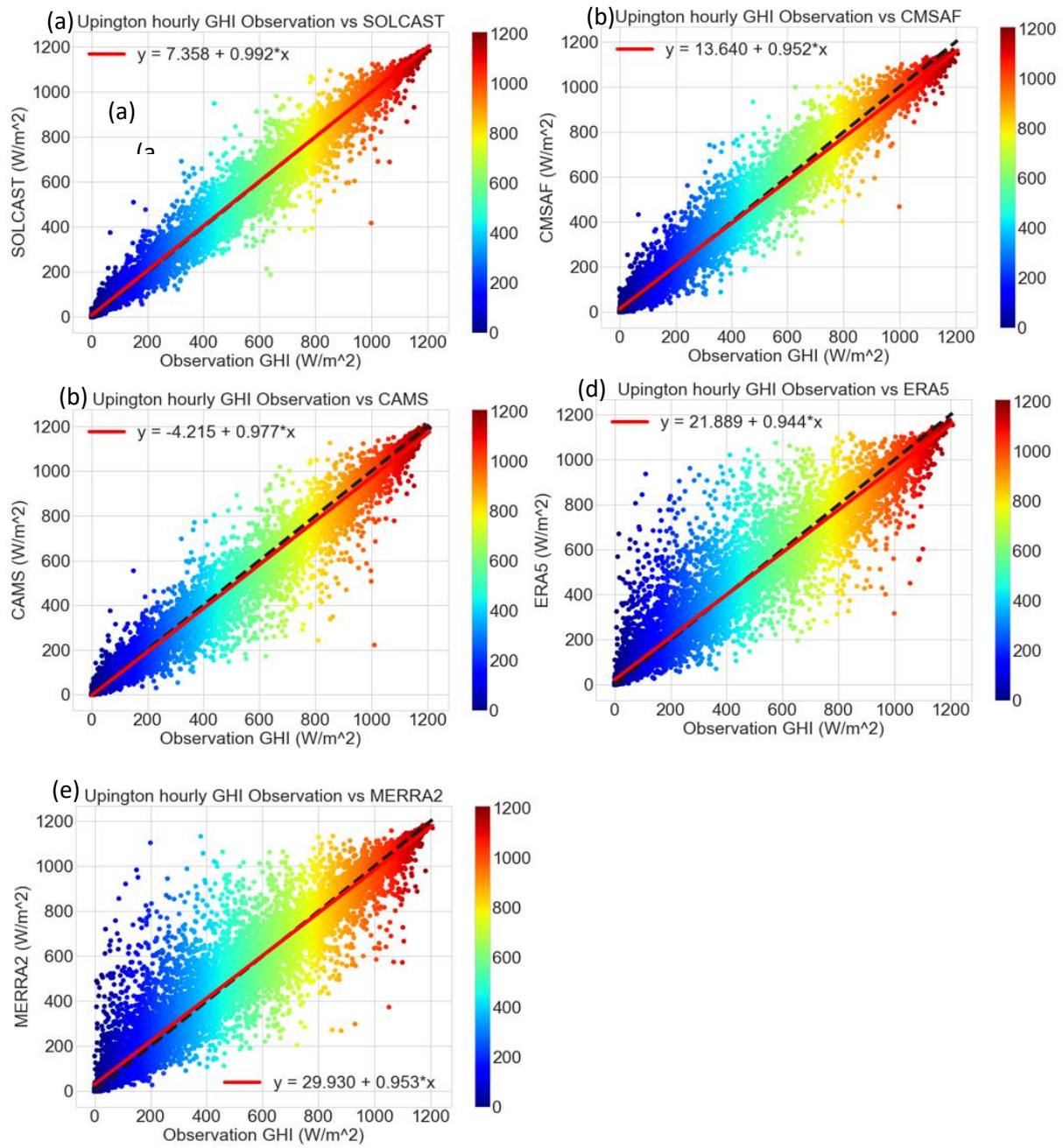


Figure 6.9. Scatter plot of measured and estimated daily global irradiance at Uppington station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

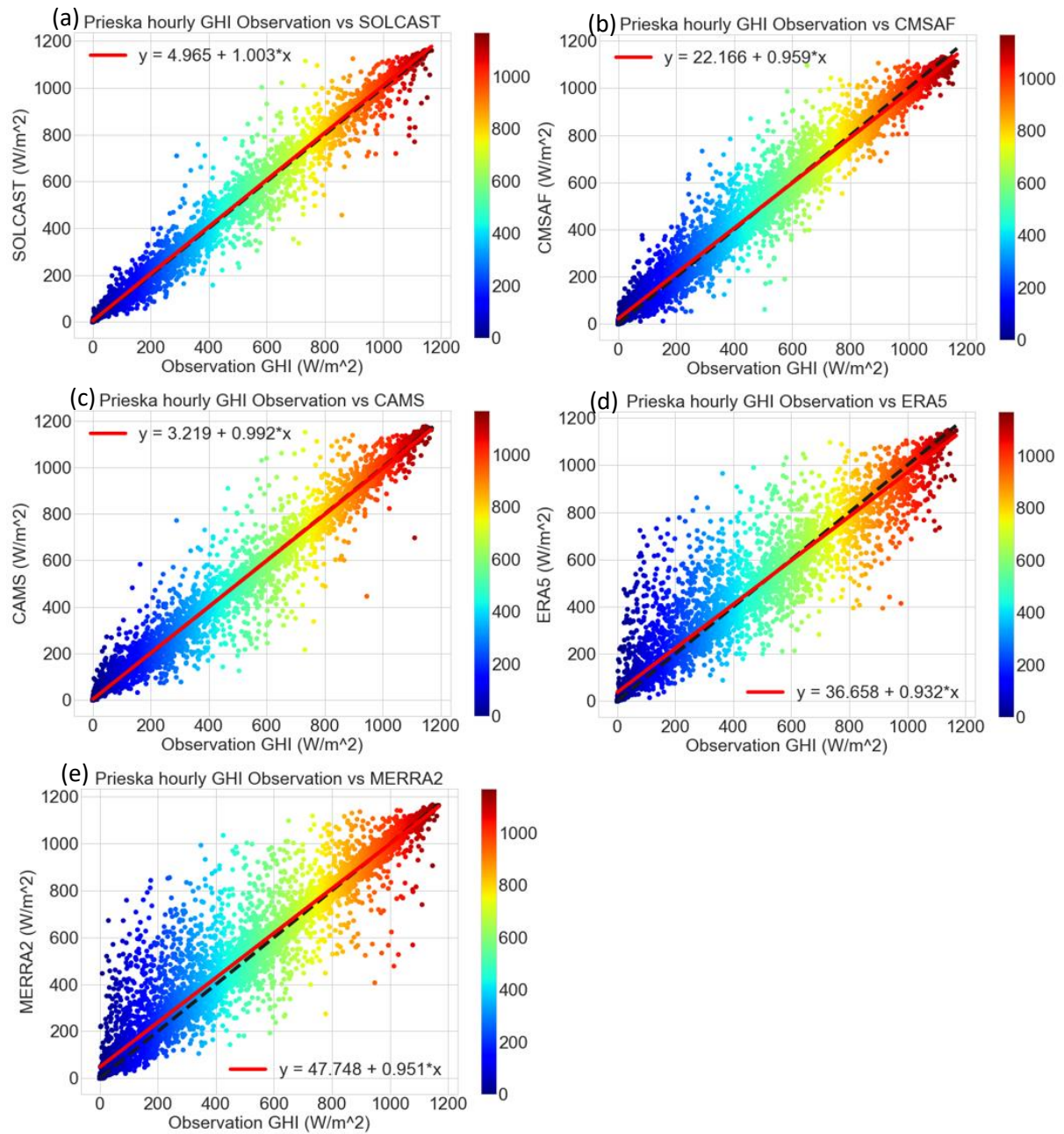


Figure 6.10. Scatter plot of measured and estimated daily global irradiance at Prieska station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

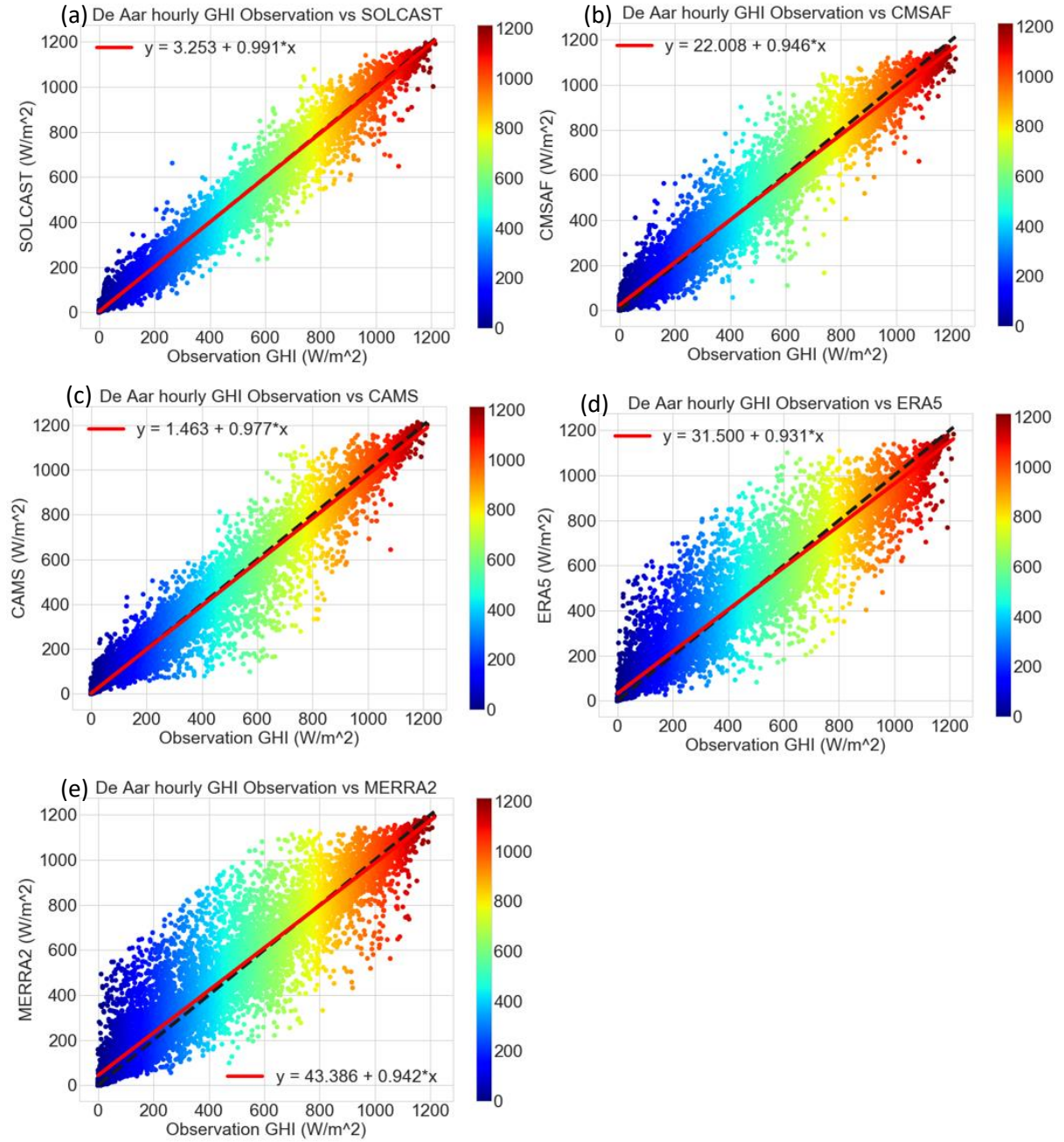


Figure 6.11 Scatter plot of measured and estimated daily global irradiance at De Aar station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

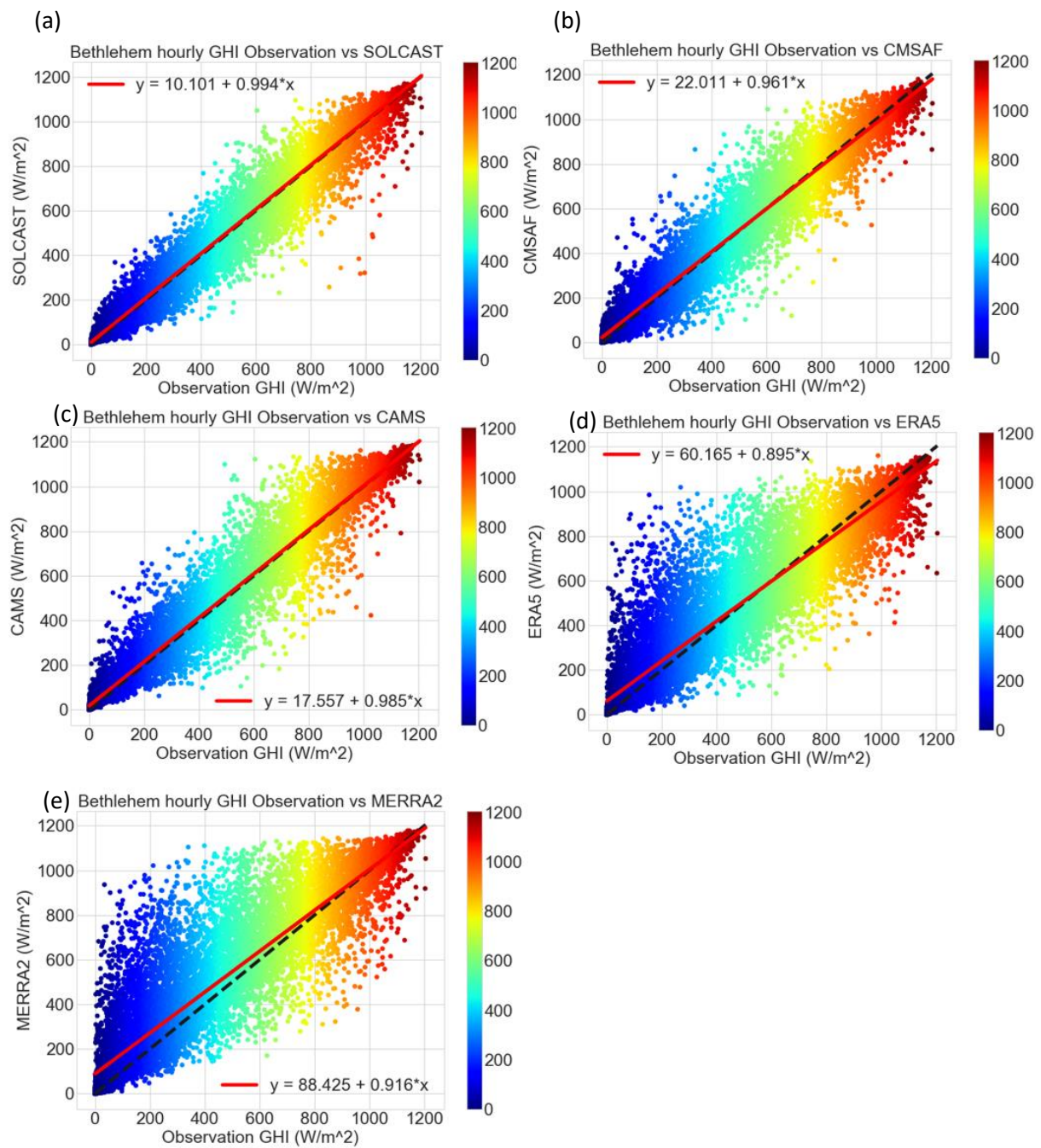


Figure 6.12. Scatter plot of measured and estimated daily global irradiance at Bethlehem station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

