# Modelling and Validation of Satellite-Derived Hourly Global Horizontal Irradiance in Northeast Iraq Based On a New Quality Control Procedure

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

**Bikhtiyar Mohammed Ameen (MSc)**

School of Geography, Geology, and the Environment

University of Leicester

2019

# Abstract

Modelling and Validation of Satellite-Derived Hourly Global Horizontal
Irradiance in Northeast Iraq Based On a New Quality Control Procedure

**Bikhtiyar M Ameen**

The thesis presents a procedure for obtaining high quality hourly Global Horizontal
Irradiance (GHI) in areas with limited ground measurements of GHI such as northeast
Iraq. This was done by comparing two broadly Quality Control (QC) approaches,
developing a local test and investigating a new one utilising Air Temperature (AT) and
Sunshine Duration (SD) variables for stations when only GHI is available. This also was
achieved by using QC GHI data to validate two Satellite-Derived Datasets (SDDs) in a
new area and correcting SDDs under cloudy-sky conditions with a simple method and by
combing SDD with climate variables as new input combinations to model hourly GHI
data in Artificial Neural Network (ANN) models. The results showed vast differences
between the two QC procedures, while the two QC are applied in literature separately.
The results of the local test showed better quality check than the other two. The new QC
test using SD and AT has detected additional questionable data, which were not flagged
by the other QC approaches. The SD tests can be used as a consistency check in GHI
data, unlike AT tests for the same target. The hourly GHI ground data were well presented
by the two SDDs, while the overall performance of HelioClim-3 version-5 (Hc3v5) is
better than Copernicus Atmosphere Monitoring Service (CAMS) radiation service
version-5 (CRSv3). The results of the bias correction of SDDs under cloudy-sky
condition revealed high accuracy by decreasing bias by 10–80%. The results of modelling
GHI hourly with ANN models were improved slightly by using the new input
combinations generation on bias (0.0%) and root mean square error (9.5%). Overall, the
results suggest that the SD can be used partly as a consistency test in QC of hourly GHI
data, and AT can be used for a general check for the same target. The Hc3v5 and CRSv3
are considered to be a reliable source of GHI data in northeast Iraq. The results of ANN
models with the new input combinations support their continued and future use to
improve the outcomes of the hourly GHI modelling.

# Dedication

This thesis is dedicated to the spirit of all the martyrs of the genocide in Iraq [1], especially for my two brothers (Majed and Wasfy) and my four uncles (Ali, Salam, Karam and Mustafa), and to the spirit of my father and my stepmother who died due to that tragedy.

---

[1] Black, G. Genocide in Iraq: The anfal campaign against the Kurds. Human Rights Watch: 1993.

# Acknowledgements

# Table of Contents

# List of Figures

# List of Tables

# List of Acronyms and Abbreviations

| | |
|---|---|
| *°C* | *Degree Celsius* |
| *ANNs* | *Artificial Neural Networks* |
| *AT* | *Air Temperature* |
| *BSh* | *Hot semi-arid climate* |
| *BSRN* | *Baseline Surface Radiation Network* |
| *BSRN-P* | *BSRN physically possible limit* |
| *BSRN-R* | *BSRN extremely rare limit* |
| *BWh* | *Hot desert climate* |
| *CAMS* | *Copernicus Atmosphere Monitoring Service* |
| *Cosθ* | *Cosine Solar Zenith Angle* |
| *CRS* | *CAMS Radiation Service* |
| *CRSv3* | *CRS version-3* |
| *Cs* | *Clear sky irradiance on the horizontal surface* |
| *Csa* | *Hot-summer Mediterranean Sea climate* |
| *Csb* | *Warm-summer Mediterranean Sea climate* |
| *DHI* | *Diffuse Horizontal Irradiance* |
| *DNI* | *Direct Normal Irradiance* |
| *El* | *Sun elevation angle* |
| *Eq.* | *Equation* |
| *F* | *Flag* |
| *GD* | *Ground Data* |
| *GHI* | *Global Horizontal Irradiance* |
| *GMS* | *The Japanese Geostationary Meteorological Satellite* |
| *GOES* | *Geostationary Operational Environmental Satellite system* |
| *H1* | *Heliosat-1* |
| *H2* | *Heliosat-2* |
| *H4* | *Heliosat-4* |
| *HC3* | *HelioClim-3* |
| *HC3v5* | *HC3 version-5* |
| *ISO* | *International Organization for Standardization* |
| *km* | *kilometre* |
| *KRG* | *Kurdistan Regional Government* |

| | |
|---|---|
| *Kt* | *clearness index* |
| *m/s* | *meter per second* |
| *M1* | *Model-1 (The same until M10)* |
| *MAE* | *Mean Absolute Error* |
| *MAPE* | *Mean Absolute Percentage Error* |
| *MATM* | *Monthly mean of daytime hourly AT in each month* |
| *M-diff* | *Possible max diffuse irradiance equal to 35% of TOA in this study.* |
| *MFG* | *Meteosat First Generation* |
| *MLP* | *multi-layer perceptron* |
| *mm* | *Millimetre* |
| *MODIS* | *Moderate Resolution Imaging Spectroradiometer* |
| *MSG* | *Meteosat Second Generation* |
| *MTSAT* | *Japanese Multifunctional Transport Satellites* |
| *NA* | *not applicable* |
| *NASA* | *The National Aeronautics and Space Administration* |
| *nm* | *Nanometre* |
| *NOAA* | *National Oceanic and Atmospheric Administration* |
| *NREL* | *National Renewable Energy Laboratory* |
| *NSRDB* | *National Solar Radiation Database* |
| *PVGIS* | *Photovoltaic Geographical Information System* |
| *QC* | *Quality Control* |
| *r* | *Correlation coefficient* |
| *$R^2$* | *Coefficient of determination* |
| *rbias* | *Relative bias* |
| *RF* | *Random Forest* |
| *RH* | *Relative Humidity* |
| *RMSE* | *Root Mean Square Error* |
| *rRMSE* | *Relative RMSE* |
| *SARAH* | *The Surface Solar Radiation Dataset- Heliosat* |
| *SD* | *Sunshine Duration* |
| *SDDs* | *Satellite-Derived Datasets* |
| *SE* | *The ratio of The Sun-Earth distance in the astronomical unit* |
| *SEVIRI* | *Spinning Enhanced Visible and Infrared Imager* |

*So*          *Solar constant equal to 1367 W/m$^2$*

*SoDa*        *Solar Radiation Data*

*SOLEMI*      *Solar Energy Mining*

*SPA*         *Solar Position Algorithm*

*SR*          *Solar irradiance*

*SVM*         *Support Vector Machines*

*SZθ*         *solar zenith angle*

*TOA*         *top-of-atmosphere irradiance on the horizontal surface*

*TOACs*       *The name of quality procedure based on rates of TOA and Cs in this thesis*

*TRMM*        *The Tropical Rainfall Measuring Mission*

*USGS*        *United States Geological Survey*

*W/m$^2$*     *watt per square meter*

*WMO*         *World Meteorological Organization*

*WS*          *Wind Speed*

Chapter 1 :   **Introduction and Thesis overview**

## 1.1  Introduction

This chapter gives the general background of the topic, the sun and its energy with defining some terms that are used in this study. It concludes with topics, which lead to draw the literature review map and the structure of the thesis.

## 1.2  General context

The sun is the main source of energy for the earth, which runs the physical, chemical, biological, climatic, meteorological, photosynthetic processes and life on earth (Bojanowski, 2014; Chiras, 2016; Haigh and Cargill, 2015). The energy from the sun comes to the earth as electromagnetic waves. 97% of that energy has a spectral range of 290–3000 nm (WMO, 2008), which generally consists of 1% ultraviolet, visible 45% and near infrared 54% (Soulayman, 2017). The incoming energy from the sun to the earth is based on the blackbody radiation, the sun-earth distance and the surface temperature of the sun. Based on those factors the emitted energy of the sun reaching the earth (Box, 2017; Petty, 2006; Iqbal, 2012). The suns irradiance on the top of the atmosphere on the normal plane perpendicular to the array of the sun is called the solar constant, which is 1367 W/m$^2$ (Rigollier et al., 2000; WMO, 2008; Iqbal, 2012). The solar constant is a general term, and it is calculated based on long-term data from models and sensors aboard satellites at the top of the atmosphere (Bojanowski, 2014; Muller, 2014). It is different by plus or minus of watts based on the activation of the sun every 11 years and the means by which this is calculated is not considered to be a concern for solar energy project on the earth (Box, 2017; Petty, 2006), and in the general context for this study. However, the solar constant is different on the horizontal plane owing to the spherical shape of the earth, and it is also different based on the sun-earth distance from January to March (Iqbal, 2012; Box, 2017). Therefore, the top of atmosphere irradiance on the horizontal plane (TOA) is calculated by multiplying the solar constant by the cosine solar zenith angle (Equation 1.1), according to the location, day of the year and time of the day for each minute or an hour for any areas (Muneer, 2007; Blanc and Wald, 2016). The solar zenith angle is the angle between the beam of the sun and the vertical line on the horizontal surface. The zenith angle is similar to the sun's elevation angle, whereas it is measured from the vertical rather than from the horizontal and it is between 0-90° (Figure 1.1).

$$TOA = (So/SE^2)*(Cos\theta) \hspace{4cm} (1.1)$$

where:

So: Solar constant equal to 1367 W/m$^2$.