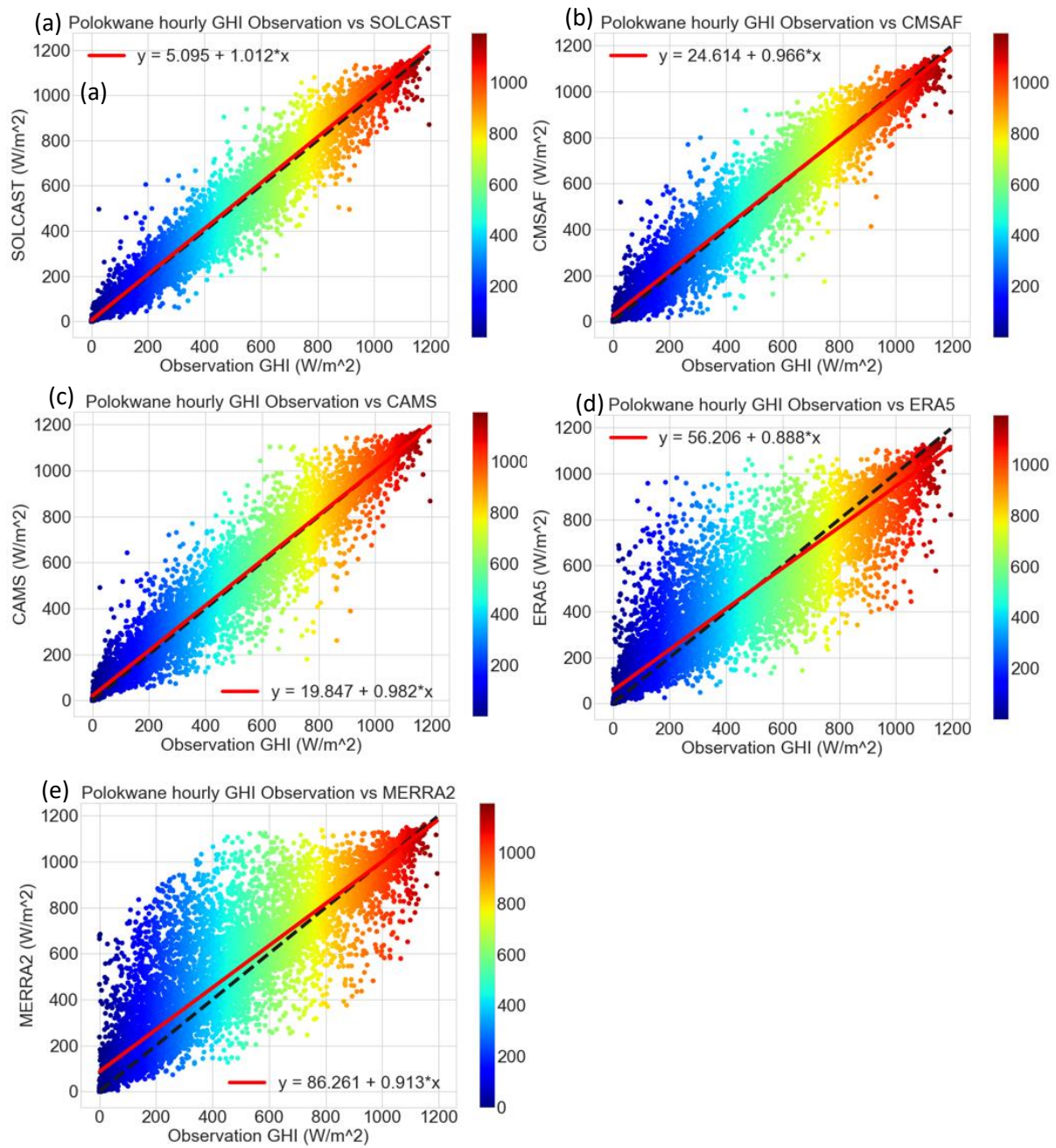


Figure 6.13: Scatter plot of measured and estimated daily global irradiance at Polokwane station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

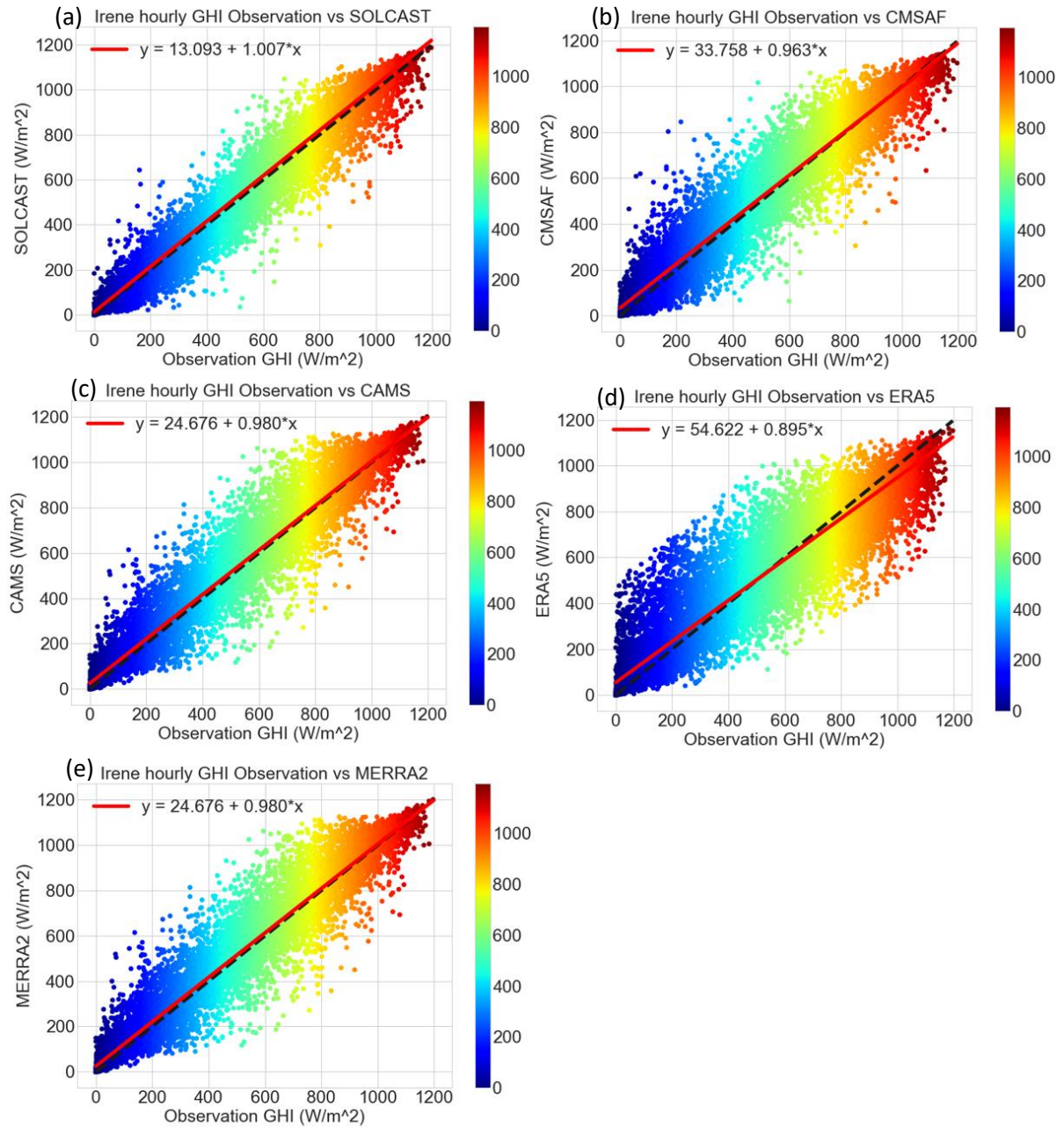


Figure 6.14 Scatter plot of measured and estimated daily global irradiance at Irene station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

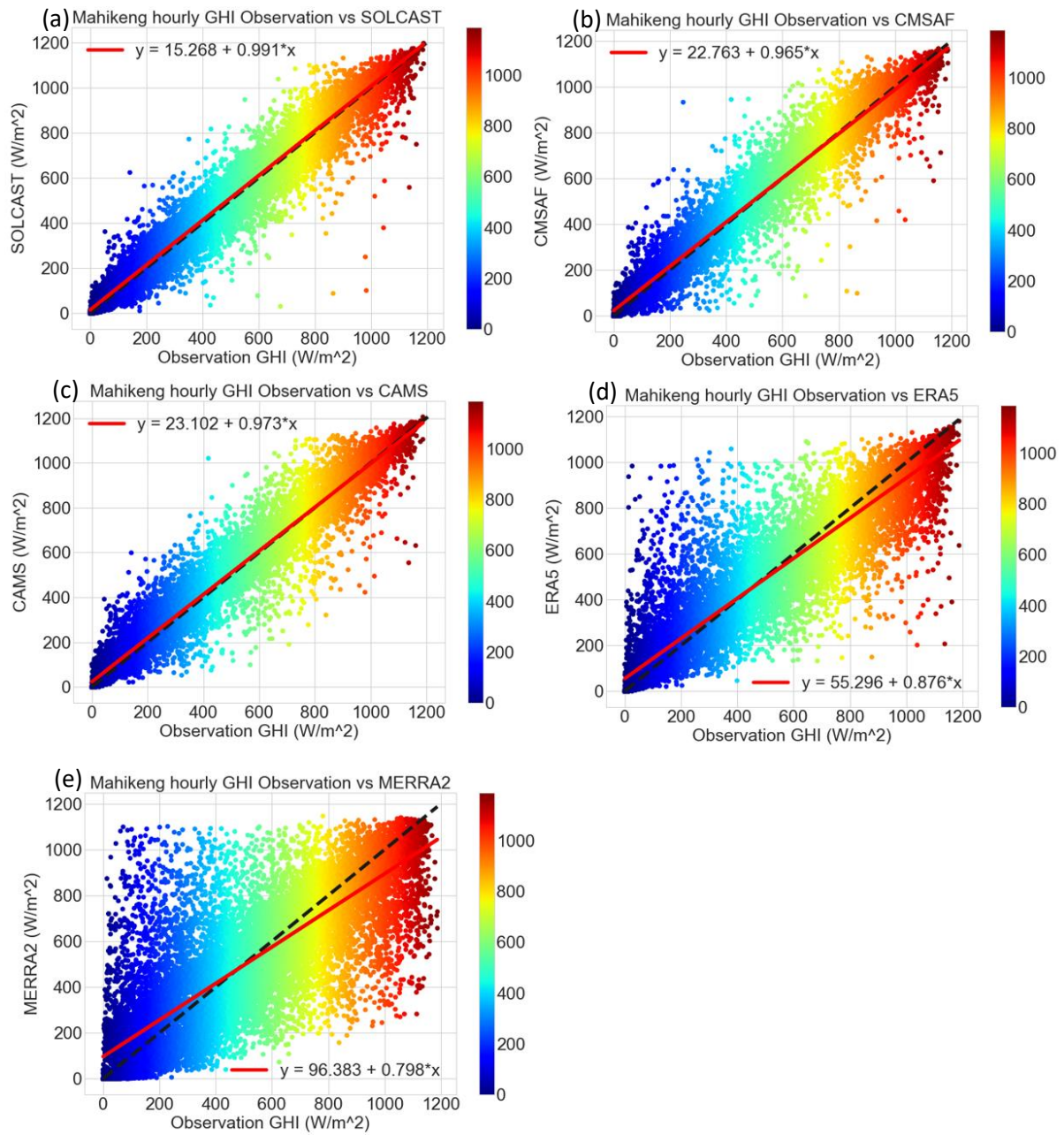


Figure S6.15 Scatter plot of measured and estimated daily global irradiance at Mahikeng station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

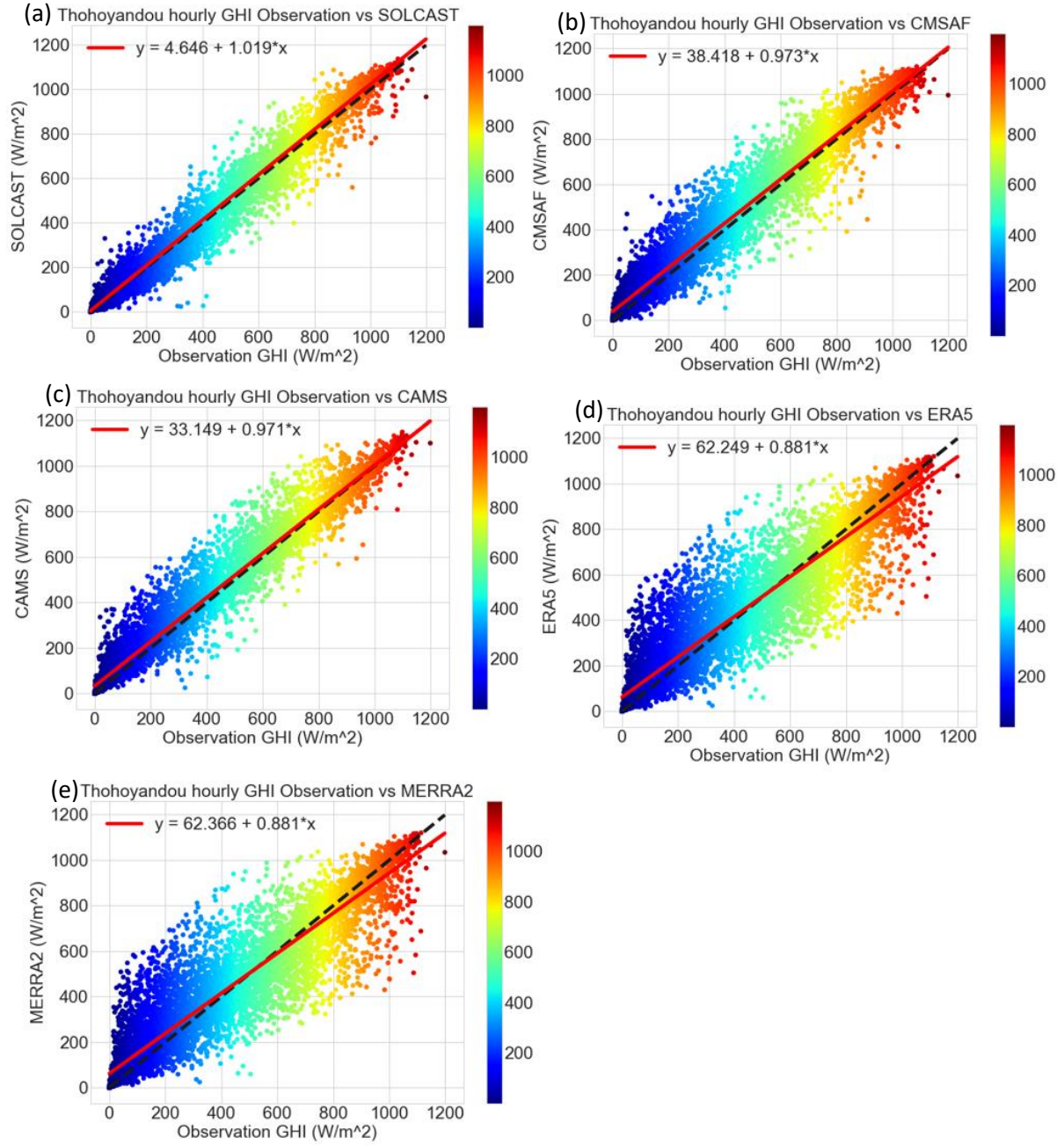


Figure S6.16 Scatter plot of measured and estimated hourly global irradiance at Thohoyandou station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