Cosθ: Cosine of solar zenith angle.

SE: The Sun-Earth distance in astronomical units. This is used to adjust the solar constant over a year. For instance, its highest value is 1.016682 in July, and the lowest value is 0.983277 in January (Long and Dutton, 2002; NREL, 2017).
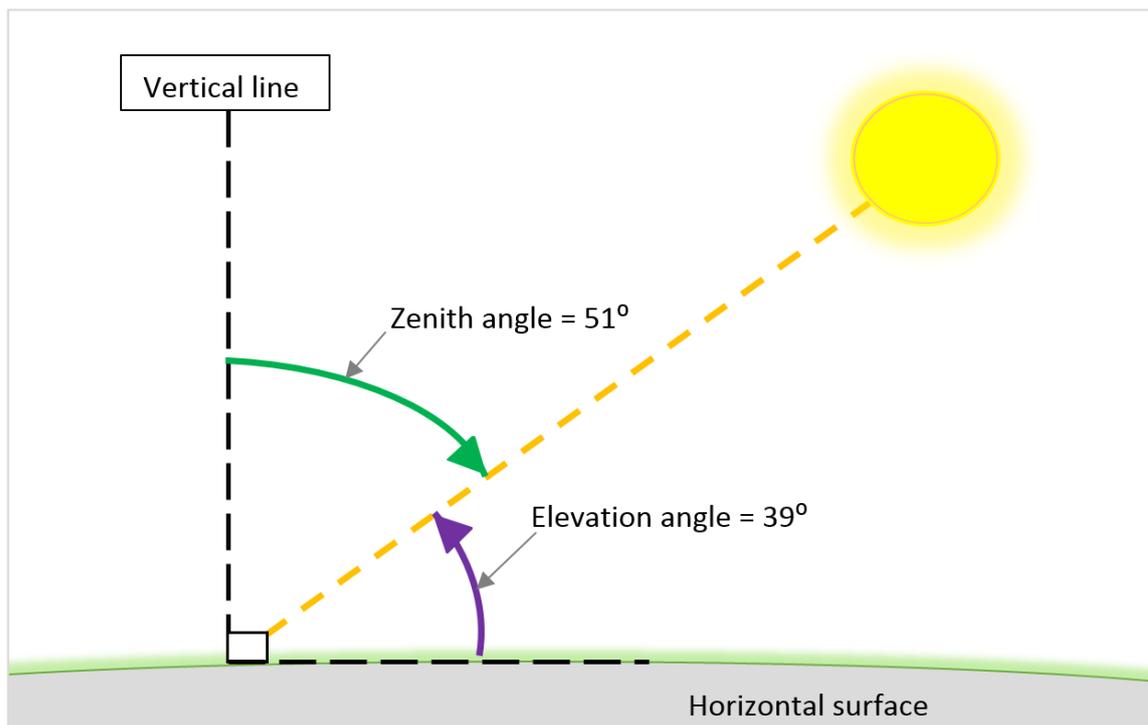


**Figure 1.1:** An example of zenith and elevation angles of the sun's beam irradiance at the earth's surface.

The rate of the TOA irradiance reaching the surface of the earth differs owing to the degree to which it is scattered, absorbed and reflected by clouds, aerosols and gases in the atmosphere (Muneer, 2007; Box, 2017). Figure 1.2 shows the overall irradiance, which reaches the atmosphere for all land and ocean for the earth energy budget. However, the amount is temporally and spatially variable (Box, 2017; Haigh and Cargill, 2015). Therefore, knowing the amount of irradiance, which reaches the earth surface as direct and diffuse radiation scattered by clouds and other atmosphere materials, based on

a location and time, is crucial for fields such as agriculture, architecture, solar energy and hydrology.
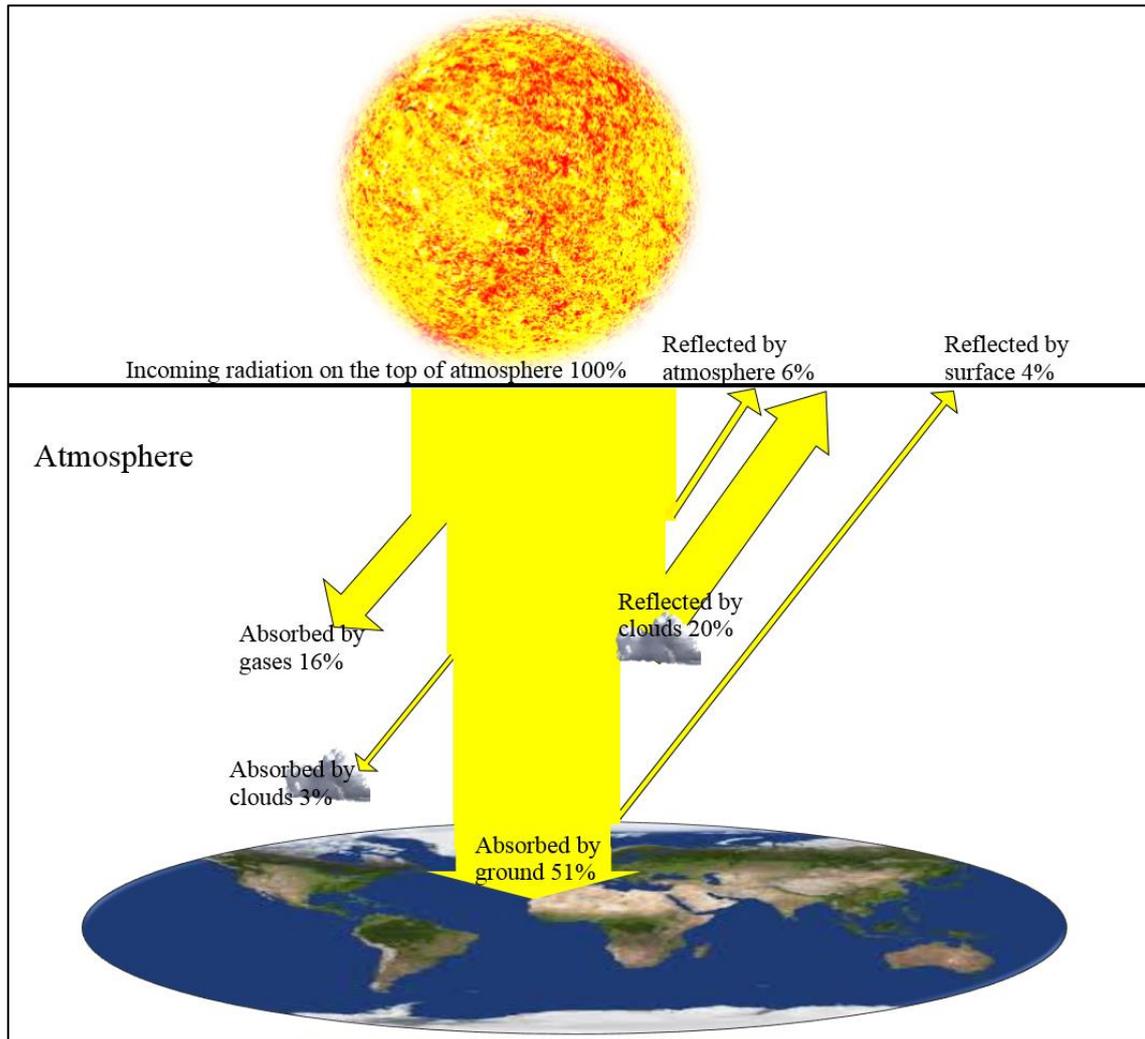


**Figure 1.2:** Overall global annual rates of solar irradiance which reaches the atmosphere and the ground, adapted from (Grüter et al., 1986; Petty, 2006).

This study takes into consideration the amount of solar energy, which reaches the horizontal surface of the earth, which consists of all the solar irradiance components (direct and diffuse), these are called Global Horizontal Irradiance (GHI). There are also some alternative terms and names for GHI such as surface solar irradiance, short wave radiation, insolation, downward radiation and global solar radiation (Alton et al., 2005; Box, 2017; Goswami et al., 2000; Korany et al., 2016; Zhang et al., 2018a). The sun's energy or irradiance from the sun on the earth's surface is calculated as an energy unit (radiation) joule per square centimeter (joule/cm$^2$), which is an accumulative amount of

the energy per unit area in time or as power (irradiance) in watts per square meter (W/m$^2$), which is a mean of irradiance over a given time interval (Badescu, 2008; Goswami et al., 2000; Haigh and Cargill, 2015; Iqbal, 2012). The GHI (W/m$^2$) on the hourly time scale is a concern of this study.

The sources of obtaining or calculating GHI on the earth's surface are ground measurements, satellite data and modelling in several ways.

## 1.2.1    Ground Measurements of Solar Irradiance

The most accurate measurement of irradiance at the earth's surface comes from pyranometers (Bojanowski, 2014; Palmer et al., 2018). There are different types of instruments to measure solar irradiance, which are based on the wavelength targeted detection (KAZ, 2018). The focus of this study is on pyranometers, which generally measure the shortwave irradiance between (285–3000 nm) of spectral ranges, which are UV, visible and infrared (Harrison and Harrison, 2014; Badescu, 2008). There are two main types of pyranometers which measure irradiance; these are blackbody thermopile pyranometers and silicon-cell pyranometers (Harrison and Harrison, 2014). The better instruments are the thermopile pyranometers, which convert nearly all the energy spectrum, unlike silicon-cell pyranometers (Goswami et al., 2000). A thermopile pyranometer consists of an outer and inner glass dome to protect it from long wavelengths and thermal radiations, and a cable connection to a data logger (Figure 1.3-a). The inside contains a thermopile sensor, which has a black coating; this sensor converts the temperature to a voltage. The voltage is converted to watts per meter by dividing it by the sensitivity of a pyranometer (KAZ, 2018). Silicon-Cell pyranometers use a silicon device (photodiode) to convert irradiance from the sun to an electrical current (Figure 1.4) (Goswami et al., 2000; Walter-Shea et al., 2018). The principles of the working of pyranometers vary.

According to solar components, there are three types of instruments, these record Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI) and GHI (Figure 1.3-c, 1.3-b and 1.3-a respectively) (Harrison and Harrison, 2014). The main use of the pyranometer is to record GHI, as shown in Figure 1.3-a; the pyranometer can also be used to record diffuse irradiance along with a shadowing device, which prevents the sensor from the direct irradiance Figure 1.3-b. The beam irradiance is recorded by a

pyrheliometer, which is designed to capture only the direct irradiance from the sun (Figure 1.3-c). The sun tracker, shadowing device and the pyrheliometer need special maintenance, which means few stations have those instruments for recording solar components individually (Mueller et al., 2012). However, the recording GHI pyranometer is also limited owing to the high cost and maintenance (Bojanowski et al., 2013; Hassan et al., 2016).
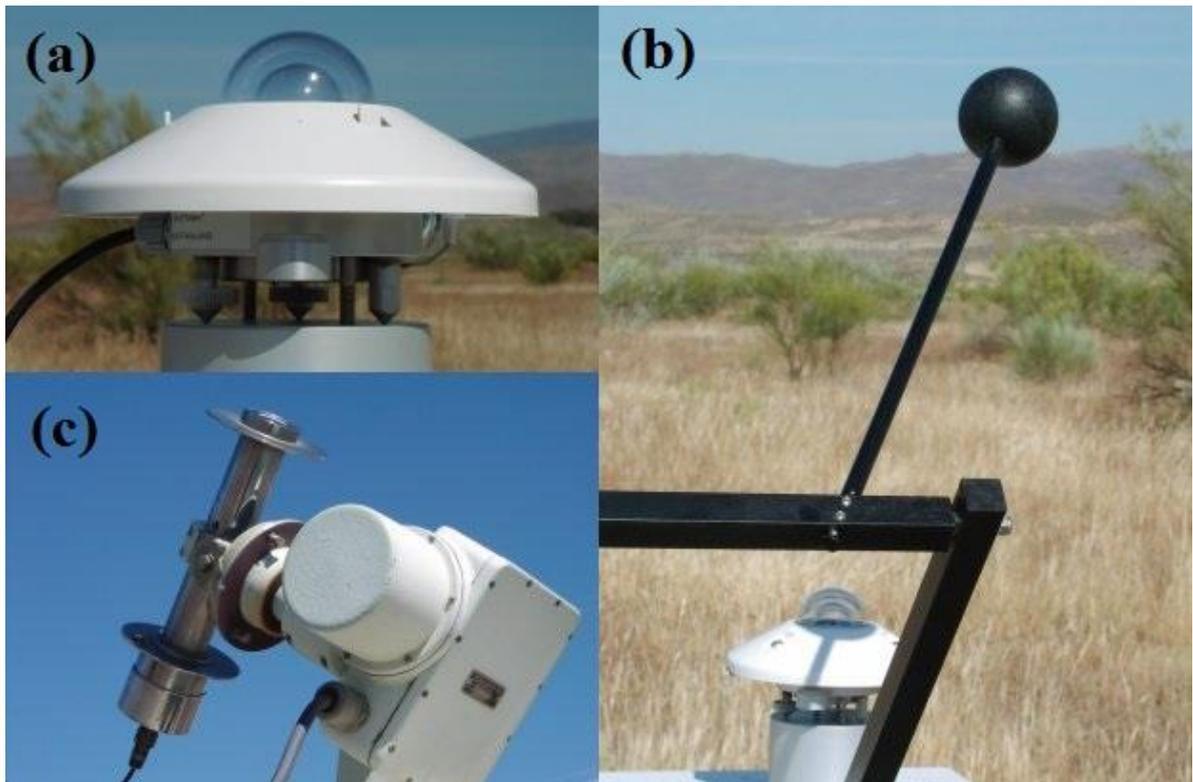


**Figure 1.3.** Thermopile Pyranometer used to measure GHI (a), Same pyranometer with shading ball for DHI measurements (b) and Two-axis tracked pyrheliometer for DNI measurements (c) (REW, 2003).

**Figure 1.4:** Vaisala pyranometer at Dukan station (Taken 2018).

Pyranometers are manufactured by various companies around the world. The broadly used instruments are from Kipp and Zonen, Apogee, Campbell Scientific and Vaisala firms. The instruments have been developed and have been modified several times and know there are various types of instruments in the market. Therefore, the World Meteorological Organization (WMO) and the International Organization for Standardization (ISO) have some specific terms and classifications as classes to show the reliability of pyranometers according to the purpose of use. The classes are Moderate quality, Good quality and High quality for WMO and similarly ISO has Second Class, First Class and Secondary Standard (ISO, 1990; Geiger et al., 2002; WMO, 2008). This study utilised GHI data, recorded by a good quality WMO pyranometers (ISO first class).

All the pyranometers are considered to be sensitive equipment; for this reason regular checks, cleaning, high maintenance and calibration every one-two years are required (KAZ, 2018; Younes et al., 2005; WMO, 2008; Balzarolo et al., 2011). The recorded data needs to be checked or passed with several filters to detect errors in them before use for scientific purposes (Moradi, 2009; Journée and Bertrand, 2011a). The sources of errors include equipment failure, cosine effect, thermal offset, spectral range, non-linearity, sensitivity, power off and other operationally related issues (Muneer and Fairooz, 2002; Younes et al., 2005; Schwandt et al., 2014). Those sources of errors are contributing to

some uncertainty of recorded GHI data according to various types of instruments based on manufacturing development. The sources of errors are as follows.

The cosine effect is that the detector inside pyranometers response to the angle at which radiation strikes the sensor's area. The more acute angle of the sun's irradiance to horizontal surface the more significant errors will be mainly at the sunrise and the sunset when the sun is low in the horizon (Muneer and Fairooz, 2002; Younes et al., 2005).

The thermal offset is a false signal because of the temperature difference between the inner dome and the detector of a pyranometer. This different temperature leads to a potentially significant imbalance in the net infrared radiation budget of the detector and consequently passes it to the output signal. This is because the sensor absorbs not only solar irradiance but also thermal radiation from the instrument body of the instrument. The rate of this error is under 20 $W/m^2$ and it reached to under 3 $W/m^2$ in very recent and developed pyranometers (KAZ, 2018; Sanchez et al., 2015).

The spectral range depends on the pyranometer's manufacturing for detecting the sun's irradiance Spectral. For example, the spectral range of the CMP6 Pyranometer lies between 285-2800 nm and the QMS101 Pyranometer is between 400-1100 nm, which means the error rates of the second one is higher because it does not observe all spectacles from the sun's irradiance. The ability of the detector for the instruments to observe all spectral is also essential to minimise these kinds of errors (KAZ, 2018; VAISALA, 2019).

The non-linearity and non-stability are also related to the capacity of the instruments for work according to their design and their lifetime, which they cause some errors in the data. It should not reach above 1% per year (KAZ, 2018; Muneer and Fairooz, 2002).

Each pyranometer has a specific sensitivity value to calculate the final output from a pyranometer (by dividing millivolt output by the sensitivity). The high difference of temperature and instruments lifetime affects the sensitivity resulting in a decrease in the quality of the data if they are not recalibrated or modified with those factors.