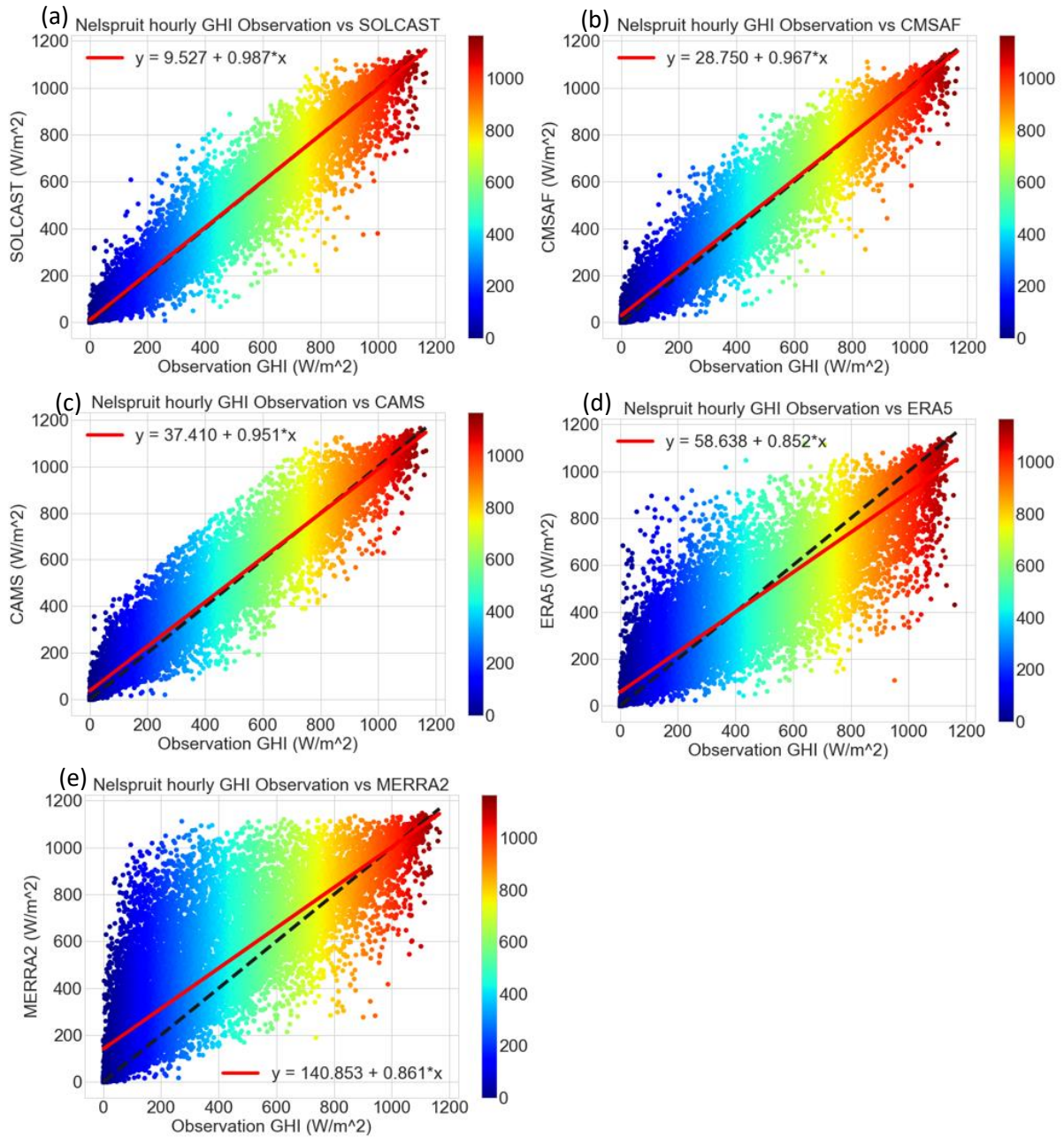


Figure S6.17 Scatter plot of measured and estimated hourly global irradiance at Thohoyandou station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

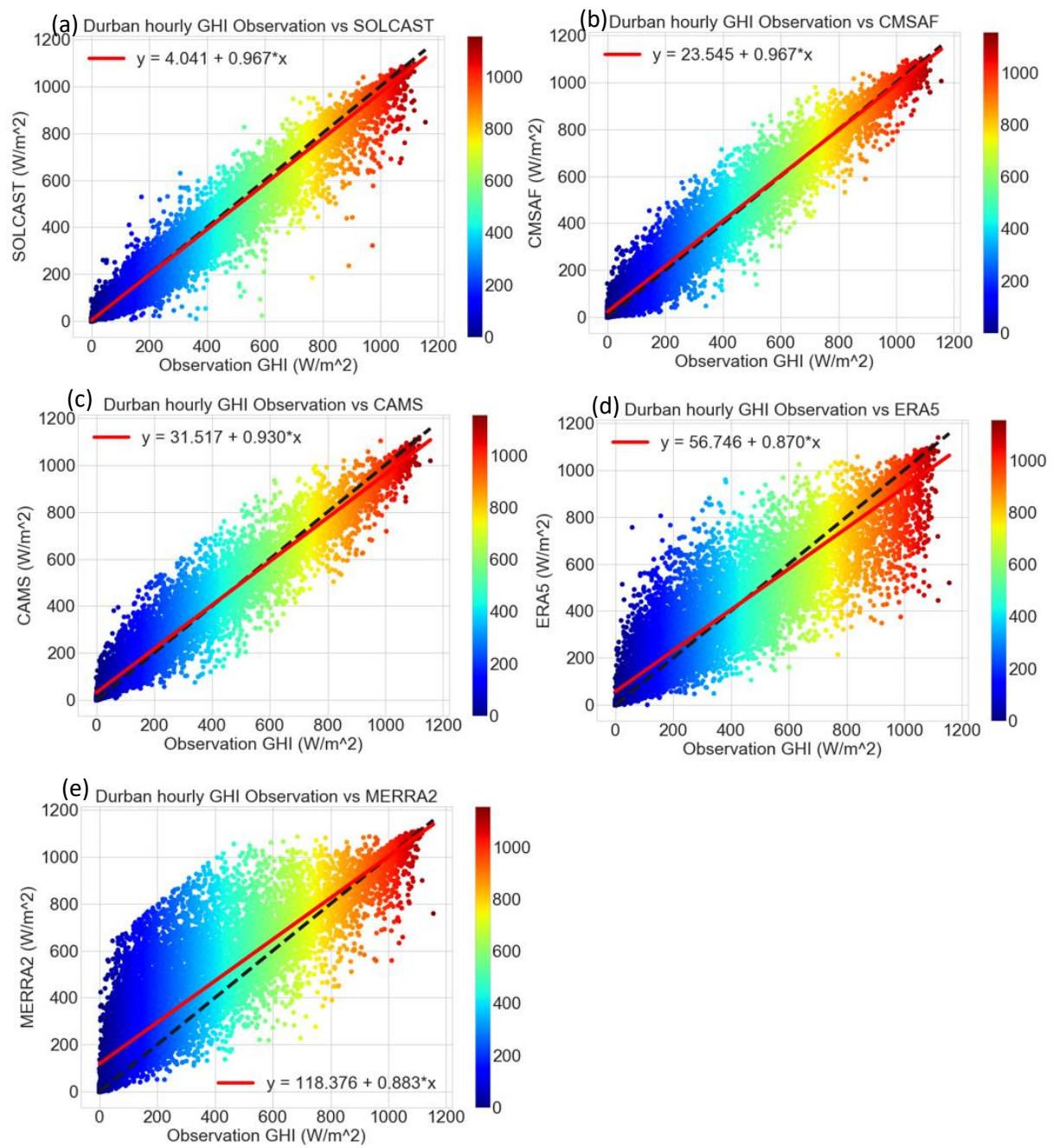


Figure S6.18 Scatter plot of measured and estimated hourly global irradiance at Durban station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

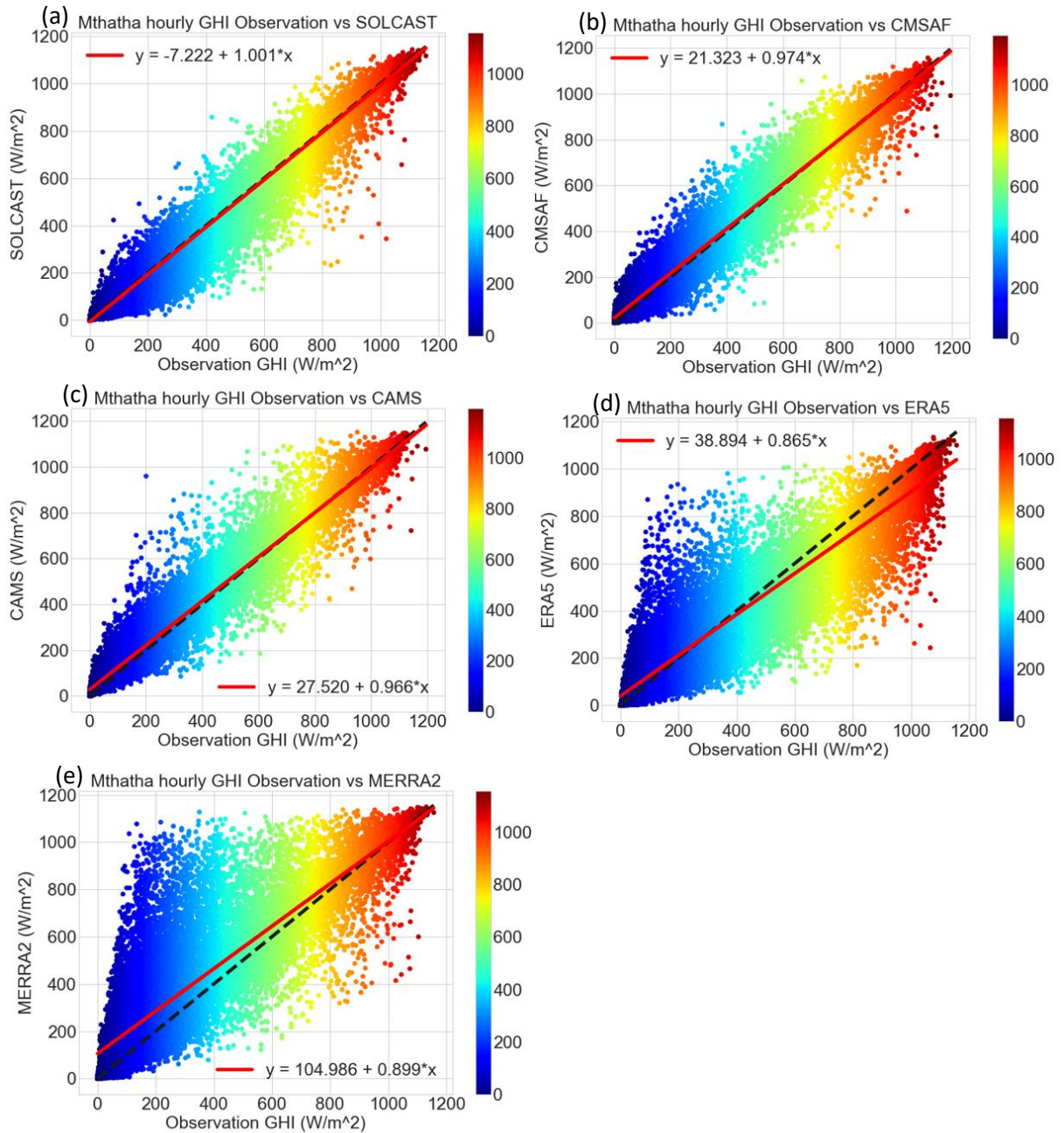


Figure S6.19 Scatter plot of measured and estimated hourly global irradiance at Mthatha station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

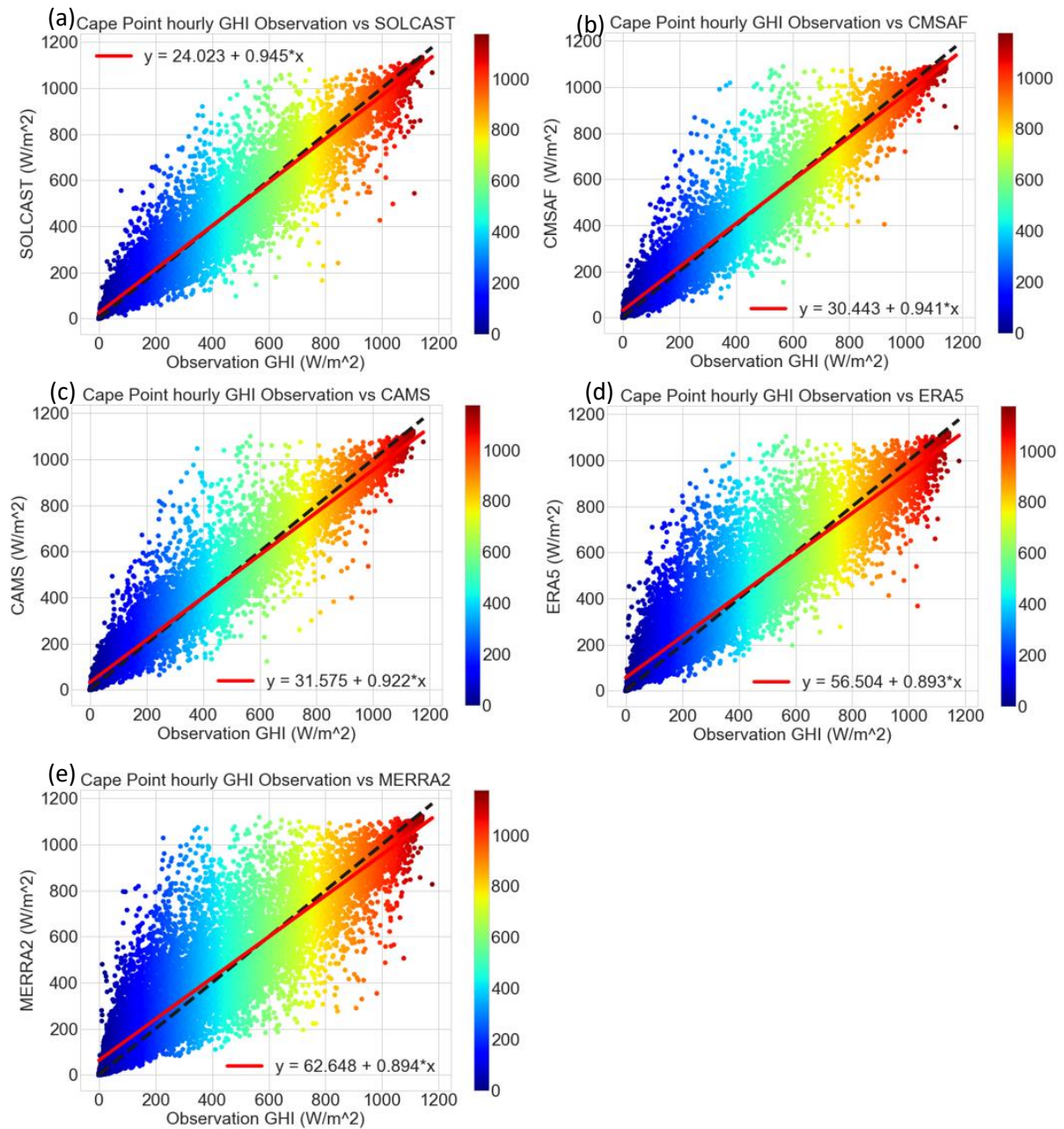


Figure S6.20 Scatter plot of measured and estimated hourly global irradiance at Cape Point station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

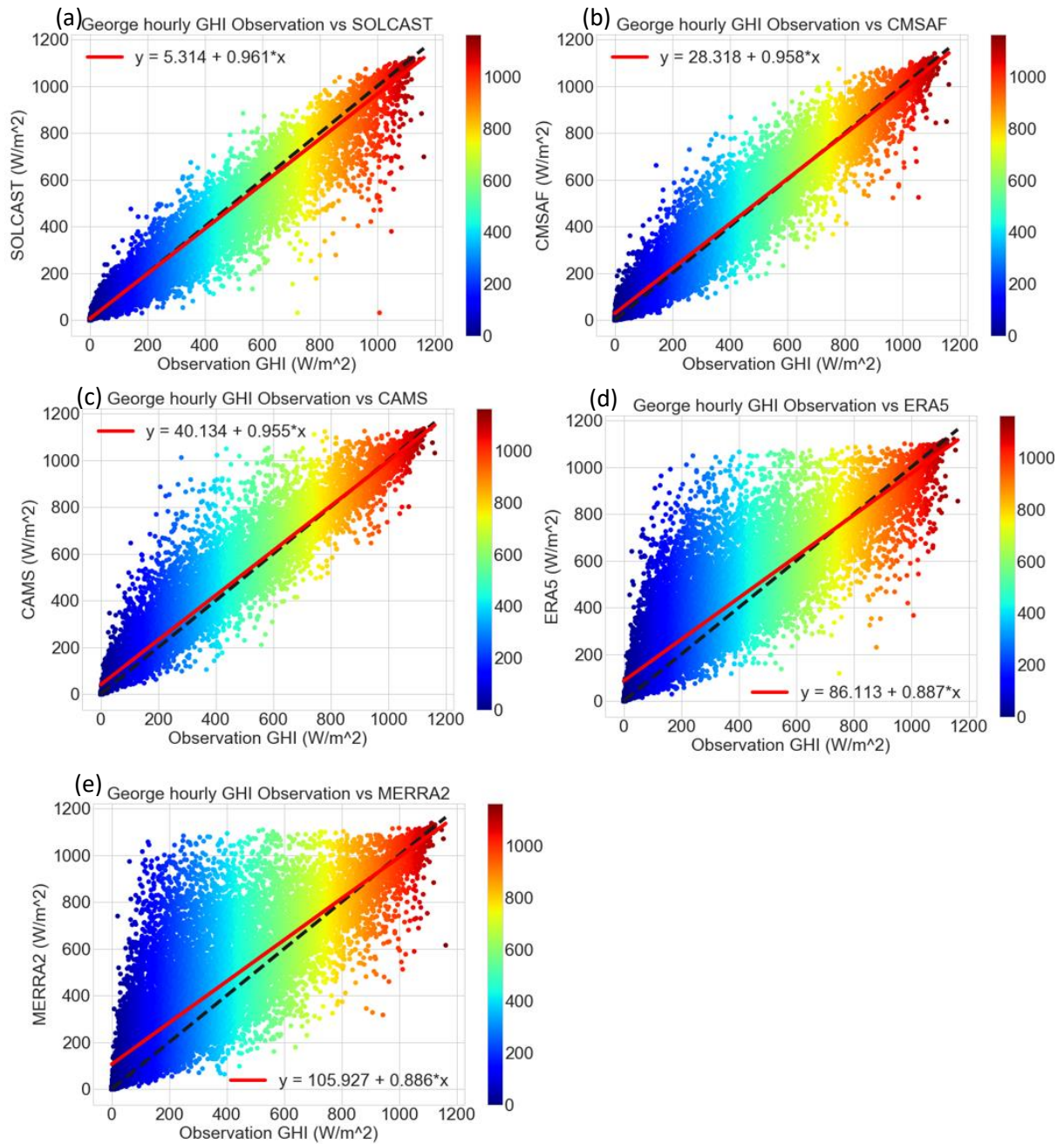


Figure 6.21. Scatter plot of measured and estimated hourly global irradiance at George station, SOLCAST (a), CMSAF (b), CAMS (c), ERA5 (d) and MERRA2 (e).