Therefore, there are various Quality Control (QC) methods to detect those errors and issues in the GHI, including statistical measurements, graphs, using assumptions based on the various limits of the TOA and using other climate variables. Those QC approaches will be discussed in the literature section. The ground data need to be quality checked to

reveal the reliability of ground measurements, to evaluate Satellite-Derived Datasets (SDDs) and to test the results of GHI modelling. This thesis deals with QC of GHI in a chapter to check the ability of QC tests to detect errors in the GHI and compares the result of error detection of different methods with developing a new QC method for detecting errors in GHI by utilising Sunshine Duration (SD) and Air Temperature (AT). The literature will be reviewed in the next chapter regarding this.

## 1.2.2 GHI from Satellite Data

Satellite images are considered to be a good source of GHI data at high spatial resolution to cover the limited spatial resolution in the ground measurements (Blanc et al., 2011a; Eissa et al., 2013; Janjai et al., 2009). The most affordable satellite images to cover large areas of the earth are geostationary satellites, namely Meteosat First Generation (MFG) which comprises Meteosat 2–7 providing coverage until 2006 and the Meteosat Second Generation (MSG) which comprises Meteosat 8–10. The Japanese Geostationary Meteorological Satellite (GMS) is also known as Himawari series or Multifunctional Transport Satellites (MTSAT), the current satellite is Himawari-8, and the Geostationary Operational Environmental Satellite system (GOES) (Gherboudj and Ghedira, 2016; Ineichen et al., 2009). These geostationary satellites have a fixed position on the equator, and they move with the movement of the earth providing images for a fixed position every 15 minutes. Hence, the resolution is decreased the further from the centre of the image (Bouchouicha et al., 2016; Marchand et al., 2017). The polar-orbiting satellites such as the National Oceanic and Atmospheric Administration (NOAA) are also utilised for the same purpose (Riihelä et al., 2015; Urraca et al., 2017b).

The methods for converting satellite images to GHI started in the 1960s (Grüter et al., 1986; Bojanowski, 2014). There are two general methods for utilising satellite images to obtain GHI, these are physical and statistical approaches (Cano et al., 1986). The physical approach is based on physical laws, which use radiative transfer models to formulate a relationship between ground and satellite data, to determine the sun's irradiance on the earth's surface (Janjai, 2010; Janjai et al., 2009; Pinker and Ewing, 1985; Gautier et al., 1980). The statistical approach is based on finding a relationship between ground data and information from visible and Infrared bands of satellite images (Mueller et al., 2012; Rigollier et al., 2004). The older methods and early steps are fully described in Grüter et al. (1986).

The methods that are broadly and currently used to estimate GHI from geostationary satellite images are the family of Heliosat methods, which are based on a statistical relationship between cloud properties and ground albedo with using the Clear Sky irradiance on the horizontal plane (Cs) model (Schroedter-Homscheidt et al., 2017; Urraca et al., 2017b; Diallo et al., 2018). The Heliosat method was originally devised by Cano et al. (1986). The method has since been modified several times (Beyer et al., 1996; Mueller et al., 2004; Rigollier et al., 2002; Schroedter-Homscheidt et al., 2017). The modifications and developments to the method involve reducing the number of empirical parameters from Heliosat-1 (H1) to Heliosat-2 (H2) and changing the calculation approach for the cloud index as from H2 to Heliosat-4 (H4). Further improvements of the inputs for calculating the Cs model, which affect the results of the model or change the Cs model among the available Cs models (Cano et al., 1986; Qu et al., 2014; Marchand et al., 2017; Schroedter-Homscheidt et al., 2017). The broadly used model of the Heliosat family models until now is H2, which was published by Rigollier et al. (2004).

The principle of the H2 is that the reflectance in the cloudy pixel is higher than other pixels. The law is based on comparing a cloudy satellite pixel to what should be observed on that pixel if the sky were clear. In this way, the cloud index, which is related to the 'clearness' of the atmosphere, is calculated as in Equation 1.1 as first step. The Cs index is defined as the ratio of the actual GHI to the GHI that would be received if the sky were clear (Equation 1.2). In the second step, the cloud index is corrected based on the Cs irradiance model (Equation 1.3). The method is illustrated in Figure (1.5).

$$n = \frac{p - pg}{pc - pg} \tag{1.1}$$

$$kc = 1 - n \tag{1.2}$$

$$GHI = kc * \text{Cs} \tag{1.3}$$

where $n$ = cloud index, $p$ = albedo of the pixel, $pg$ = ground albedo, $pc$ = albedo of the bright clouds, $kc$ = clear sky index and Cs = GHI under clear sky.
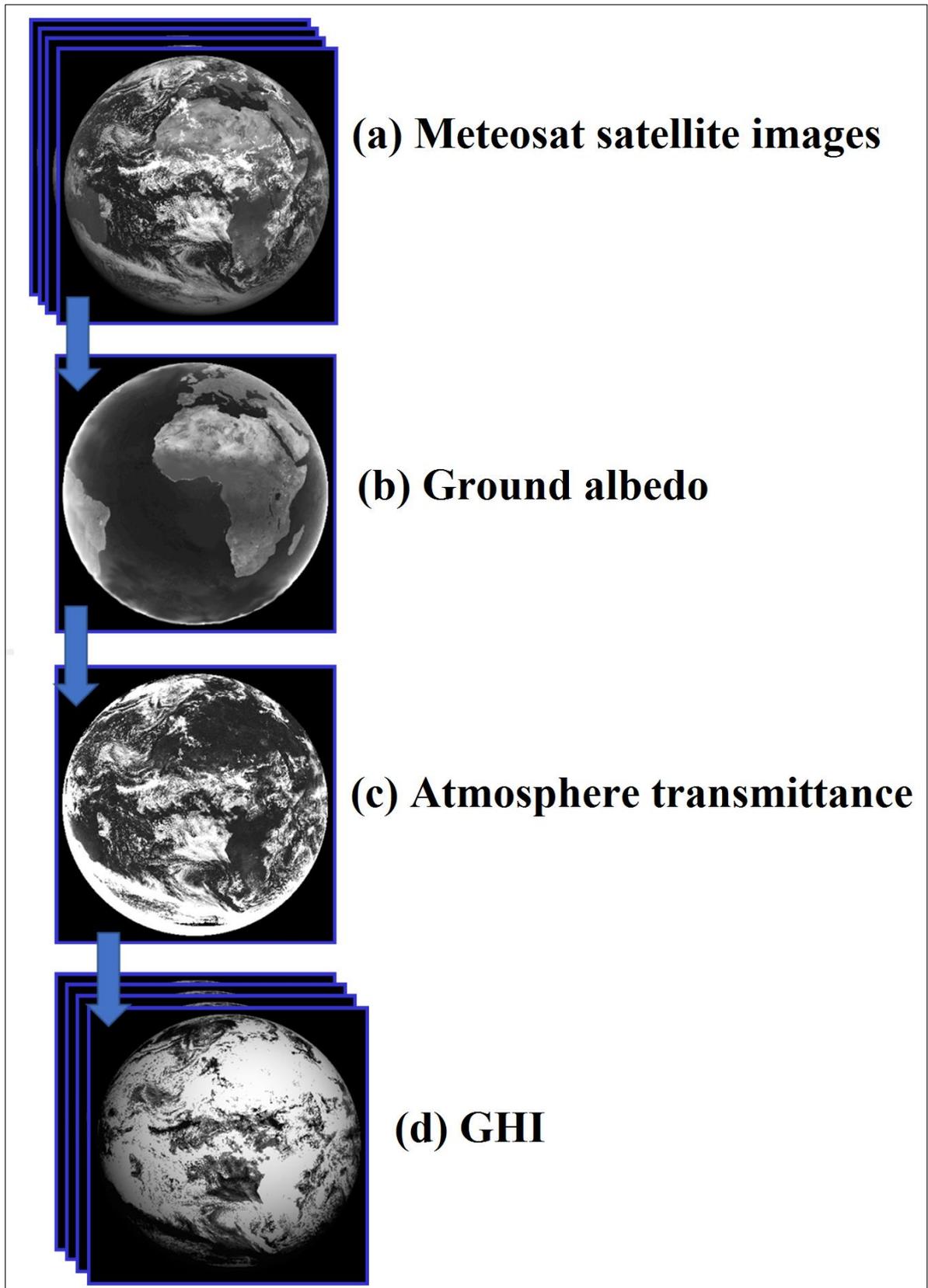
**Figure 1.5**: H2 method adapted from (Rigollier and Wald, 1999).

GHI in cloud-free conditions is estimated as a Cs model with various ways, the basic idea is the comparison of GHI on the earth surface compared to the amount of TOA. Several variables namely solar zenith angle, site elevation, aerosol and gases in the atmosphere need to be addressed to determine that rate on the ground level. Several studies have tried to estimate Cs from a simple algorithm (Reno et al., 2012; Gueymard and Myers, 2008). Cs model is in progress by improving the quality of inputs to the model, which they determine the atmosphere optical depth and its properties from various sources (Zhang et al., 2018b; Munkhammar and Widén, 2018; Eissa et al., 2015a). The simple Cs model is to use TOA with the cosine of solar zenith angle and empirical coefficient by:

$$Cs = 0.70*TOA*(Cos\theta) \tag{1.4}$$

Full information on several Cs models and their development can be found at (Reno et al., 2012; Scharmer and Greif, 2000; Rigollier et al., 2000; Gueymard and Myers, 2008). Other developed Cs models are Bird (Bird and Hulstrom, 1981) and The European Solar Radiation Atlas (ESRA) (Scharmer and Greif, 2000). A recent development is the McClear Cs model (Lefèvre et al., 2013), which uses all inputs (solar zenith angle, ground albedo, ozone and water vapour column, aerosol, gases, time interval, location and elevation), some of these inputs are calculated from satellite images. The McClear Cs model has been validated in several areas (Eissa et al., 2015b; Lefèvre and Wald, 2016), and has a better performance compared to previous Cs models (Schroedter-Homscheidt et al., 2017). The model has been used to improve HelioClim-3 (HC3) (Qu et al., 2014) and it has been used in Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service (CRS) (Schroedter-Homscheidt et al., 2017). Full details about the McClear Cs model is published in Lefèvre et al. (2013). This thesis has utilised hourly GHI from the McClear Cs model in various way in Chapters 4, 5 and 6.

Based on the H2 and Heliosat family and other methods (Habte et al., 2017; Gautier et al., 1980; Janjai, 2010; Janjai et al., 2009) for converting satellite images to estimate GHI and other solar components, several datasets provide solar irradiance components at the earth's surface on different time scales. Full detail and information about those datasets can be found in Polo et al. (2016) and Urraca et al. (2017b). Generally, the solar irradiance estimated from satellite sources and the above-mentioned methods are called SDDs. For example, SDDs namely HC3, Photovoltaic Geographical Information System

(PVGIS), The Surface Solar Radiation Dataset- Heliosat (SARAH) and Solar Energy Mining (SOLEMI) are created from H2 and have differing numbers of inputs or use inputs from satellites or new Cs models. Another example of SDDs is CRS, which is created by H4, which has the same principle of H2, but uses cloud properties from APOLLO (AVHRR Processing scheme Over cLouds, Land and Ocean). It is a series term for analysing various cloud types and cloud properties from geostationary satellite images (Schroedter-Homscheidt et al., 2017; Klüser et al., 2015).

SDDs provide GHI, DNI and DHI data in sub-hourly, hourly, daily and monthly based on the original satellite image resolution, and they are an essential source for obtaining solar irradiance data especially in the area with few ground measurements. They have been validated in several areas, but further research is required to evaluate them in the new areas to demonstrate their accuracy. Therefore, in the literature review section, the extensive studies that examine those SDDs will be discussed to identify knowledge gaps.

### 1.2.3   Modelling Global Horizontal Irradiance (GHI)

GHI is required for several fields such as hydrology, architecture, agriculture and solar energy applications with different time scales. The ground measurement is limited, owing to high maintenance and a high cost of instruments. In the early twentieth century, the equipment for directly measuring GHI was rare. Currently, the *in-situ* measurements are still not wide (Spencer, 1982; Fan et al., 2018). Those reasons encouraged the researcher to model GHI in a variety of techniques. The most common equation to estimate GHI from SD was developed in 1924 by Angstrom (1924). This has since been developed and is known as the Angstrom-Prescott equation (Njau, 1996; Liu et al., 2009; Paulescu et al., 2016b). Other studies have used cloud cover and minimum and maximum AT to estimate GHI (Supit and Kapper, 1998). Even due to the limitation of those variables, other studies tried to estimate GHI from AT which is available at most stations (Bristow and Campbell, 1984; Ododo et al., 1995; Jong and Stewart, 1993; Hassan et al., 2016). Several other climate variables (SD, AT, Relative Humidity (RH), Wind Speed (WS), and cloud cover) and set of variables from sun's position and a geographical location with a variety of algorithms, statistical techniques and empirical models have been used to estimate GHI (Mohanty et al., 2016; Zhang et al., 2017; Yadav and Chandel, 2014a; Chukwujindu and Ogbulezie Julie, 2017; Besharat et al., 2013; Chukwujindu,

2017). The GHI ground data record is limited even in developed countries, studies have tried to interpolate station data to cover a large area, and other studies have used satellite data to cover those limitations (Palmer et al., 2017; Urraca et al., 2017c; Zou et al., 2016; Loghmari et al., 2018; Gutierrez-Corea et al., 2014).

Using various methods and several variables to model GHI leads studies that have classified GHI models. For example, the empirical, statistical and physical model types are broadly mentioned (Grüter et al., 1986; Mohanty et al., 2016; Bojanowski, 2014). The models are classified as five major approaches for obtaining GHI, which are: parametric, statistical, reanalysing, SDD and interpolation (Urraca et al., 2017c). In recent studies, empirical models (Fan et al., 2019), interpolation techniques (Loghmari et al., 2018), linear regression with various combination of variables (Keshtegar et al., 2018; Doorga et al., 2019) and physical models with satellite images (Kumler et al., 2019; Wang et al., 2019) are used to estimate GHI. Machine learning algorithms are broadly used to model GHI either with satellite images or with a variety of ground based climate variables (Koo et al., 2019; Jadidi et al., 2018; Khosravi et al., 2018b; Antonanzas-Torres et al., 2015). Applying any of these approaches for the purpose of GHI modelling depends on the availability and the limitation of the inputs, e.g. climate variables, satellite data or model assumptions.

The direction of the new research in the field is going to model GHI in a high spatial-temporal resolution such as sub-hourly or hourly timescales with machine learning approaches, and obtaining high quality GHI by applying various QC tests and combining SDDs and ground based measurements for the target. Therefore, the QC approaches, SDDs and GHI modelling with Artificial Neural Network (ANN) models will be discussed in detail in the literature review chapter.

## 1.3    Thesis structure

The thesis is comprised of seven chapters. They can be considered in three main sections (Figure 1.6). The first three chapters in Section one consist of an introduction to the study with general context, background information and outlining key concepts (Chapter 1). The literature is reviewed broadly in three parts; these are QC approaches to test GHI, validation of SDDs and modelling GHI with various methods and time scales, to identify the gaps and informing research questions and objectives (Chapter 2). The

geography of the study area is described regarding location and climate variables in Chapter 3 and with a description of the evaluation criteria used in this study.

Section 2 consists of the analyses and results chapters:

Chapter 4 compares the results of three QC procedures testing GHI against each other and investigates a new QC test methodology by utilising SD and AT variables to check hourly GHI data for consistency test.

Chapter 5 uses the quality checked GHI data from Chapter 4 to validate the spatial and temporal variability of the two SDDs (HC3 version-5 [HC3v5] and CRS version-3 [CRSv3]) in all-sky, clear-sky and cloudy-sky conditions in northeast Iraq. This chapter also uses ground data to evaluate other pixel point data from the two SDDs. The chapter also applied a simple method for bias correction in SDDs under cloudy-sky condition.

Chapter 6 describes a method to estimate hourly GHI data with ANN and regression models in areas with limited ground measurements by using new inputs from SDDs (Chapter 5), climate variables and GHI (Chapter 4) in ten input combinations.