6.6. Discussion

6.6.1. CAMS

The above statistical metrics suggest that the CAMS dataset has good performance in South Africa. The findings are like the studies by Marchand et al. (2018) [16] in Morocco; Ameen et al. (2018) [4] in Iraq; Yang and Bright [12] for 57 BSRN stations; Thomas et al. [17] in Brazil; and Trolliet et al. [18] in the tropical Atlantic Ocean, which found that CAMS could accurately estimate the hourly GHI and can therefore be used with quantitative confidence as a reliable alternative source of estimated GHI data. The relatively high rRMSE observed for Nelspruit, Thohoyandou, Mthatha, Durban, Cape Point, and George may be due to the characteristic high annual humidity (greater than 60 % in Table 3.2; Table 3.3). The same tendency was also observed by Thomas et al. [17]. George and Nelspruit stations also had a high diffuse fraction ($DF > 0.32$). This implies that the performance of CAMS is affected by high aerosols, high humidity, and hence, many days with diffuse skies. When considering areas where there is infrequent cloud occurrence, there was no significance difference in bias, meaning that McClear can accurately estimate clear sky conditions in the study area. The previous study (Chapter 4) by Mabasa et al. [19] showed that the McClear clear sky model had a good performance in South Africa. For CAMS, cloud properties were derived from Meteosat satellites with 15-minute temporal resolution.

For CAMS, the main inputs to Heliosat-4 are aerosol properties, total column water vapor, and ozone content, as provided by the CAMS global services every three hours. The lower temporal resolution in aerosol, total column vapor, and ozone content may account for the observed biases between ground truth and CAMS estimated GHI. The overall good performance of CAMS datasets might be attributed to having a high spatial resolution 3–5 km in Southern Africa.

6.6.2. CMSAF

The overall results show that there is relatively good performance by the CMSAF satellite-based dataset, which suggests that CMSAF is a viable tool to estimate GHI for sites such as the 13 stations in this study. CMSAF satellite-based dataset showed a relatively poor performance at Nelspruit Station. Nelspruit is a station with the highest diffuse fraction (Table 3.2 and 3.3). CMSAF satellite-based dataset uses aerosol climatology as input to satellite retrieval algorithms as given by Riihelä et al. [20], however, aerosol climatology might not capture the aerosol climate variability. Mueller et al. [21] showed that aerosol climatologies used in CMSAF satellite retrievals algorithms were underestimated when compared to real aerosol measurements. The poorer metrics for Nelspruit might be due to the use of aerosol climatology information in the CMSAF satellite retrieval algorithm. No significant bias has been observed for the stations with high humidity and more frequent cloud occurrence compared to areas with low humidity and less frequent cloud occurrence. This means that the CMSAF cloud and water vapor parameterization scheme is effective in South Africa.

The CMSAF satellite-based dataset outperformed CAMS, ERA5, and MERRA2 at all 13 stations under this study (Table 6.3). CMSAF was outperformed by SOLCAST in 11 stations, this might be due to the use of hourly average from only two intervals (half hour and hour) instead of four intervals (15 minutes, 30 minutes, 45 minutes, and hour). The good performance of CMSAF datasets might also be attributed to having a high spatial resolution 5 km.

6.6.3. SOLCAST

These results show an overall very good performance of SOLCAST satellite-based dataset in all 13 stations in this study. The results agreed with the study by Yang and Bright [12] and Bright [13], who found that the SOLCAST satellite-based dataset performed well. Overall SOLCAST was outperformed by CMSAF (from Table 6.3) at Mahikeng and Cape Point

Stations; this was similar to the findings by Yang and Bright [12], where SOLCAST did not outperform some freely available products at all sites. The stations where rRMSE was greater than 15 % when referencing from Table 3.2; Table 3.3. They all had a low number of clear sky days (less than 5 %), high humidity (greater than 60 %), and high diffuse fraction (greater than 0.33), except Nelspruit Station with (more than 10 %) number of clear sky days. This implies that frequent cloud occurrence, higher humidity, and higher diffuse fraction slightly affected the performance of the SOLCAST satellite-based dataset.

The excellent performance of the SOLCAST dataset might be due to the use of the REST2v5 [14] clear sky model to calculate the clear sky index when converting the satellite image to GHI. Sun et al. [22] found that the REST2v5 clear sky model had an excellent worldwide performance. The use of very high spatial resolution satellite images 1–2 km enabled almost all features (e.g., terrain difference) in an area of interest or a grid to be properly identified and properly interpolated.

6.6.4. ERA5

Overall, the ERA5 reanalysis dataset showed a poor performance in estimating GHI in South Africa. The results were similar to Yang and Bright [12] and Trolliet et al. [18], who found that ERA5 datasets had poor performance, overestimated GHI for most sites, and were outperformed by the satellite-based dataset; rMBE was 72 % (Table 1.1). ERA5 reanalysis data showed a very poor performance in areas with frequent cloud occurrences, high humidity, and high diffuse fraction. From in Figures 6.9 to 6.21, ERA5 estimates tended to estimate cloud conditions while observations were showing actual conditions as clear conditions were shown by irradiance values, this might be contributing to the higher biases. Basically, ERA5 cloud models struggle to differentiate non cloud and cloud conditions. ERA5 uses climatological aerosol information [2] instead of measured aerosols which captures changes in atmospheric

constituents. This might also be one of the reasons for poor performance. The low spatial resolution ($0.25^\circ \times 0.25^\circ$) of the ERA5 reanalysis data might also be a contributing factor to the poor performance of the ERA5 datasets in South Africa.

6.6.5. MERRA2

Overall, MERRA2 reanalysis showed a very poor performance and it overestimated hourly GHI in all 13 sites under study. The performance of MERRA2 in South Africa was like the findings by Yang and Bright [12] for 57 BSRN stations and Trollet et al. [18] in the tropical Atlantic Ocean (i.e., poor performance, overestimating GHI, and being outperformed by satellite-based datasets; rMBE was 76 %) (Table 1.1). MERRA2 data showed a very poor performance in areas with frequent cloud occurrences, high humidity, and high diffuse fraction. The very low spatial resolution ($0.625^\circ \times 0.5^\circ$) of MERRA2 reanalysis data might also be a contributing factor to the poor performance of MERRA2. MERRA2 was the overall worst performing dataset in the study.

6.7. Chapter Summary

In chapter 6, Hourly global horizontal irradiance (GHI) from three satellite-based GHI datasets (namely SOLCAST, CAMS, and CMSAF SARA) and two reanalysis-based GHI datasets (namely ERA5 and MERRA2) were validated against quality-controlled hourly in-situ GHI recorded at 13 radiometric stations in South Africa. The study demonstrated that GHI from the satellite-based datasets had better performance than reanalysis-based datasets in South Africa. The overall statistical metrics used to gauge the performance of the datasets varied, as tabulated below (Table 6.4).

Table 6.4. Summary of the overall validation results. Colors are used to show the overall performance ranking of the five datasets with green (1/5), yellow (2/5), blue (3/5), orange (4/5), and red (5/5).

Dataset	rMBE	rRMSE	rMAE	R ²
SOLCAST	−3 % to 4 %	8 % to 19%	5 % to 12 %	0.969 to 0.996
CAMS	−3 % to 6 %	10 % to 21 %	6 % to 13 %	0.962 to 0.995
CMSAF	−2 % to 7 %	10 % to 33 %	7 % to 19 %	0.908 to 0.995
ERA5	−4 % to 11 %	15 % to 38 %	9 % to 25 %	0.868 to 0.987
MERRA2	−1 % to 23 %	16 % to 50 %	9 % to 32 %	0.823 to 0.985

SOLCAST was the best performing overall, while MERRA2 was the overall worst performing dataset. Freely available satellite-based datasets (CAMS and CMSAF) are recommended for use with quantitative confidence in diverse solar energy applications that require GHI data. Reanalysis based GHI datasets (ERA5 and MERRA2) are not good enough to be used in South Africa. Low spatial resolution, weak cloud parameterization schemes, and the use of climatological inputs instead of real in situ measurement in reanalysis GHI deriving algorithms might be some of the reasons behind the poor performance of reanalysis based GHI estimates in the study.

Based on the distribution of the results in study, the threshold or the range of the statistical metrics that are used to benchmark the optimal and/or the best applicable hourly gridded products based on their performance when compared to measured hourly data in the study area are summarized in (Table 6.5).

Table 6.5. Range of the statistical metrics used to benchmark the applicability of the gridded datasets. Green represents excellent performance, blue good performance, and orange poor performance).

Skill	rMBE	rRMSE	rMAE	R ²
Poor	$rMBE > \pm 10 \%$	$rRMSE > 20\%$	$rMAE > 15\%$	$R^2 < 0.9$
Good	$ \pm 5 \% < rMBE \leq \pm 10 \%$	$10\% < rRMSE \leq 20\%$	$10\% < rMAE \leq 15\%$	$90 < R^2 < 0.95$
Excellent	$rMBE \leq \pm 5 \%$	$rRMSE \leq 10\%$	$rMAE \leq 10\%$	$R^2 > 0.95$

References

1. Peng, X.; She, J.; Zhang, S.; Tan, J.; Li, Y. Evaluation of multi-reanalysis solar radiation products using global surface observations. *Atmosphere* 2019, 10, 42. <https://doi.org/10.3390/atmos10020042>.
2. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 2020, 146, 1999–2049. <https://doi.org/10.1002/qj.3803>
3. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* 2017, 30, 5419–5454. <https://doi.org/10.1175/jcli-d-16-0758.1>.
4. Ameen, B.; Balzter, H.; Jarvis, C.; Wey, E.; Thomas, C.; Marchand, M. Validation of hourly global horizontal irradiance for two satellite-derived datasets in Northeast Iraq. *Remote Sens.* 2018, 10, 1651. <https://doi.org/10.3390/rs10101651>
5. Pfeifroth, U.; Kothe, S.; Trentmann, J.; Hollmann, R.; Fuchs, P.; Kaiser, J.; Werscheck, M. Surface Radiation Data Set—Heliosat (SARAH), 2.1 Ed; Satellite Application Facility on Climate Monitoring: Offenbach, Germany, 2019. https://doi.org/10.5676/EUM_SAF_CM/SARAH/V002_01
6. Pfeifroth, U.; Kothe, S.; Trentmann, J. EUMETSAT Satellite Application Facility on Climate Monitoring, Validation Report. Meteosat Solar Surface Sarah 2. Available online: https://www.cmsaf.eu/SharedDocs/Literatur/document/2016/saf_cm_dwd_val_meteosat_hel_2_1_pdf.pdf?__blob=publicationFile (accessed on 17 January 2020).
7. Copernicus Portal. Available online: <https://atmosphere.copernicus.eu/data> (accessed on 16 March 2020).

8. Solar Radiation Data (SoDa) Service. Available online: <http://solar.atmosphere.copernicus.eu/cams-radiation-service> (accessed on 02 March 2020).
9. Qu, Z.; Oumbe, A.; Blanc, P.; Espinar, B.; Gesell, G.; Gschwind, B.; Klüser, L.; Lefèvre, M.; Saboret, L.; Schroedter-Homscheidt, M.; et al. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat 4 method. *Meteorol. Z.* 2017, 26, 33–57. <https://doi.org/10.1127/metz/2016/0781>.
10. CAMS Validation. Available online: https://atmosphere.copernicus.eu/sites/default/files/custom-uploads/EQC-solar/CAMS72_2018SC2_D72.1.3.1-2021Q2_RAD_validation_report_SON2020_v1.pdf (accessed on 18 January 2021).
11. SOLCAST. Available online: <https://solcast.com> (accessed on 15 April 2020).
12. Yang, D.; Bright, J.M. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. *Sol. Energy* 2020, 210, 3–19. <https://doi.org/10.1016/j.solener.2020.04.016>.
13. Bright, J.M. Solcast: Validation of a satellite-derived solar irradiance dataset. *Sol. Energy* 2019, 189, 435–449. <https://doi.org/10.1016/j.solener.2019.07.086>.
14. Gueymard, C.A. REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation—Validation with a benchmark dataset. *Sol. Energy* **2008**, 82, 272–285. <https://doi.org/10.1016/j.solener.2007.04.008>
15. SOLCAST Historical Data Validation. Available online: <https://solcast.com/historical-and-tmy/validation-and-accuracy> (ac-cessed on 15 April 2020).
16. Marchand, M.; Ghennioui, A.; Wey, E.; Wald, L. Comparison of several satellite-derived databases of surface solar radiation against ground measurement in Morocco. *Adv. Sci. Res.* 2018, 15, 21–29. <https://doi.org/10.5194/asr-15-21-2018>
17. Thomas, C.; Wey, E.; Blanc, P.; Wald, L. Validation of three satellite-derived databases of surface solar radiation using measurements performed at 42 stations in Brazil. *Adv. Sci. Res.* 2016, 13, 81–86. <https://doi.org/10.5194/asr-13-81-2016>.
18. Trolliet, M.; Walawender, J.P.; Bourlès, B.; Boilley, A.; Trentmann, J.; Blanc, P.; Lefèvre, M.; Wald, L. Downwelling surface solar irradiance in the tropical Atlantic Ocean: A comparison of re-analyses and satellite-derived data sets to PIRATA measurements. *Ocean Sci.* 2018, 14, 1021–1056. <https://doi.org/10.5194/os-14-1021-2018>.
19. Mabasa, B.; Lysko, M.; Tazvinga, H.; Zwane, N.; Moloi, S. The performance assessment of six global horizontal irradiance clear sky models in six climatological regions in South Africa. *Energies* 2021, 14, 2583. <https://doi.org/10.3390/en14092583>.

20. Riihelä, A.; Kallio, V.; Devraj, S.; Sharma, A.; Lindfors, A.V. Validation of the SARA-E satellite-based surface solar radiation estimates over India. *Remote Sens.* 2018, 10, 392. <https://doi.org/10.3390/rs10030392>.
21. Mueller, R.; Behrendt, T.; Hammer, A.; Kemper, A. A new algorithm for the satellite-based retrieval of solar surface irradiance in spectral bands. *Remote Sens.* 2012, 4, 622–647. <https://doi.org/10.1016/j.rse.2011.11.016>.
22. Sun, X.; Bright, J.M.; Gueymard, C.A.; Acord, B.; Wang, P.; Engerer, N.A. Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis. *Renew. Sustain. Energy Rev.* 2019, 111, 550–570. <https://doi.org/10.1016/j.rser.2019.04.006>.

Chapter 7: Comparison between satellite-based datasets and Ångström–Prescott (AP) model in estimating daily GHI estimates in South African locations

7.1. Introduction

This chapter aims to answers a problem statement:

1. Which is the overall best between the satellite-based datasets and Ångström–Prescott (AP) model in estimating daily GHI estimates in South African locations? What are the benchmarks for daily (rMBE, rRMSE, rMAE and R^2) in area of study?

The problem statement was addressed by setting the following objectives:

- Compare daily satellite based global horizontal irradiance (GHI) datasets estimates and Ångström–Prescott (AP) model estimates against daily measured in situ GHI to determine the overall best method in approximating daily GHI data for sites in South Africa.

Ångström–Prescott (AP) model (Chapter 5) showed that it is capable of accurately estimating daily GHI data. Satellite based data sets (Chapter 6) also showed that they can accurately estimate hourly GHI data. Reanalysis data sets ERA5 and MERRA2 were not included in this chapter because they did not perform well in approximating GHI in Chapter 6. NASA SSE was not included in Chapter 6 because they are only available as daily GHI averages. The chapter is aimed at comparing daily GHI estimated by Ångström–Prescott (AP) model and daily GHI estimated by satellites (NASA SSE, CAMS, CMSAF and SOLCAST) to determine the overall best between the satellite based and the Ångström–Prescott (AP) daily estimates.

7.2. Study Area

The area of study is eight sites located in all six macro-climate regions in South Africa, as described in Chapter 3 (Subsections 3.1 and 3.2). Four stations Bethlehem, Nelspruit, Mahikeng, Cape Point and Prieska were excluded in the study because there were no sunshine observation data to be used as an input to Ångström–Prescott (AP) model.

7.3. Data

Observation of daily GHI data as described in Chapter 3 (subsection 3.2), Ångström–Prescott (AP) model data described Chapter 5, satellite data (CMSAF, CAMS, SOLCAST) described in Chapter 6.

7.3.1. NASA SSE data

The National Aeronautics and Space Administration (NASA)’s Surface meteorology and Solar Energy (SSE), produces primary solar radiation data on a global spatial resolution of $1^{\circ} \times 1^{\circ}$ grid and then remap to a $0.5^{\circ} \times 0.5^{\circ}$ grid via bilinear interpolation. NASA SSE solar parameters are available in a daily average value, monthly averages, annual averages as well as climatological average values. NASA SSE was initiated to improve the current renewable energy data set and to create new data sets from new satellite systems. It was designed to provide access to parameters specifically tailored to assist in the design of solar and wind powered renewable energy systems. The data, from July 1983 to about 4 days of near real time, is processed, archived, and made available as daily averages [1].

7.4. Methodology

For the daily GHI averages, NASA SSE, CAMS, CMSAF SARA H were sourced as daily averages while for SOLCAST, ERA5 and MERRA2 daily averages were calculated from sourced hourly averages. The daily average GHI data was calculated from the hourly average. The daily average also included the night values i.e., it is the average calculated over 24 hours.

7.4.1. Statistical Metrics

As described in section 3.3.

7.4.2. Most Feasible Gridded Dataset

As described in 6.4.3.

7.5. Results

To allow for better comparison between daily satellite-based datasets results and the Ångström–Prescott (AP) model, the results from chapter 5 were converted from $\text{MJ m}^{-2}\text{d}^{-1}$ to W/m^2 .

7.5.1. CAMS

From Figure 7.1, CAMS underestimated GHI in Upington (−3 %) and De Aar (−2 %); and overestimated GHI at the remaining six stations, rMBE ranged from 2 %–6 %. From Figure 7.2, daily rRMSE was less than and slightly above 10 % for all stations. When considering the daily correlation R^2 results in Table 7.1, CAMS had $R^2 > 0.95$ in 7 stations, except Polokwane (0.944) suggesting that the daily measured and the estimated data correlated well. From Figure 7.1–7.3 and Table 7.1, the daily metrics for CAMS varied as given below:

- $-3\% \leq \text{daily rMBE} \leq 6\%$;
- $5\% \leq \text{daily rRMSE} \leq 12\%$;

- $3\% \leq \text{daily rMAE} \leq 8\%$ and
- $0.944 \leq \text{daily } R^2 \leq 0.990$.

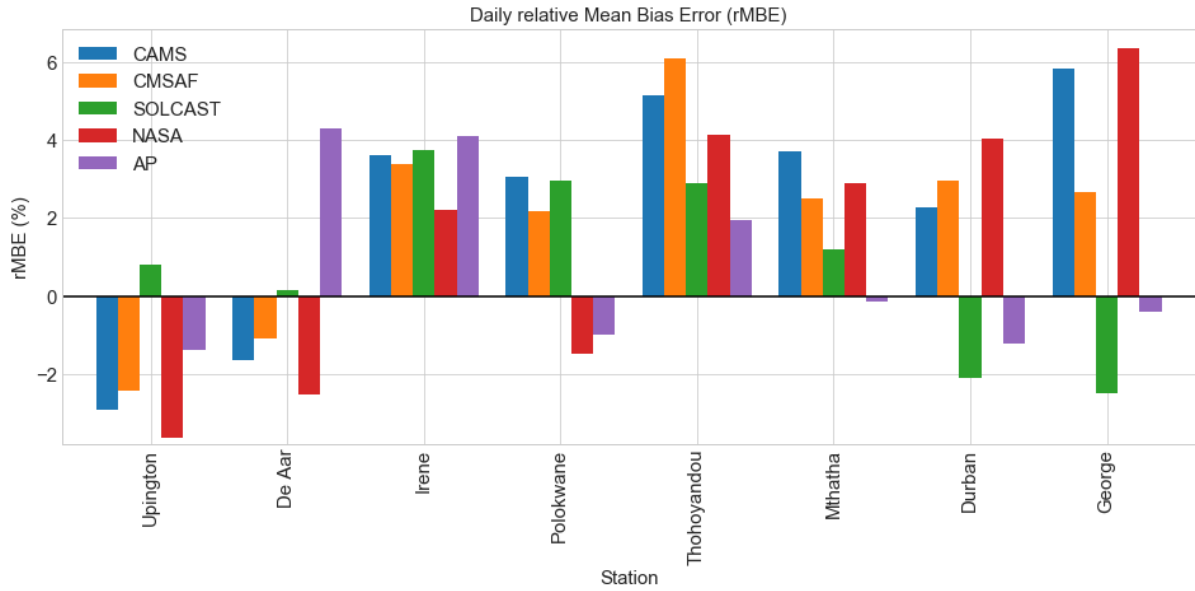


Figure 7.1: Daily relative mean bias error of gridded datasets against measured in situ GHI.

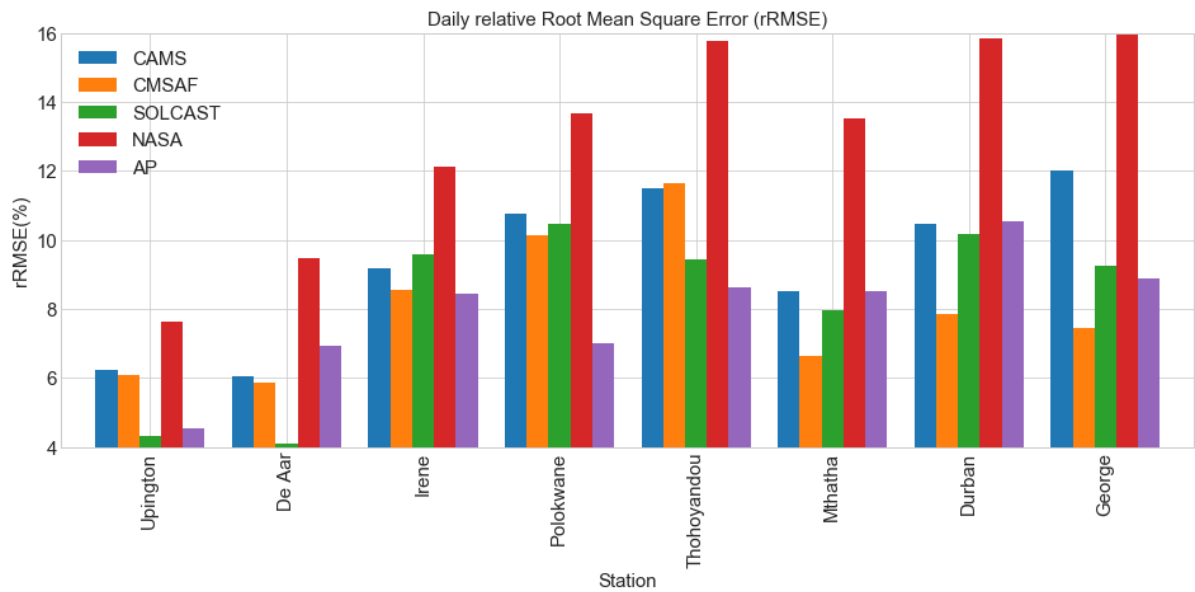


Figure 7.2: Daily relative root mean square error of gridded datasets against measured in situ GHI.

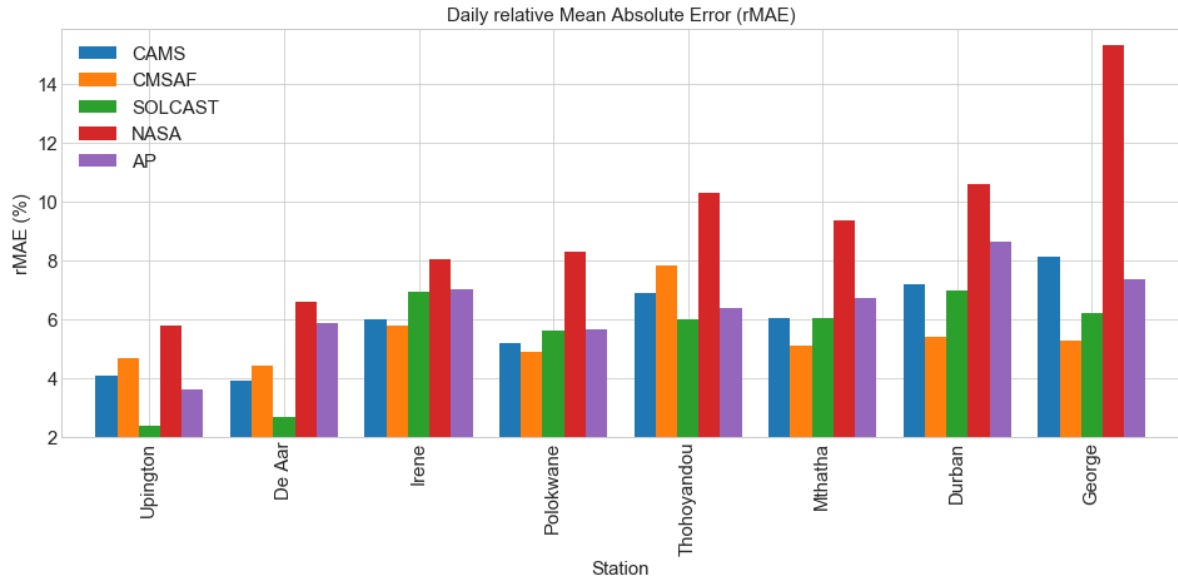


Figure 7.3: Daily relative mean absolute error of gridded datasets against measured in situ GHI.

Table 7.1 Daily mean measured GHI (W/m^2) and correlation (R^2) of gridded datasets against measured in situ GHI. Green color represents the best ($R^2 > 0.975$), blue color represents the intermediate ($0.925 \leq R^2 \leq 0.975$), and orange color represents the poor correlation ($R^2 < 0.925$).

Station	MEAN	CAMS	CMSAF	SOLCAST	NASA	Ångström–Prescott
Upington	260.81	0.985	0.987	0.991	0.979	0.930
DeAar	246.15	0.988	0.990	0.994	0.972	0.930
Irene	230.79	0.964	0.968	0.961	0.925	0.912
Polokwane	230.30	0.944	0.948	0.948	0.900	0.91
Thohoyandou	200.19	0.967	0.968	0.975	0.926	0.937
Mthatha	193.86	0.985	0.990	0.984	0.953	0.951
Durban	184.77	0.977	0.987	0.976	0.943	0.915
George	193.87	0.977	0.990	0.983	0.896	0.948

7.5.2. CMSAF

From Figure 7.1, CMSAF underestimated GHI in Upington (−2 %) and De Aar (−1 %) and overestimated GHI at the remaining six stations; rMBE ranged from 2 %–6 %. In Figure 4, daily rRMSE was less than and slightly above 10 % for all stations. From Table 7.1, when considering the daily correlation R^2 results, CMSAF had $R^2 > 0.96$ in 7 stations except in

Polokwane (0.948), which shows a very good correlation between the in-situ measured and estimated data. From Table 7.2, which gives the most feasible daily dataset per station by combining all the daily metrics, CMSAF was most prominent in 5 out of 8 stations for the daily aggregates. From Figure 7.1-7.3 and Table 7.1, the daily metrics for CMSAF varied as given below:

- $-2 \% \leq \text{daily rMBE} \leq 6 \%$;
- $5 \% \leq \text{daily rRMSE} \leq 12 \%$;
- $4 \% \leq \text{daily rMAE} \leq 8 \%$ and
- $0.948 \leq \text{daily } R^2 \leq 0.990$.

7.5.3. SOLCAST

From Figure 7.1, the SOLCAST dataset slightly underestimated daily GHI at George (-2%), and Durban (-2%) and it overestimated hourly GHI at the rest of the stations; rMBE ranged from 0.15% to 4% . From Table 7.1, SOLCAST had the correlation $R^2 > 0.96$ in 7 stations, except Polokwane (0.948) suggesting all 7 stations, showing a very good agreement between the measured and estimated data. From Table 7.2, which gives the most feasible daily dataset per station by combining all the daily metrics, SOLCAST was the most prominent in 3 out of 8 stations for the daily aggregates. From Figure 7.1-7.3 and Table 7.1, the daily metrics for SOLCAST varied as given below:

- $-2 \% \leq \text{daily rMBE} \leq 4 \%$;
- $4 \% \leq \text{daily rRMSE} \leq 10 \%$;
- $2 \% \leq \text{daily rMAE} \leq 7 \%$ and
- $0.948 \leq \text{daily } R^2 \leq 0.995$.

Table 7.2: Best performing gridded dataset per hourly metric, most feasible model, and level of performance (rating). The colors were used to show the best performing gridded dataset based on the hourly statical metric per station and the most feasible dataset per station-based level of rating out of 4 (the number of statistical metrics used). Green represents SOLCAST, yellow CAMS, blue CMSAF, orange AP, and red NASA.

Daily	Minimum rMBE	Minimum rRMSE	Minimum rMAE	Maximum R ²	Most Feasible	Rating
Upington	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
De Aar	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
Irene	NASA	AP	CMSAF	CMSAF	CMSAF	2/4
Polokwane	AP	AP	CMSAF	CMSAF	AP/CMSAF	2/4
Thohoyandou	AP	AP	SOLCAST	SOLCAST	AP/SOLCAST	2/4
Mthatha	AP	CMSAF	CMSAF	CMSAF	CMSAF	3/4
Durban	AP	CMSAF	CMSAF	CMSAF	CMSAF	3/4
George	AP	CMSAF	CMSAF	CMSAF	CMSAF	3/4

7.5.4. Ångström–Prescott (AP) model

From Figure 7.1, the AP model slightly underestimated daily GHI at Upington (-1.4 %), Polokwane (-1 %), George (-0.4 %), Mthatha (-0.15 %) and Durban (-1.2 %) and it overestimated hourly GHI at the rest De Aar (4 %), Irene (4 %) and Thohoyandou (2 %). In Figure 7.2, daily rRMSE was less than and slightly above 10 % for all stations. From Table 7.1, SOLCAST had the correlation $R^2 > 0.9$ in all the stations, showing a very good agreement between the measured and estimated data. From Table 7.3, which gives the most feasible daily dataset per station by combining all the daily metrics, the AP model was most prominent in 2 out of 8 stations for the daily aggregates. From Figure 7.1 to Figure 7.3 and Table 7.2, the daily metrics for AP model varied as given below:

- $-1 \% \leq \text{daily rMBE} \leq 4 \%$;
- $5 \% \leq \text{daily rRMSE} \leq 11 \%$;
- $4 \% \leq \text{daily rMAE} \leq 9 \%$ and
- $0.910 \leq \text{daily } R^2 \leq 0.951$.