The final section is Chapter 7, which provides a synopsis of the overall thesis contributions and general discussion, highlighting the major conclusions and outcomes. Limitations of the study are identified with recommendations for further research based on the limitations and the findings.
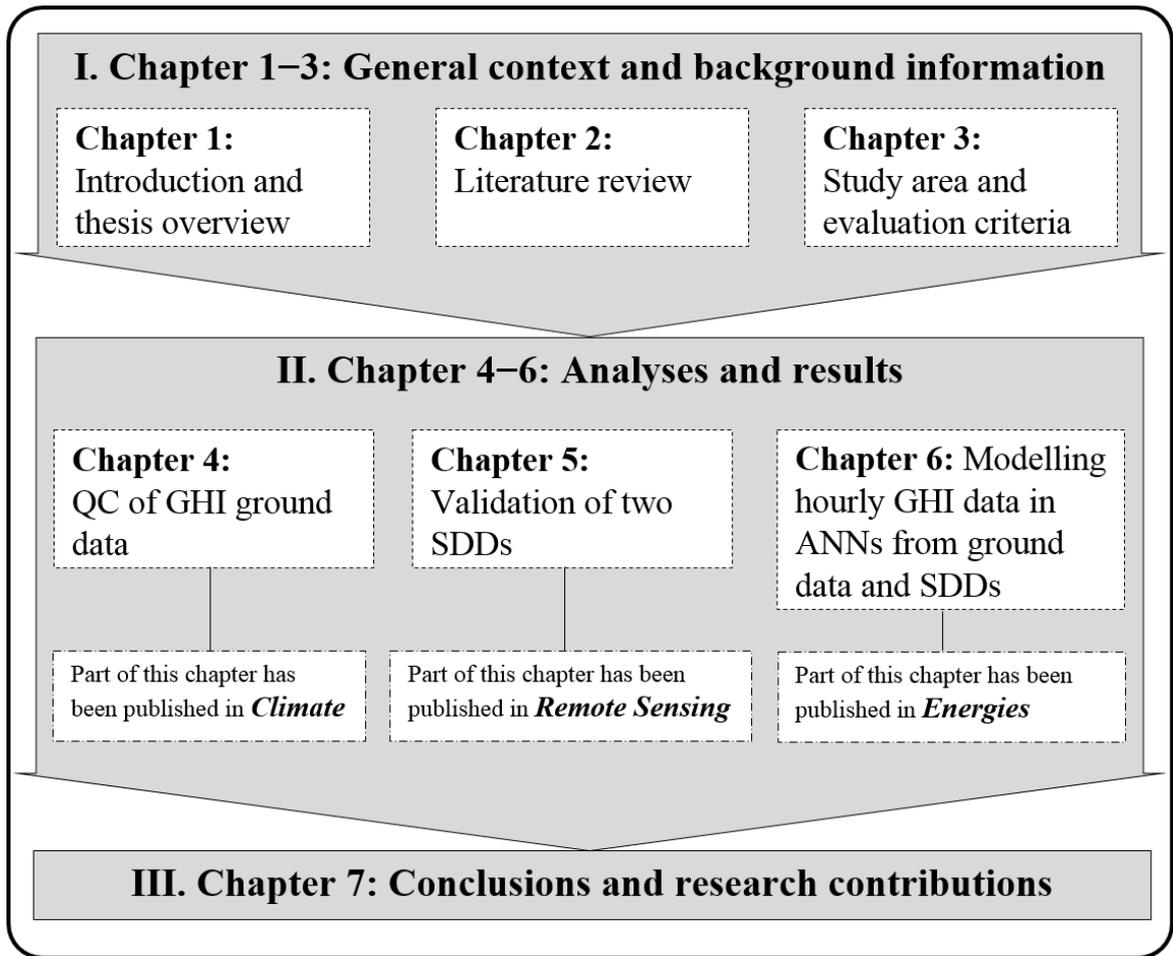
**Figure 1.6:** Thesis structure.

Chapter 2 :  **Literature Review**

## 2.1 Introduction

GHI data is crucial for several fields such as solar energy, agriculture, architecture and hydrology, and thus there are a high number of studies on this subject. Indeed, there are weekly publications about GHI in its various aspects and branches of knowledge. This thesis focuses in particular on three topics in the literature; these are QC of ground measurements, validation of SDD and modelling GHI with several inputs.

In the first part of this chapter, literature on QC methods of GHI is reviewed. Next, the literature regarding the validation of satellite data and SDDs in various climate regions is explored, and then the literature concerning modelling GHI with various algorithms and inputs with a focus on ANN models is detailed. Finally, the knowledge gaps from each of the review sections are identified and presented, and these formulate the research questions and objectives of the thesis.

## 2.2 Quality Control of Ground Data

### 2.2.1 Introduction

Solar irradiance is considered to be one of the most critical climate variables. It affects other climate variables and it is crucial for research fields including climate change, renewable energy, agriculture, architecture and hydrology (Pashiardis and Kalogirou, 2016; Zo et al., 2017). Therefore, high quality solar irradiance data are needed. It can be estimated from satellite images (Rigollier et al., 2002; Janjai et al., 2009; Dubayah, 1992; Qin et al., 2011) and predicted it using other climate variables (Ampratwum and Dorvlo, 1999; El-Metwally, 2004; Hassan et al., 2016; Yadav et al., 2014b). In both cases, high quality ground data are required for validation. Solar irradiance is measured at the ground level with precision by pyranometers, albeit with some uncertainty due to technical issues of the instruments, which include the cosine effect, temperature response, sensitivity, non-linearity, spectral range and thermal offset (Younes et al., 2005). In addition, there are operational and installation errors from miscalibration, a lack of regular cleaning of dust, snow, water droplets and bird droppings, and shadows cast on the equipment by nearby trees and buildings (Muneer and Fairooz, 2002). These factors cause systematic and random errors in the data, which have been reported in the literature (Long and Shi, 2008; Moreno-Tejera et al., 2015; Roesch et al., 2011; Schwandt et al., 2014; Shi et al., 2008). Therefore, sensors that measure solar

irradiance are unlike other meteorological instruments in that they need high maintenance to sustain performance and collect of high quality data, and their data recording needs to be checked regularly before they are used in scientific studies (Younes et al., 2005; Pashiardis and Kalogirou, 2016; Perez-Astudillo et al., 2018). Hence, new pyranometers have been developed, and some of the equipment errors have been almost eliminated, e.g. snow melt on the pyranometer dome (Blonquist et al., 2009). The QC approaches and the way for checking in literature are listed as follows:

### 2.2.2 Quality Control methods

Several studies and organisations have proposed models for the QC of solar irradiance data to detect errors using a variety of tests. The tests recommended by the Baseline Surface Radiation Network (BSRN) of the WMO are the most widely used tests in the literature (García Cabrera et al., 2018; Long and Shi, 2008; Moreno-Tejera et al., 2015; Ntsangwane et al., 2018; Perez-Astudillo et al., 2018; Roesch et al., 2011; Yang et al., 2018; Zo et al., 2017). This method was proposed by Long and Dutton (Long and Dutton, 2002). The test consists of several sub-tests, which are used to check errors of the physically possible limit, extremely rare limit and comparison test. The comparison test independent on the availability of all solar components, but this study is dealing with GHI only not DNI and DHI. Some other QC tests are formed in equations based on some assumptions by utilising the data of solar constant and solar zenith angle.

Other prominent tests are those, which depend on the TOA and Cs model for testing the physically possible limit and extremely rare limit, respectively. This test is named TOACs in this study and is presented by Geiger et al. (2002). It has been applied widely in the literature (Moradi, 2009; Tang et al., 2010; Khaliliaqdam and Soltani, 2012). This comparison of BSRN and TOACs tests has not been documented in depth in the literature. The National Renewable Energy Laboratory (NREL) of the United States has also developed QC software to quality check solar irradiance data based on the ratio between global and beam irradiance (Maxwell et al., 1993), that QC software cannot be applied when only GHI is available.

Others QC tests have set the maximum limit of solar irradiance data as 120% of TOA to check a major problem in the data by flagging any value above that limit. (Muneer and Fairooz, 2002; Younes et al., 2005). Some studies (Pashiardis and Kalogirou, 2016;

Lemos et al., 2017) have used a combination of tests in the literature such as a subtest of BSRN, a subtest of TOACs and other tests as a quality check without comparing and investigating the different results of each test for the same target.

### 2.2.3   Consistency Test or Comparison Test

Comparisons between solar irradiance components as a test to check the consistency and plausibility of the solar irradiance data based on the relationship between diffuse, direct (beam) and global irradiance have been studied intensely (Muneer and Fairooz, 2002; Younes et al., 2005; Long and Shi, 2008; Ineichen, 2013). Some studies have assessed the relationship between beam, diffuse, and GHI as an index for detecting errors. For example, several studies have used checks such as direct irradiance is lower than GHI, diffuse is lower than 110% of GHI and the sum of direct and diffuse is within ± 8% of GHI (Maxwell et al., 1993; Muneer and Fairooz, 2002; Younes et al., 2005; Journée and Bertrand, 2011a; Ineichen, 2013). Pyranometers, pyrheliometer and sun trackers need high maintenance and are costly (WMO, 2008). Therefore, meteorological stations often lack the capability of recording solar irradiance data, particularly all of its components. This means that most of the above tests cannot be applied when just GHI is available. This study addresses this problem area.

### 2.2.4   Statistical Index

GHI data have also been checked by utilising statistical indexes such as the ratio between the first and third quartiles to determine the rates of lower and upper outliers in the data. This is to check the normal operation in a station based on those rates. If the outlier rates are low, it means data quality is good and vice versa (Muneer and Fairooz, 2002; Shi et al., 2008; Younes et al., 2005). The ratio of standard deviation and the mean of the GHI with TOA data have also been used as conditional operators of an equation for a persistence test of GHI; for further details, see (Journée and Bertrand, 2011a). This study has also utilised some statistical indices for setting the test limits.

### 2.2.5   Satellite-Derived Datasets (SDDs) for Testing GHI

GHI of SDD is the same as GHI ground measurements, but are estimated from satellite images. The role of SDDs for checking GHI ground measurements have been addressed by two similar studies as a method (Urraca et al., 2018a; Urraca et al., 2017a).

They have mentioned that the minor deviation and some operational errors (snow, soiling and shading on sensor and data logger fault) and equipment errors (miscalibration and low quality sensor) can be detected by the method. They proved that most of these errors cannot be detected by conventional QC approaches.

Similarly, Urraca et al. (2018c) address uncertainty of annual ground measurements as yearly bias by comparing them with SDDs. The QC methods in their study were applied for daily or annual data. Therefore, it is not clear how this works for hourly data as per this research. In addition, the SDDs need to be checked with high quality controlled data before using them as a reference to check GHI ground, which is limited in the study area where GHI ground data are rare.

### 2.2.6 Graphical Representation to Detect Errors

Graphical plots are considered a good way to demonstrate data trends. In most of the previous studies, graphics are used to display the errors by flagging the GHI data with a specific flag number for each error type, which is useful for the user to see the errors and question the data easily. For instance, Moreno-Tejera et al. (2015) investigated a new kind of graphic to show the errors as flags in BSRN tests. They state that their graphic is a good representation for the errors. Other kinds of graphics have also been used to show the errors of daily GHI among several datasets based on QC methods (Urraca et al., 2018a; Urraca et al., 2017a). Others have shown that various scatter plots can be used to determine the border of tests and to demonstrate the uncertainty in the GHI data (Yang et al., 2018; Younes et al., 2005; Moradi, 2009). However, a few studies (Pashiardis and Kalogirou, 2016; Moreno-Tejera et al., 2016) mention a figureprint plot to determine errors. It is hard to find studies, which have investigated the use of general plot to detect systematic errors in GHI by SD and AT variables. In general, some of those graphics are not easily readable and understandable for a user to examine the errors in the solar irradiance data. Therefore, this study uses a new style of fingerprint plot and uses general plots to plot variables of SD, AT and GHI for detecting systematic errors and other questionable data.

### 2.2.7 Comparison between Pyranometers

Several studies have dealt with faults and issues regarding instruments. For instance, Blonquist et al. (2009) have compared several pyranometers in the field to select their

quality and their work. Some other studies have investigated the errors regarding the thermopile and photodiode pyranometers (Al-Rasheedi et al., 2018; Habte et al., 2015; Geuder et al., 2014; Wilbert et al., 2016). The others have addressed specific issues namely, temperature response and thermal offset (Sanchez et al., 2015; Ji and Tsay, 2010), cosine response (Michalsky et al., 1995), spectral response (Vignola et al., 2016), and calibration (Kim et al., 2018; Mathijssen et al., 2018; Olano et al., 2015). In all the above cases, the reference pyranometer is required, which is unavailable in the case study. Therefore, this study is addressing other ways for QC of GHI data.

### 2.2.8 Climate Variables for Test GHI

Regarding the use of other climate variables for testing the quality of GHI data, SD and AT are used as below.

### 2.2.8.1 Sunshine Duration for Test GHI

SD has been used for testing daily GHI data by Muneer and Fairooz (2002) and monthly GHI data by Younes et al. (2005). Moradi (2009) also investigated a model, which is based on the lower limit of the SD index, to test daily global irradiance. Moradi's (2009) method is suited especially for those stations for which direct and diffuse irradiance components are not available, as (Khaliliaqdam and Soltani, 2012) confirm. Recently, Journée and Bertrand (2011a) have also tested hourly DNI data with SD. In Chapter 4 of this thesis another way to test hourly GHI with SD is presented. The use of upper and lower limit of SD with some conditionals of the TOA rates to check hourly GHI data is limited in the literature.

### 2.2.8.2 Air Temperature for Test GHI

Uses of AT to assess GHI data are scarce in the literature, despite the known relationship between the two variables (Zahumenský, 2004; Hassan et al., 2016; Prieto et al., 2009) and several models have also used AT to estimate GHI (Hassan et al., 2016). AT has been used for testing longwave radiation (Long and Shi, 2008). The same study also used the lower bound of AT for snow melting to test the sum of global shortwave radiation, whereas utilising AT within different circumstances to test hourly GHI is not seen in the literature.

## 2.3 Validation of Satellite-Derived Datasets (SDDs)

### 2.3.1 Introduction

The ground measurements of GHI have high accuracy and high temporal availability, whereas the high spatial resolution of recorded data and the number of stations with solar irradiance data are limited in most geographical areas. The reasons for this are the high purchase and high maintenance costs of pyranometers. Satellite images have been analysed to estimate GHI in order to cover the scarcity of ground measurement data. Most of the affordable satellite images for that purpose are the geostationary satellite images (Gherboudj and Ghedira, 2016). Others such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Qin et al., 2011) and Landsat images have been used (Xie et al., 2012), but their temporal resolution is not acceptable. The basic idea of estimating GHI from satellite images is to find the relationship between satellite images and ground measurements, either with statistical or physical approaches (Cano et al., 1986). The popular method for that is Heliosat-2 method, which was described in (Chapter 1, Section 2.2). Other studies also used satellite imagery with different techniques for GHI estimations (Janjai et al., 2009; Janjai, 2010; Janjai et al., 2011).

There are several SDDs for establishing, measuring, modelling and estimating GHI (Polo et al., 2016; Zhang et al., 2016). Some of those well-known SDDs are described in (Table 2.1). In addition, SDDs have been utilised in several ways, for example, an SDD from MSG has been used to create a solar map (Polo, 2015). Studies which merged ground data with the SDD for the same above purpose, reveal that the merging technique for producing a solar map is better than interpolating ground data (Journée and Bertrand, 2011b; Journée et al., 2012). SDDs have also been combined with meteorological data to calibrate a GHI model (Bojanowski et al., 2013). The same data combinations have also been analysed to create GHI datasets for crop modelling over Europe (Roerink et al., 2012). SDDs have been utilised to assess long-term trends of a GHI time series (Sanchez-Lorenzo et al., 2015). SDDs are quite useful because of the limitations of ground data for GHI applications.

There are also some reanalysis datasets for GHI and other solar components, which are described in (Zhang et al., 2016). More broadly, Zhang et al. (2016) have evaluated the result of the six re-analysed datasets for obtaining GHI with ground data from

measurement networks such as the BSRN and others for 674 cities around the globe, with an overall bias found to be 11–50 W/m$^2$. They used a large volume of data in various climate regions and countries; however, the results are shown according to the datasets and measurement networks, rather than for each station. This is useful for comparison between the SDDs and measurement networks, but it does not reflect a real situation for the individual stations. Urraca et al. (2018b) have also validated three global reanalyses of GHI data worldwide, and the result shows positive bias. Another study has merged SDDs and reanalysis products to create a long-term GHI dataset (Feng and Wang, 2018a).

**Table 2.1:** Datasets and services providing solar irradiance time series derived from satellite information

| Name | Time scale | Coverage | Method | Web site |
|------|-----------|----------|--------|----------|
| NSRDB | 30 min | USA-India | Physical Solar Model | http://rredc.nrel.gov/solar/old_data/nsrdb/ |
| NASA | 3- hourly | World | - | http://gewex-srb.larc.nasa.gov/ |
| PVGIS | Hourly | Europe-Africa-Asia | Heliosat family | http://re.jrc.ec.europa.eu/pvgis/ |
| CM – SAF(SAHRA) | Hourly | Europe-Africa | Heliosat family | http://www.cmsaf.eu/ |
| SOLAMI | 15 min | Europe-Africa-Asia | H2 | http://wdc.dlr.de/data_products/SERVICES/SOLARENERGY/ |
| HelioClim | 15 min | Europe-Africa-Asia | H2 | http://www.soda-pro.com/web-services/radiation/helioclim-3-archives |
| CAMS | 15 min | Europe-Africa-Asia | H4 | http://www.soda-pro.com/web-services/radiation/cams-radiation-service |

SDDs are necessary for many fields of research because they provide GHI for many areas and countries. Therefore, the validation of SDDs is crucial to investigate their reliability by using various methods in different geographical and climate areas. Several prior studies achieved that based on the following datasets.

### 2.3.2 National Solar Radiation Database (NSRDB)

National Solar Radiation Database (NSRDB) is a database for providing GHI, other solar components and other climate variables in the United States and surrounding

countries. The GHI by NSRDB was created using the Physical Solar Model (PSM), which utilises satellite images and information from the satellites sources from GEOS, NOAA, MODIS and the National Ice Centre and Modern Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) (Sengupta et al., 2018). The PSM has been developed by NREL with their collaborators, and it shares similar principles with H2 for calculating cloud properties and calculating GHI (Sengupta et al., 2015).