7.5.5. NASA SSE

From Figure 7.1, the NASA SSE slightly underestimated daily GHI at Upington (-4 %), De Aar (-3 %) and Polokwane (-1.5 %) and it overestimated daily GHI at the remaining five stations, rMBE ranged from 2 %–6 %. In Figure 7.2, daily rRMSE was less than 10 % in Upington (8 %) and De Aar (10 %), it was greater than 10 % but less than 15 % in three stations and greater than 15 % in Thohoyandou (15.8 %), Durban (15.8 %) and George (22 %). From Table 7.2, NASA SSE had the correlation $R^2 > 0.9$ in 7 stations and less than 0.9 in George (0.896), showing a good agreement between the measured and estimated data. From Figure 7.1 to Figure 7.3 and Table 7.2, the daily metrics for NASA SSE varied as given below:

- $-4 \% \leq \text{daily rMBE} \leq 6 \%$;
- $8 \% \leq \text{daily rRMSE} \leq 23 \%$;
- $5 \% \leq \text{daily rMAE} \leq 15 \%$ and
- $0.896 \leq \text{daily } R^2 \leq 0.979$.

7.6. Chapter Summary

In Chapter 7, daily global horizontal irradiance (GHI) from four satellite-based GHI datasets (namely SOLCAST, CAMS, CMSAF SARA and NASA SSE) and the Ångström–Prescott (AP) model were validated against quality-controlled daily in situ GHI recorded at 8 radiometric stations in South Africa. The chapter further investigated the overall best method between the satellite-based and Ångström–Prescott (AP) model in estimating daily GHI. The chapter demonstrated that daily GHI estimates from the satellite-based and Ångström–Prescott (AP) model were comparable. The overall statistical metrics used to gauge the performance of the datasets varied, as tabulated below (Table 7.3).

Table 7.3 Summary of the overall validation results. Colors are used to show the overall performance ranking of the five datasets with green (1/5), yellow (2/5), blue (3/5), orange (4/5), and red (5/5).

Dataset	rMBE	rRMSE	rMAE	R ²
SOLCAST	−2% to 4%	4% to 10%	2% to 7%	0.948 to 0.995
CAMS	−3% to 6%	5% to 12%	3% to 8%	0.944 to 0.990
CMSAF	−2% to 6%	5% to 12%	4% to 8%	0.948 to 0.990
Ångström–Prescott (AP)	−1% to 4%	5% to 11%	4% to 9%	0.910 to 0.951
NASA SSE	−4% to 6%	8% to 23%	5% to 15%	0.896 to 0.979

SOLCAST was the best performing overall in estimating daily GHI, while NASA SSE was the overall worst performing dataset. CAMS, CMSAF and Ångström–Prescott (AP) model performed better, and their performance was almost the same in the study area. CAMS, CMSAF and Ångström–Prescott (AP) model are recommended for estimating daily GHI in South Africa with quantitative confidence in diverse solar energy applications that require GHI data. NASA SSE daily GHI dataset is not good enough in estimating daily GHI in South Africa. Low spatial resolution (0.5 ° x 0.5 °) might be one of the reasons behind its poor performance. AP model outperforms satellite based in some stations, but they can only be used in areas where there in measured sunshine duration data, satellites based can be derived at any South African location.

Based on the distribution of the results in the area of study, the threshold or the range of the statistical metrics that are used to benchmark the optimal and/or the best applicable daily gridded products based on their performance when compared to measured daily data in the study area are summarized in (Table 7.4).

Table 7.4 Range of the statistical metrics used to benchmark the applicability of the gridded datasets. Green represents excellent performance, blue good performance, and orange poor performance).

Skill	rMBE	rRMSE	rMAE	R ²
Poor	$rMBE > \pm 5 \%$	$rRMSE > 15\%$	$rMAE > 10\%$	$R^2 < 0.975$
Good	$ \pm 3 \% < rMBE \leq \pm 5 \%$	$5\% < rRMSE \leq 15\%$	$5\% < rMAE \leq 10\%$	$0.925 \leq R^2 \leq 0.975$
Excellent	$rMBE \leq \pm 3 \%$	$rRMSE \leq 5\%$	$rMAE \leq 5\%$	$R^2 > 0.925$

References

[1] NASA SSE. Available online: <https://power.larc.nasa.gov/data-access-viewer/> (accessed on 02 February 2020). These data were obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program.

Chapter 8. Conclusion

This chapter summarizes the solution or findings to each problem statement (s).

8.1. Problem Statement 1

Which clear sky model performs better in estimating minute global horizontal irradiance (GHI) in each location/station under study and in each macro-climatological region of South Africa? Which GHI clear sky model is better overall in estimating minute clear sky GHI in South Africa?

The research problem statement 1, was addressed by setting the following aims and objectives

8.1.1. Aims and Objectives 1

To Calibrate (using meteorological input parameters) and validate six global horizontal irradiance (GHI) clear sky models namely: Bird, Simple Solis, McClear, Ineichen–Perez, Haurwitz and Berger–Duffie against GHI data recorded by a reference pyranometer only on clear sky days, to determine the overall the best performing clear sky model for sites in South Africa.

8.1.2. Meeting the Aims and Objectives 1

The Aim and objectives 1 were meet by calibrating and validating six GHI clear sky models, namely: Bird, Simple Solis, McClear, Ineichen–Perez, Haurwitz and Berger–Duffie in Chapter 4. All the clear sky models assessed were capable of reproducing GHI irradiance, but with different accuracy in different stations. Complex clear sky models performed better than the simple clear sky models. Bird was the overall best model, followed by Simple Solis, McClear, Ineichen–Perez, Haurwitz and Berger–Duffie.

8.2. Problem Statement 2

Are the Ångström–Prescott (AP) coefficients which are used as inputs to the model that is used daily estimate GHI using daily sunshine duration (SD) and Top of the atmosphere (TOA) GHI as an input different from default values $a=0.25$ and $b=0.75$? What are the values of AP coefficients in each location/station under study and in each macro-climatological region of South Africa? How accurate are the calibrated AP coefficients in estimating daily GHI in each location/station under study and in each macro-climatological region of South Africa?

The research problem statement 2, was addressed by setting the following aims and objectives:

8.2.1. Aims and Objectives 2

Calibrate the Ångström–Prescott (AP) model regression coefficients a and b that could be used to estimate GHI in different climatic zones of South Africa, thus increasing the density of available solar radiation data in the country.

8.2.2. Meeting the Aims and Objectives 2

The Aim and objectives 2 were met by calibrating the annual Ångström–Prescott model coefficients a and b . AP model coefficients were calculated using the linear relationship between ratio of daily radiation on a horizontal surface to the daily extra-terrestrial radiation on that surface and the ratio of daily sunshine duration to the theoretical sunshine duration. The calculated Ångström–Prescott model coefficients were calculated for each station and then used to estimate global horizontal irradiance. There was a very close agreement with the corresponding observation global horizontal irradiance. The agreement was quantified by statistical metrics. This implies that the Ångström–Prescott coefficients calculated in this study could enable the estimation of daily global horizontal irradiance data at any location in South

Africa, where daily observation sunshine duration data are available, and the climate is correctly classified.

8.3. Problem Statement 3

Which is the best freely available (for research purposes) GHI satellite-based datasets and GHI reanalysis-based dataset in each location/station under study and overall, in South Africa? What are the benchmarks for hourly (rMBE, rRMSE, rMAE and R^2) in study area based on the distribution of the results from the study area ?

The research problem statement 3, was addressed by setting the following aims and objectives:

8.3.1 Aims and Objectives 3

Validate and rank (based on the performance) the hourly satellite-based global horizontal irradiance (GHI) datasets namely: SOLCAST, Copernicus Atmosphere Monitoring Service (CAMS), and Satellite Application Facility on Climate Monitoring (CMSAF SARA) and the two reanalysis-based, namely, Fifth generation European Center for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) and Modern-Era Retrospective Analysis for Research and Applications (MERRA2) for sites in South Africa.

8.3.2 Meeting the Aims and Objectives 3

In chapter 6, Hourly global horizontal irradiance (GHI) from three satellite-based GHI datasets (namely SOLCAST, CAMS, and CMSAF SARA) and two reanalysis-based GHI datasets (namely ERA5 and MERRA2) were validated against quality-controlled hourly in-situ GHI recorded at 13 radiometric stations in South Africa. The study demonstrated that GHI from the satellite-based datasets had better performance than reanalysis-based datasets in South Africa.

SOLCAST was the best performing overall, while MERRA2 was the overall worst performing dataset. Freely available satellite-based datasets (CAMS and CMSAF) are recommended for use with quantitative confidence in diverse solar energy applications that require GHI data. Reanalysis based GHI datasets (ERA5 and MERRA2) are not good enough to be used in South Africa.

8.4. Problem Statement 4

Which is the overall best between the satellite-based datasets and Ångström–Prescott (AP) model in estimating daily GHI estimates in South African locations? What are the benchmarks for daily (rMBE, rRMSE, rMAE and R^2) in study area based on the distribution of the results from the study area ?

The research problem statement 4, was addressed by setting the following aims and objectives:

8.4.1. Aims and Objectives 4

Compare daily satellite based global horizontal irradiance (GHI) datasets estimates and Ångström–Prescott (AP) model estimates against daily measured in situ GHI to determine the overall best method in estimating daily GHI data for sites in South Africa. Determine the interval metrics based on the distribution of the daily results

8.4.2. Meeting the Aims and Objectives 4

In Chapter 7, daily global horizontal irradiance (GHI) from four satellite-based GHI datasets (namely SOLCAST, CAMS, CMSAF SARA and NASA SSE) and the Ångström–Prescott (AP) model were validated against quality-controlled daily in situ GHI recorded at 8 radiometric stations in South Africa. The chapter further investigated the overall best method between the satellite-based and Ångström–Prescott (AP) model in estimating daily GHI. The

chapter demonstrated that daily GHI estimates from the satellite-based and Ångström–Prescott (AP) model were comparable.

SOLCAST was the best performing overall in estimating daily GHI, while NASA SSE was the overall worst performing dataset. CAMS, CMSAF and Ångström–Prescott (AP) model performed better, and their performance was almost the same in the study area. CAMS, CMSAF and Ångström–Prescott (AP) model are recommended for estimating daily GHI in South Africa with quantitative confidence in diverse solar energy applications that require GHI data. NASA SSE daily GHI dataset is not good enough in estimating daily GHI in South Africa.

8.5. Study limitations

- The study was limited but not having in-situ sunshine duration data in all thirteen stations under study. Sunshine duration data was only available at nine stations, it was not available at four stations, namely Bethlehem, Nelspruit, Mahikeng, Cape Point and Prieska. The stations where sunshine duration data was not available were excluded in the study.
- The challenge encountered in the study was the absence of high spatio-temporal resolution model inputs. Low spatio-temporal resolution model input parameters were used, even though some studies showed that using low spatio-temporal resolution model inputs does not have significant effect, there might be some uncertainties introduced which might increase the biases between observed and estimated datasets. It is recommended that GHI monitoring stations also measure high temporal and high spatial resolution water vapour, aerosols, ozone and albedo to aid model assessment studies.

- Absence of in-situ cloud data led the author to use hourly ERA5 reanalysis cloud information.
- Data gaps in the reference GHI data due to operational problems and further data removal (outliers) after quality checks, reduced the sample size of the study but the study period was sufficient to make thorough investigation of the performance of the models, reanalysis and satellite-based data sets.

8.6. Future studies

- One of the study limitations was the absence of in-situ data in four stations. To enable representation of all the study sites, satellite-based sunshine duration data will be validated against in-situ data.
- Comparison between daily satellite-based GHI estimates and AP model estimates with coefficients calculated and validated from satellite-based sunshine duration data will be investigated.
- Further research will focus on extended monitoring of the stability of the AP model coefficients over time in each climate zone.
- Future studies or activities will focus on close monitoring and maintaining of the existing radiometric network, to minimise data gaps and reduce the outliers in the observed data. Additional of sensors to measure high spatio-temporal resolution water vapour, aerosols, ozone and albedo to aid model assessment studies in existing radiometric network will be proposed to meteorological institutions.
- Calibration and validating of humidity-based, cloud-based and temperature-based models and compare the validation results to the sunshine-based models. Calibration and validation of other models will enable solar radiation data to be estimated even in areas with no in-situ sunshine duration data.

- Validation of models, satellites and reanalysis-derived datasets that can estimate diffuse horizontal irradiance and direct normal irradiance will be done in the future studies.
- Applying machine learning techniques to remove the bias between estimated datasets and observed datasets (validated in the study) will form part of future studies.
- Applying machine learning techniques to forecast solar radiation.
- Solar radiation climatological studies and creating high resolution GHI maps will be done after bias removal techniques have been applied on the validated datasets.
- Future studies will focus on optimising or calibrating simple clear sky models' coefficients for each South African site under study.

APPENDIX

List of Acronyms and Abbreviations

Abbreviation	Full Description	Units
AP	Ångström–Prescott model	
GHI	Global Horizontal Irradiance	W/m ²
DNI	Direct Normal Irradiance	W/m ²
DIF	Diffuse Horizontal Irradiance	W/m ²
CMSAF	Satellite Application Facility on Climate Monitoring	-
SARAH	Surface Solar Radiation Data Record—Heliosat Edition 2	-
ERA5	Fifth generation European Center for Medium-Range Weather Forecasts atmospheric reanalysis	-
MERRA2	Modern-Era Retrospective Analysis for Research and Applications	-
SAWS	South African Weather Services	
MBE	Mean Bias Error	W/m ²

rMBE	relative Mean Bias Error	Percentage (%)
MAE	Mean Absolute Error	W/m ²
rMAE	relative Mean Absolute Error	Percentage (%)
RMSE	Root Mean Square Error	W/m ²
rRMSE	relative Root Mean Square Error	Percentage (%)
R ²	Coefficient of Determination	-
ESRA	European Solar Radiation Atlas	-
BSRN	Baseline Surface Radiation Network	-
QC	Quality Check	-
K_T	Clearness index threshold,	-
ECMWF	European Centre for Medium-Range Weather Forecasts	-
GHI_{TOA}	Extraterrestrial (top of the atmosphere) GHI radiation	W/m ²
a and b .	Ångström–Prescott model coefficients	-
GPI	Global Performance Indicator	-
PIRATA	Prediction and Research Moored Array in the Tropical Atlantic	-
SD	Sunshine Duration	Hours
TOA	Top of the Atmosphere	W/m ²
I_{sc}	Solar constant	W/m ²
WRC	World Radiation Centre	-
WMO	World Meteorological Organization	-
DNI_{TOA}	Direct normal extra-terrestrial irradiance	W/m ²
DNI'_{TOA}	Enhanced Top of the atmosphere Direct Normal Irradiance	W/m ²
R	Instantaneous distance of the Earth from the Sun	Kilometres (km)
D	D is the Julian day	
θ_z	Solar zenith angle	Degrees
UV	Ultraviolet radiation	-
AM	Relative Air Mass	-
AM_a	Absolute Air Mass	-
SH	Southern Hemisphere	-
δ	Solar declination angle	Degrees
α_s	Solar altitude	Degrees
γ_s	Solar azimuth	Degrees
ϕ	Degrees of latitude	Degrees
E_o	Eccentricity factor	-
ω_s	Hour angle	Degrees
ST	Solar Time	Minutes
STD	Local Standard Time	Minutes
E	Equation of time	Minutes
N_d	Daylength	Hours
n	Daily measured sunshine duration	Hours
U_{emf}	Output of a pyranometer or pyrliometer	Microvolt (μV)
S	Sensitivity or calibration factor of the pyrliometer or pyranometer's thermopile.	$\mu V / W/m^2$.
ISO	International Standard	-
ACR	Absolute Cavity Radiometer	-

WRR	World Radiometric Reference	-
WRC	World Radiation Centre	-
PV	Photovoltaic	-
CSIR	Council for Scientific and Industrial Research	-
KGCC	Köppen-Geiger climate classification	-
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites	-
AVHRR	Advanced Very High-Resolution Radiometer	-
NOAA	National Oceanic and Atmospheric Administration	-
SoDA	Solar Radiation Data	-
GOME	Global Ozone Monitoring Experiment	-
GEOCART	Global Ozone Chemistry Aerosol Radiation and Transport	-
APPOLLO	AVHRR Processing scheme Over cLOUDs, Land and Ocean	-
MODIS	Moderate Resolution Imaging Spectroradiometer	-
NASA	The National Aeronautics and Space Administration	-
SSE	Surface meteorology and Solar Energy	-
AWS	Automatic Weather Station	-
SPA	Solar Position Algorithm	-
SAST	South African Standard Time	-
a_1, a_2	Kasten model's altitude coefficients	
f_{h1}, f_{h2}	Coefficients that relate the site altitude with the altitude of the atmospheric interactions	
h	Altitude of the site	Metres (m)
T_A	Transmittance after aerosol absorptance and scattering	
T_O	Transmittance after ozone absorptance in the stratosphere	
T_W	Transmittance after water vapor absorptance	
T_{UM}	Transmittance after absorptance by uniformly mixed gases	
T_R	Transmittance after Rayleigh scattering	
R_g	Ground albedo	
R_s	Sky albedo	
T_{AS}	Transmittance of aerosol scattering	
T_{AA}	Transmittance of aerosol absorbance	
B_A	Asymmetric factor or the forward scattering ratio	
τ	Total atmospheric optical depth	
b	Constant of adjustment for Direct Normal Irradiance	
τ_g	Global Horizontal Irradiance total optical depth	
g	Global Horizontal Irradiance fitting parameter	
AOD	Aerosol optical depth	
AOD_{380}	Aerosol optical depth at 380 nanometres	
AOD_{500}	Aerosol optical depth at 500 nanometres	
AOD_{550}	Aerosol optical depth at 550 nanometres	
AOD_{700}	Aerosol optical depth at 700 nanometres	
AOD_{1240}	Aerosol optical depth at 1240 nanometres	
τ_λ	Optical thickness at unknown wavelength	
τ_{λ_0}	Optical thickness at known wavelength	
λ	Wavelength of the unknown aerosol optical depth	Angstrom
λ_0	Wavelength of the known aerosol optical depth	Angstrom

α	Angstrom of the aerosol	nanometres (nm)
P_w	Precipitable water or water vapour	Centimetres (cm)
H_v	Apparent water vapor scale height	Kilometres (km)
P_v	Surface water vapor density	$(\frac{g}{m^3})$.
T	Temperature	Degrees Celsius (°C)
RH	Relative Humidity	Percentage (%)
p	Surface pressure	Pascals (Pa)
TL	Linke Turbidity factor	
O_3	Atmospheric ozone	Dobson units

Python scripts

-Used to calculate solar geometry angle, model configuration, data analysis and plotting

Data Analysis Script

1 minute to 15-minute averages

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

obs_file = 'DAA2020_min.csv'
#columns = ['DATE', 'HOUR', 'MIN', 'GHI', 'TOA']
columns = ['DATE', 'HOUR', 'MIN', 'GHI', 'DIF', 'DNI', 'DHI', 'TOA']
obs = pd.read_csv(
    obs_file,          # path to your input obs file
    header = None,
    skiprows = 1,      # ignore first 5 lines (headers)
    names = columns,   #dtype='object', # name the columns
)
obs
```

```
ValidityDatetime = pd.to_datetime(
    obs['DATE'] + \
    obs['HOUR'].astype(int).astype(str).apply(lambda x: x.zfill(2)) + \
    obs['MIN'].astype(int).astype(str).apply(lambda x: x.zfill(2)),
    format='%Y-%m-%d%H%M'
)
obs['ValidityDatetime'] = ValidityDatetime
```

```

obs.index = obs['ValidityDatetime'] # setting obs table index to be the
ValidityDatetime columns
obs_15m = obs.resample('0.25H').mean()
obs_15m = obs_15m.reset_index(level=None)
obs_15m

obs_15m.to_csv('DAA2020_15min.csv')

```

15-minute average to 1 hour average

```

obs_file = 'DAA2020_15min.csv'

columns = ['DATE', 'HOUR', 'SZA', 'GHI', 'CAMS']
obs = pd.read_csv(
    obs_file,          # path to your input obs file
    header = None,
    skiprows = 1,      # ignore first 5 lines (headers)
    names = columns,   #dtype='object', # name the columns
)
obs

```

```

ValidityDatetime = pd.to_datetime(
    obs['DATE'] + \
    obs['HOUR'].astype(int).astype(str).apply(lambda x: x.zfill(2))
    ,
    format='%Y-%m-%d%H'
)
obs['ValidityDatetime'] = ValidityDatetime
obs.index = obs['ValidityDatetime'] # setting obs table index to be the
ValidityDatetime columns
obs_hour = obs.resample('H').mean()
obs_hour = obs_hour.reset_index(level=None)
obs_hour

```

```

obs_hour.to_csv('DAA2020_hour.csv')

```

Hourly averages to Daily averages

```

obs_file = 'DAA2020_hour.csv'
#columns = ['DATE', 'HOUR', 'MIN', 'GHI', 'TOA']
columns = ['DATE', 'HOUR', 'SZA', 'GHI', 'CAMS']
obs = pd.read_csv(
    obs_file,          # path to your input obs file
    header = None,
    skiprows = 1,      # ignore first 5 lines (headers)
    names = columns,   #dtype='object', # name the columns
)
obs

```

```

ValidityDatetime = pd.to_datetime(
    obs['DATE'] + \

```

```

    obs['HOUR'].astype(int).astype(str).apply(lambda x: x.zfill(2))
    ,
    format='%Y-%m-%d%H'
)
obs['ValidityDatetime'] = ValidityDatetime
obs.index = obs['ValidityDatetime'] # setting obs table index to be the
ValidityDatetime columns
obs_day = obs.resample('D').mean()
obs_day = obs_day.reset_index(level=None)
obs_day

obs_day.to_csv('DAA2020_DAY.csv')

```

Clear Sky Model configuration

Example: De Aar station on 16 August (08) 2008 (D=228)

Latitude: -30.67; Longitude: 22.39; Altitude: 1284 m; Time zone: Africa/Johannesburg (SAST or GMT+2)

Air temperature: 12.4; Relative Humidity: 32.3; Ozone: 0.284; Albedo:0.193; Linke
Turbidity:2.5; aod550nm=0.0430125; aod1240nm =0.0188125

```

import datetime
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import pvlib
from pvlib.location import Location
import math
from math import cos

#Station and Date
DeAar = Location(-30.67, 23.99, 'Africa/Johannesburg', 1284, 'DeAar')
times = pd.date_range(start=datetime.datetime(2018,8,16), end=datetime.date
time(2018,8,17), freq='1Min')

#Meteorological input parameters
pressure = pvlib.atmosphere.alt2pres(1284)
dni_extra=pvlib.irradiance.get_extra_radiation(228, solar_constant=1367,
method='spencer', epoch_year=2018)
altitude=1284
temp_air=12.4
relative_humidity=32.3
precipitable_water=pvlib.atmosphere.gueymard94_pw(temp_air, relative_humidi
ty)

```

```

ozone=0.284
albedo=0.193
linke_turbidity=2.5

times_loc = times.tz_localize(DeAar.pytz)
spaout = pvlib.solarposition.spa_python(times_loc, DeAar.latitude, DeAar.longitude)
apparent_zenith = spaout['apparent_zenith']
apparent_elevation = spaout['apparent_elevation']
airmass_relative=pvlib.atmosphere.get_relative_airmass(apparent_zenith, model='kastenyoung1989')
absolute_airmass = pvlib.atmosphere.get_absolute_airmass(airmass_relative, pressure)
asymmetry=0.5*(1+pvlib.tools.cosd(apparent_zenith))
alpha_exponent = pvlib.atmosphere.angstrom_alpha(aod1240nm, 1240, aod550nm, 550)
aod700nm = pvlib.atmosphere.angstrom_aod_at_lambda(aod1240nm, 1240, alpha_exponent, 700)
aod380nm = pvlib.atmosphere.angstrom_aod_at_lambda(aod550nm, 550, alpha_exponent, 380)
aod500nm = pvlib.atmosphere.angstrom_aod_at_lambda(aod550nm, 550, alpha_exponent, 500)
aod_bh = pvlib.atmosphere.bird_hulstrom80_aod_bb(aod380nm, aod500nm)

Bird=pvlib.clearsky.bird(apparent_zenith, airmass_relative, aod380nm, aod500nm, precipitable_water, ozone, pressure, dni_extra, asymmetry, albedo)
Bird.to_csv('BIRDDAA20180816M.csv')

Haurwitz =pvlib.clearsky.haurwitz(apparent_zenith)
Haurwitz.to_csv('HZDAA20180816M.csv')

cosa= pvlib.tools.cosd(apparent_zenith)
# Berger-Duffie = 1350 * (0.70*cosa)
Berger-Duffie = 945 *cosa #945=1350*0.7
Berger-Duffie [Berger-Duffie < 0] = 0
df_out = pd.DataFrame({'ghi': Berger-Duffie})
Berger-Duffie to_csv('BDDAA20180816M.csv')

Simple Solis = pvlib.clearsky.simplified_solis(apparent_elevation, aod700nm, precipitable_water, pressure, dni_extra)
solis.to_csv('SOLISDAA20180816M.csv')

Ineichen & Perez = pvlib.clearsky.ineichen(apparent_zenith, absolute_airmass, linke_turbidity, altitude, dni_extra, perez_enhancement=True)
Ineichen.to_csv('INPDAA20180816M.csv')

```

Merging Clear Sky Model Output

After running a script on a clear sky day to generate 1-minute values, 1-minute values for several number of days is merged by the script below example DeAar station (DAA) Bird clear sky model.

```
import datetime
import numpy as np
import pandas as pd
import matplotlib.pyplot as pylab
import glob
import os

myfiles = glob.glob('C:/Users/brighton.mabasa/CLEAR SKY MODELS/DeAar/Bird/*.csv')

#myfiles = glob.glob("/Bird")
for file in myfiles:
    lines = open(file).readlines()
    open(file, 'w').writelines(lines[:-1])

path = ("C:/Users/brighton.mabasa/CLEAR SKY MODELS/DeAar/Bird/")

#path = "BDD"
all_files = glob.glob(os.path.join(path, "*.csv"))
all_df = []
for f in all_files:
    df = pd.read_csv(f, sep=',')
    #df['file'] = f.split('/')[-1]
    all_df.append(df)

merged_df = pd.concat(all_df, ignore_index=True, sort=True)

merged_df.to_csv("DAA_Bird.csv")
```

Clear Sky Detection

```
import datetime
import numpy as np
import pandas as pd
%matplotlib inline
import pylab
import os
import itertools
from pylab import clearsky, atmosphere, solarposition
from pylab.location import Location
import math
from math import cos
```

```

obs_file = 'DAA_CLEARDETECTION.csv'
columns = ['DATE', 'HOUR', 'MIN', 'GHI', 'CAMS', 'BIRD', 'SOLIS', 'IPN', 'HWZT',
'BD']

obs = pd.read_csv(
    obs_file,          # path to your input obs file
    header = None,
    skiprows = 1,      # ignore first 5 lines (headers)
    names = columns,   #dtype='object', # name the columns
)
obs

```

```

ValidityDatetime = pd.to_datetime(
    obs['DATE'] + \
    obs['HOUR'].astype(int).astype(str).apply(lambda x: x.zfill(2)) + \
    obs['MIN'].astype(int).astype(str).apply(lambda x: x.zfill(2)),
    format='%Y-%m-%d%H%M'
)
obs['ValidityDatetime'] = ValidityDatetime
obs.index = obs['ValidityDatetime'] # setting obs table index to be the
ValidityDatetime columns

```

```

obs_1min = obs.resample('T').mean()

```

```

obs_1min.insert(0, 'month', obs_1min.index.month)
obs_1min.insert(0, 'year', obs_1min.index.year)

```

```

CAMS_clear_1min=pvlib.clearsky.detect_clearsky(obs_1min['GHI'], obs_1min['CAMS'], obs_1min.index, window_length=8, mean_diff=100, max_diff=75, lower_line_length=-5, upper_line_length=10, var_diff=0.005, slope_dev=8, max_iterations=20, return_components=False)

```

```

pvlib.clearsky.detect_clearsky(measured, clearsky, times, window_length, mean_diff=75, max_diff=75, lower_line_length=-5, upper_line_length=10, var_diff=0.005, slope_dev=8, max_iterations=20, return_components=False)

```

```

obs['CAMS_clear_1min']=CAMS_clear_1min

```

```

CAMS_clear_1min = pvlib.clearsky.detect_clearsky(obs_1min['GHI'], obs_1min['CAMS'], obs_1min.index, 8)

```

```

obs.to_csv('DAA_CLEARDETECTION_RESULTS.csv')

```

Calculating the theoretical sunshine duration (N) an input to the Ångström–Prescott (AP) model

```
import pandas as pd
import matplotlib.pyplot as plt
import ephem
observer = ephem.Observer()

#Example(e.g De Aar station, 2017)
#Latitude :-34.01; Longitude:22.38 ; Altitude:192
observer.lat = "-30.67"
observer.lon = "22.39"
observer.elev = 1284

current_year = 2017
all_days = pd.date_range(str(2017) + "/01/01", str(current_year) + "/12/31"
)

sun = ephem.Sun()
day_length = []

for ad in all_days:
    observer.date = ad
    day_length += [observer.next_setting(sun) - observer.next_rising(sun)]

days = 24.*pd.DataFrame(day_length, index=all_days, columns=["day_length"])
days.to_csv('De Aar_2017_N.csv')
```

Calculating the ratio between GHI and GHI at the top of the atmosphere ($Y = \frac{GHI}{GHI_{TOA}}$) and ($X = \frac{n}{N}$) the ratio between in-situ sunshine duration (n) and the theoretical sunshine duration (N) that will be used to calculate *a* and *b*.

```
from statistics import mean
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

obs_file = 'Station Name.csv'
columns = ['DATE', 'GHI', 'TOA', 'n', 'N']
obs = pd.read_csv(
    obs_file,
    header = None,
    skiprows = 1, # ignore first line (headers)
    names = columns)

ValidityDatetime = pd.to_datetime(
```

```

    obs['DATE'] ,
    format='%Y-%m-%d')
obs['ValidityDatetime'] = ValidityDatetime
obs.index = obs['ValidityDatetime']

obs['X'] = obs['n'] / obs['N']
obs['Y'] = obs['GHI'] / obs['TOA']

obs.to_csv('Station Name.csv')

```

Calculating and displaying the Ångström–Prescott (AP) model coefficients *a* and *b*.

```

import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib import style
from statistics import mean
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn import metrics
from scipy import stats
from scipy.stats import linregress
import matplotlib.pyplot as pylab
matplotlib.rcParams['figure.figsize'] = (12,8)
sns.set_palette(sns.color_palette("hls", 20))
sns.set_style("whitegrid")

obs_file = 'STATION NAME.csv'
columns = ['DATE', 'X', 'Y']
obs = pd.read_csv(obs_file, header = None, skiprows = 1, names = columns)

ValidityDatetime = pd.to_datetime(obs['DATE'], format='%Y-%m-%d')
obs['ValidityDatetime'] = ValidityDatetime

obs = obs.replace(np.inf, np.nan).replace(-np.inf, np.nan).dropna()

x = obs['X']
y = obs['Y']

def best_fit_slope(x,y):
    b = ((mean(x)*mean(y)) - mean(x*y)) /
        ((mean(x)**2) - mean(x**2)))
    return b

b = best_fit_slope(x,y)
a = mean(y) - b*mean(x)

def xyline(x, y):
    xypts = min(x.min(), y.min()), max(x.max(), y.max())

```

```

    return xypts, xypts

np.polyfit(obs['X'], obs['Y'], 1)
regression_line=np.polyfit(obs['X'], obs['Y'], 1)
plt.plot(*xyline(x, y), 'k--')
plt.plot(x,regression_line[1]+regression_line[0]*x)
plt.title('TITLE', fontsize=25)
plt.xlabel('n/N',fontsize=25)
plt.ylabel('GHI/TOA',fontsize=25)
plt.plot(x, (a + b*x), label = 'y = {:.3f} + {:.3f}*x'.format(a, b))
plt.legend(loc='lower left',fontsize=25)
plt.scatter(x, y, c=x, cmap='jet')
plt.colorbar()
#plt.savefig("SAVEAS")

```

Statistical Metrics

Calculating MBE; rMBE; MAE; rMAE; RMSE; rRMSE and R²

```

import numpy as np
import pandas as pd
import math
from math import cos
%matplotlib inline
from sklearn.metrics import r2_score
import scipy.stats

```

```

obs_file = 'DAA2020_OBS_BIRDMODEL.csv'
columns = ['DATE', 'HOUR', 'MIN', 'SZA', 'OBS', 'BIRD']

```

```

obs = pd.read_csv(
    obs_file,          # path to your input obs file
    header = None,
    skiprows = 1,      # ignore first 5 lines (headers)
    names = columns,   #dtype='object', # name the columns
)

obs

```

```

obs = obs.replace(np.inf, np.nan).replace(-np.inf, np.nan).dropna()

```

```

x=obs['OBS'] #reference
y=obs['BIRD'] #model

```

```

MBE =np.mean(y - x)
rMBE=np.mean(y - x)/np.mean(x)*100

```

```

RMSE = np.sqrt(np.mean((y - x)**2))
rRMSE = np.sqrt(np.mean((y - x)**2)) / np.mean(x)*100

MAE = np.mean(abs(y - x))
rMAE = np.mean(abs(y - x)) / np.mean(x)*100

r, p = scipy.stats.pearsonr(x, y)
R2=r #R_Squared

summary_obs = pd.DataFrame(index = ['OBS', 'BIRD', 'MBE', 'rMBE', 'RMSE', 'rRMSE', 'MAE', 'rMAE', 'R2'])

summary_obs['Linear Regression, all variables'] = [OBS, BIRD, MBE, rMBE, RMSE, rRMSE, MAE, rMAE, R2]
summary_obs
summary_obs.to_csv('DAA2020_OBS_BIRDMODEL_RESULTS.csv')

```

Plotting (Data Visualisation)

Correlation Graphs

```

import os
import matplotlib
import seaborn as sns
from statistics import mean
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib import style
import matplotlib.pyplot as pylab

sns.set_context('notebook')
sns.set(font_scale=1.5)
sns.set_palette(sns.color_palette("hls", 20))
pylab.rcParams['figure.figsize'] = 12, 8 #image size
plt.style.use('seaborn-whitegrid')
pylab.rcParams['font.size'] = 25 # Legend font size
pylab.rcParams['xtick.labelsize'] = 25 # xaxis tick label size
pylab.rcParams['ytick.labelsize'] = 25 # yaxis tick label size
pylab.rcParams['axes.labelsize'] = 25 # general label size

obs_file = 'DAA_2020_OBS_BIRD.csv'
columns = ['DATE', 'HOUR', 'SZA', 'X', 'Y']
obs = pd.read_csv(obs_file, # path to your input obs file
                  header = None,
                  skiprows = 1, # ignore first 1 lines (headers)
                  names = columns, #dtype='object', # name the columns
                  )

obs

```

```
obs = obs.replace(np.inf, np.nan).replace(-np.inf, np.nan).dropna()
```

```
x = obs['X']
y = obs['Y']
```

```
def best_fit_slope(x, y):
    b = ((mean(x)*mean(y)) - mean(x*y)) /
        ((mean(x)**2) - mean(x**2))
    return b
```

```
b = best_fit_slope(x, y)
print(b)
```

```
a = mean(y) - b*mean(x)
print(a)
```

```
def xyline(x, y):
    """given x and y data, return the endpoints of the y=x line"""
    xypts = min(x.min(), y.min()), max(x.max(), y.max())
    return xypts, xypts
```

```
np.polyfit(obs['X'], obs['Y'], 1)
regression_line=np.polyfit(obs['X'], obs['Y'], 1)
plt.plot(*xyline(x, y), 'k--', linewidth=5)
plt.plot(x, regression_line[1]+regression_line[0]*x)
plt.title('De Aar hourly GHI Observation vs BIRD', fontsize=25)

plt.xlabel('Observation GHI (W/m^2)')
plt.ylabel(('BIRD (W/m^2)'), fontsize=25)
plt.plot(x, (a + b*x), label = 'y = {:.3f} + {:.3f}*x'.format(a, b), lin
ewidth=5, color='red') #format(slope, intercept)
plt.legend(loc='lower left')
plt.legend(fontsize=25)
plt.scatter(x, y, c=x, cmap='jet')
plt.colorbar()
plt.savefig("DAA_BIRD_HOURLY")
```

Line Graph

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib import style
import matplotlib.pylab as pylab
pylab.rcParams['figure.figsize'] = 12, 8 # this sets the default image
size for this session
plt.style.use('seaborn-whitegrid')
pylab.rcParams['font.size'] = 25 # Legend font size
pylab.rcParams['xtick.labelsize'] = 25 # xaxis tick label size
```

```
pylab.rcParams['ytick.labelsize'] = 25 # yaxis tick label size
pylab.rcParams['axes.labelsize'] = 25 # general label size
```

```
obs_file = 'DAA_20161015.csv'
columns = ['DATE', 'HOUR', 'SZA', 'Observation', 'CAMS', 'CMSAF', 'SOLCAST', 'ERA5', 'MERRA2']
```

```
obs = pd.read_csv(
    obs_file,          # path to your input obs file
    header = None,
    skiprows = 1,      # ignore first 1 lines (headers)
    names = columns,   #dtype='object', # name the columns
)
```

```
obs
```

```
fig, ax = plt.subplots()
ax=plt.subplot()
plt.title('Upington Station Hourly Mean GHI 2016-10-15', fontsize=25)
plt.xlabel('Hour', fontsize=25)
plt.ylabel(('GHI (W/m^2)'), fontsize=25)
ax.plot(obs['Observation'], color='black', label='Observation', linestyle='solid', linewidth=4)
ax.plot(obs['CAMS'], color='red', label='CAMS', linestyle='--', linewidth=3)
ax.plot(obs['CMSAF'], color='green', label='CMSAF', linestyle='dotted', linewidth=3)
ax.plot(obs['SOLCAST'], color='blue', label='SOLCAST', linestyle='-.', linewidth=3)
ax.plot(obs['ERA5'], color='m', label='ERA5', linestyle=':', linewidth=3)
ax.plot(obs['MERRA2'], color='orange', label='MERRA2', linestyle='solid', linewidth=3)
plt.legend()
plt.xticks([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14], ['5','6','7','8','9','10','11','12','13','14','15','16','17','18','19'])
plt.savefig("DAA20161015")
```

Bar Graph

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib import style
import matplotlib.pyplot as pylab
pylab.rcParams['figure.figsize'] = 12,7
plt.style.use('seaborn-whitegrid')
pylab.rcParams['font.size'] = 15 # Legend font size
pylab.rcParams['xtick.labelsize'] = 15 # xaxis tick label size
pylab.rcParams['ytick.labelsize'] = 15 # yaxis tick label size
pylab.rcParams['axes.labelsize'] = 15 # general label size
```

```

obs_file = 'DAILY_rMBE.csv'
columns = ['CAMS', 'CMSAF', 'SOLCAST', 'ERA5', 'MERRA2', 'NASA', 'AP']

obs = pd.read_csv(
    obs_file,          # path to your input obs file
    header = None,
    skiprows = 1,      # ignore first 1 lines (headers)
    names = columns,   # dtype='object', # name the columns
)

obs

```

```

ax=obs.plot.bar(figsize=(16,6), width=0.8)
#ax=plt.subplot()
plt.title('Daily relative Mean Bias Error (rMBE)', fontsize=15)
plt.xlabel('Station', fontsize=15)
plt.ylabel(('rMBE (%)'), fontsize=15)
ax.set_ylim([-5,23])
plt.xticks([0,1,2,3,4,5,6,7,8,9,10,11,12], ['Upington', 'Prieska', 'De Aar', 'Bethlehem', 'Irene', 'Mahikeng', 'Polokwane', 'Nelspruit', 'Thohoyandou', 'Mthatha', 'Durban', 'Cape Point', 'George'])
plt.axhline(y = 5, color = 'c', linestyle = '--')
plt.axhline(y = -5, color = 'c', linestyle = '--')
plt.axhline(y = 0, color = 'black', linestyle = 'solid')
plt.savefig("daily_rMBE")

```

Solar Geometry

Calculating solar zenith angle; solar elevation/altitude angle; solar azimuth angle and equation of time.

Example: De Aar station 1 January 2018 to 31 December 2018 1-minute solar angles.

```

from statistics import mean
import numpy as np
import pandas as pd
import datetime
import pvlib
from pvlib.location import Location

```

```

DeAar = Location(-30.67, 23.99, 'Africa/Johannesburg', 1284, 'DeAar')

```

```

times = pd.date_range(start=datetime.datetime(2018,1,1), end=datetime.datetime(2019,1,1), freq='1Min')
times_loc = times.tz_localize(DeAar.pytz)

```

```

spaout = pvlib.solarposition.spa_python(times_loc, DeAar.latitude, DeAar.longitude)

```

```

spaout[['zenith', 'elevation', 'azimuth', 'equation_of_time']].to_csv('DEAAR2018MHRSAST.csv')

```

BSRN data quality control of GHI data

Example: De Aar station 2017

File preparation

- Data is converted to Universal Standard time (UTC) (UTC +2 hours=SAST) i.e., moving the SAST data 2 hours up, e.g., 00:00 SAST data becomes 02:00 in UTC.
- Replace missing values with a placeholder **-999** (BSRN standard)
- Format: YYYY-MM-DDThh:mm; Latitude; Longitude and GHI (4 columns).
- Save the file as text file (.txt) e.g., De Aar_2017.txt (this represents an input file for De Aar in 2017).
- Two BSRN QC tests applied are given below (i) and (ii) with their equations:

(i) “Physically possible” limit check: $-4 W \cdot m^{-2} < GHI < S_a \cdot 1.5 \cdot \cos \theta_z^{1.2} + 100 W \cdot m^{-2}$;

(ii) “Extremely rare” values checks: $-2 W \cdot m^{-2} < GHI < S_a \cdot 1.2 \cdot \cos \theta_z^{1.2} + 50 W \cdot m^{-2}$;

Where S_a is the solar constant adjusted for Earth–Sun distance and $\cos \theta_z$ is the cosine of zenith angle ($\cos \theta_z = \sin \phi \cdot \sin \delta + \cos \phi \cdot \cos \delta \cdot \cos \omega$) given in equation (2.11).

- After quality control, before any analysis, GHI data is converted back to SAST from UTC.
- After BSRN data QC, original 1-minute GHI observation data sets with their corresponding quality codes or flag numbers will be generated, possible codes are summarised in Table A below.

Table A: Flag number/code and its description

Flag number/code	Description of the flag number/code
0	GHI observation data is of good quality, it successfully passed both the physically possible limit and the extremely rare limit tests.
5	Missing observation data or data gaps (usually GHI observation data replaced by a placeholder -999 when archiving)
1	GHI observation data failed/was below the physically possible limit ($GHI < -4 W \cdot m^{-2}$)
2	GHI observation data failed/was above the physically possible limit ($GHI > S_a \cdot 1.5 \cdot \cos \theta_z^{1.2} + 100 W \cdot m^{-2}$)

4	GHI observation data failed/was below the extremely rare limit ($GHI < -2 W \cdot m^{-2}$)
8	GHI observation data failed/was above the extremely rare limit ($GHI > S_a \cdot 1.2 \cdot \cos \theta_z^{1.2} + 50 W \cdot m^{-2}$)
10	GHI observation data failed/was above both the physically possible limit and the extremely rare limit test

- Only the GHI data coded/flagged by number **0** is regarded as good data, data coded (1;2;4;8 and 10) is discarded before any analysed and represented by a placeholder **NaN**, missing data represented by **-999** is replaced by **NaN**. The percentage of data that failed the quality check test per station is given in Table 3.1.

Visual Analysis of BSRN quality controlled GHI data

After BSRN data quality check, data was visualised to check if the data is as expected and to detect major issues, this was done by plotting a time series (Figure A, left) and a histogram (Figure A, right); GHI finger plot (Figure B) and a graphical representation of the two BSRN QC test based on their equations (Figure C).

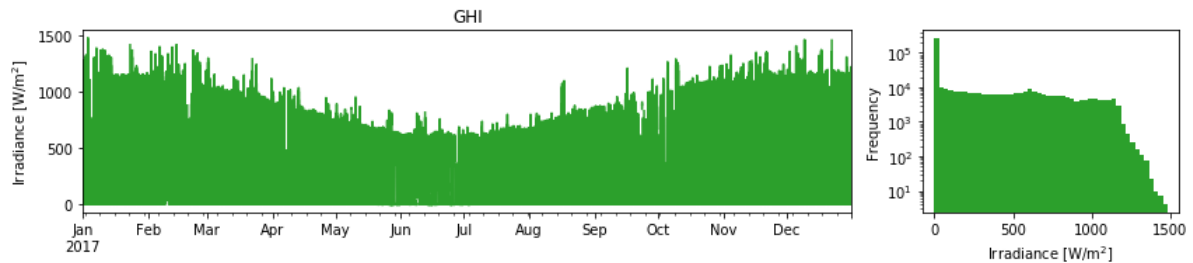


Figure A: Time series (left) and a histogram (right) of (De Aar station, 2017) plotted after data QC to check if the GHI data is as expected.

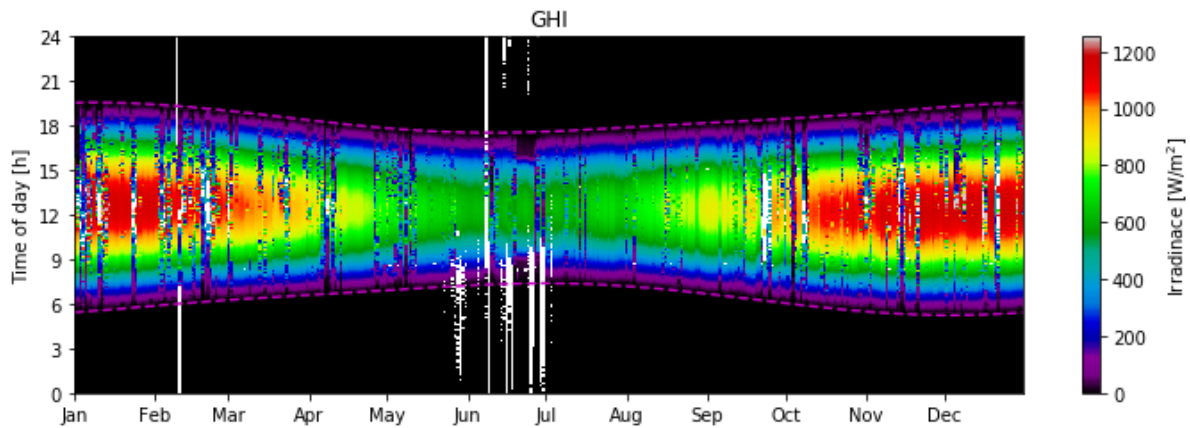


Figure B: Finger plot with horizontal axis corresponding to the day; the vertical axis corresponding to the time of day and the pixel color corresponding to the observed GHI value.

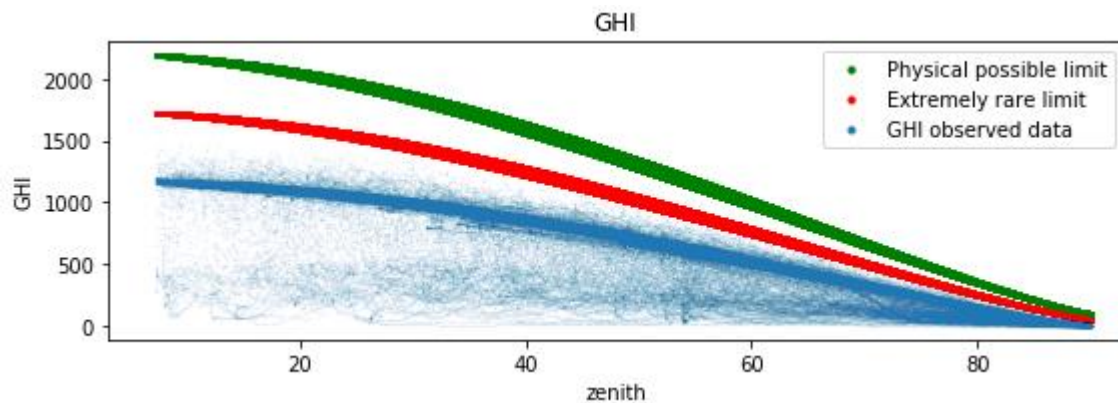


Figure C: A graphical representation of these two BSRN tests applied based on the equations and observed data.

Visual analysis of the observed quality controlled GHI data (Figures A-C) assist in checking if the data is as expected, to detect major issues, to confirm if the BSRN QC test was applied properly, enables to identify missing data and the period when it is missing and it enables to identify if there is constant shading or some disturbances in the data that might have been missed by BSRN QC limits, because the data was in the acceptable ranges.