The database has been validated by several comprehensive studies. For instance, the GHI from NSRDB has been validated with quality controlled ground data for several stations across the United States for periods of 1998-2015 (Habte et al., 2017) and 2005-2012 (Sengupta et al., 2015). The results of both studies show good agreements between ground data and SDD with a bias of $\pm$ 5%. A similar database has been validated in a United States context, but only in California with 27 stations. The overall results of validation showed overestimation by a 5% bias and showed high seasonal deviation and deviation from morning to evening in various climate types in the State (Nottrott and Kleissl, 2010). Some other studies have also evaluated the NSRDB and other SDD over the United States in different climate regions against several ground measurements (Gueymard and Wilcox, 2011; Habte et al., 2012; Lave and Weekley, 2016; Xia et al., 2017).

This demonstrates the importance of SDD and their validation in developed countries to provide a high spatial resolution of GHI owing to the limitation of ground data. Therefore, it seems clear that the validation of SDD in developing countries is considered to be highly crucial because the ground data also is limited.

### 2.3.3 GHI derived from Himawari-8

Himawari-8 is a current operator of GMS. The images from Himawari-8 have been used to estimate GHI in East Asia and Australia. The sub hourly GHI data have been validated with quality controlled GHI ground data at four stations in Japan. The result revealed good agreement by a mean bias of 20–30 W/m$^2$ (Damiani et al., 2018). The same data have been evaluated with high quality data at 36 Chinese stations, the results were improved to an overall bias of 13–14 W/m$^2$ of all the stations for the daily and monthly data but not for the sub hourly data (Shi et al., 2018). Another study has demonstrated the high accuracy of GHI derived from Himawari-8 than MERRA-2 and ERA-Interim

reanalysis datasets at 34 stations over Australia and East Asia. The results are returned to high temporal and spatial resolutions of Himawari-8 images (Yu et al., 2018). These validations also illustrate the importance of SDDs worldwide, which they can be used for various purposes.

### 2.3.4  GHI Derived From Various Meteosat Satellite Images

Some studies have validated GHI from various Meteosat satellite images either by H2 or other methods. For instance, Moradi et al. (2009) estimated daily GHI with the H2 method in Iran, evaluating the result of the model with four stations in the country, which revealed a good agreement with 12% RMSE and 2% bias. Schillings et al. (2004) validated DNI data at weather stations in eight cities in Saudi Arabia with Meteosat-7 data using the H2 method. The results indicate a good agreement with a mean bias of 4.3% from hourly data. Similarly, AL-Jumaily et al. (2010) evaluated the GHI data of two Iraqi weather stations with the same method using Meteosat-8 data for the year 2005. Positive biases of 0.024 KWh/m$^2$ and 0.012 KWh/m$^2$ GHI for daily mean values were found for both cities. The authors indicate that further research comparing Meteosat-8 data with other areas of Iraq is needed. It is necessary for studies to validate more than one SDD for comparison between them, and then select the most accurate.

### 2.3.5  Satellite Application Facility on Climate Monitoring (CM SAF)

The CM SAF provides two SDD products. First, is the CM SAF operational product, which provides daily GHI and DNI and is based on the look-up-table approach, full detail is available from (Mueller et al., 2009). The product is similar to other SDDs, using Meteosat satellite images to calculate cloud properties and other inputs (aerosol, gases, and atmosphere optical depth) to calculate GHI and estimating the other solar components DNI and DHI (Urraca et al., 2017b). The CM SAF daily GHI has been validated by several studies (Sanchez-Lorenzo et al., 2013; Ineichen et al., 2009; Mueller et al., 2009; Urraca et al., 2017b; Amillo et al., 2018).

Second, is SARAH, which provides GHI and DNI at 30 minutes (Amillo et al., 2018), it is considered the same as the Heliosat methods but is implemented differently. SARAH has two versions based on the improvement of the inputs (Urraca et al., 2018b). It has been validated worldwide in the literature. For example, SARAH GHI daily mean has been compared against ground data at 20 stations in Sweden and Norway, and the result

reveals good agreement with an bias of 8% roughly (Riihelä et al., 2015). Similar data have also been validated in 22 cities in Europe with a comparison to other SDDs (Amillo et al., 2014). In the same continent, GHI daily mean from SARAH have been evaluated with 313 stations with the comparison to other GHI products. The results show better performance of SARAH than CM SAF operational and reanalysis ERA-Interim with a bias of 4% and 7% respectively (Urraca et al., 2017b). Similarly, Amillo et al. (2018) have compared the performance of SARAH and CM SAF operational of hourly GHI at 13 stations in South Africa. The overall results show the high accuracy of SARAH (bias 1.4%) compared with CM SAF operational (bias 4.9%). The GHI data from the MFG and MSG by SARAH have also been evaluated over India, which shows an overestimation bias of 10–20% of the daily mean compared to ground observations (Riihelä et al., 2018). Recently, a study has utilised solar irradiance data from SARAH to address and show the spatial variability of solar resources in Kenya (Kariuki and Sato, 2018). The rates for the performance of SDDs in various areas indicate that the SDDs are a reliable source of GHI data and further validation in new regions is required.

### 2.3.6 Solar Energy Mining (SOLEMI)

Solar Energy Mining (SOLEMI) is also a SDD for providing solar irradiance components, that are estimated from geostationary satellite images of Meteosat west (placed at 0 longitudes) and Meteosat east (placed at 63 east over the Indian Ocean). It also applies the H2 method but with a different calculation of the inputs to the method such as Cs model as an input. SOLEMI is run by DLR (Deutsches Zentrum für Luft- und Raumfahrt, [German Aerospace Center]) in Germany (Meyer et al., 2003; Schroedter-Homscheidt et al., 2017). It has been validated over eight cities in Saudi Arabia with a bias of 4% (Meyer et al., 2003).

### 2.3.7 HelioClim-3 (HC3) and CAMS Radiation Service (CRS)

The most popular SDD is hosted by the Solar Radiation Data (SoDa) portal (SoDa, 2018), which contains several projects; one of them is HelioClim-3 version 4 and version 5 (HC3v4-5), which are based on the H2 method for converting satellite images of MSG to GHI. Another is the CRS, which is based on H4 for the same purpose. The two SDDs will be described in (Chapter 5, Section 2.4).

Those SDDs have been validated by several studies in various areas (Table 2.2). For example, Thomas et al. (2016a) have validated the hourly GHI from SDDs such as HC3v4-5 and CRS for 42 stations in Brazil. The result reveals a high correlation (an average of 96%) between HC3v4-5 and ground measurements, whereas that with CRS is lower by 2%. Similarly, $r$ values above 0.92 for 15 min and 0.98 for daily GHI, with a bias of roughly 5% were found when comparing HC3v4-5 and CRS to ground data at 14 stations over the world (Thomas et al., 2016c). Hourly GHI and DNI from HC3v4-5 for all-sky conditions, and using the McClear dataset for clear-sky conditions, have been validated with ground data in seven stations over Egypt, with RMSE ranges from 6%−22% (Eissa et al., 2015a). Marchand et al. (2017) have validated hourly GHI from HC3v4-5 and CRS with ground data at five stations in the United Arab Emirates and Oman. The overall validation result shows bias under 5% and RMSE 11%−16% on average. Similar SDDs have been validated in Morocco with five stations for hourly GHI (Marchand et al., 2018). The bias ranged from −6%−7%, whereas RMSE ranged from 12% to 21% on average. The study also revealed the superior performance of HC3v5 compared to CRS. The CRS has also been validated at ten stations in Croatia with overall $r$ values of 0.967 (Gašparović et al., 2018). This thesis focuses on those two SDDs because they are available in the case study area and they have a high rate of accuracy in several areas. They have not been validated in the study area, and they have high temporal resolution, unlike most other SDDs.

Those SDDs provide data of various regions and areas worldwide and have been validated globally, and the validation results are impressive that those SDDs are crucial for obtaining GHI. Therefore, the first step to use those SDDs in an area is to check their accuracy, which this study will do in Chapter 5. The validation is also crucial with an area that the ground measurements have not used for that purpose.

**Table 2.2:** Validation of GHI between ground measurements and HC3v5 and CAMS in literature.

| Reference | Station – Country | Temporal | Data type | Overall validation results |
|---|---|---|---|---|
| Marchand et al. (2018) | 5- Morocco | Hourly | HC3v4-5 and CRS | Bias= −6% to 7% RMSE 12% to 21% |
| Marchand et al. (2017) | 5 – United Arab Emirates and Oman | Hourly | HC3v4-5 and CRS | Bias less than 5% RMSE 11% to 16% |
| Thomas et al. (2016a) | 42- Brazil | Hourly Daily Monthly | HC3v4-5 and CRS | Bias= 1% to 10% RMAE= 15% to 36% |
| Thomas et al. (2016b) | 14- over world | 15 min Hourly Daily | HC3-4-5 and CRS | Bias= 5% on average |
| Eissa et al. (2015a) | 7- Egypt | Hourly | HC3 | RMSE=15% to 22% |
| Gašparović et al. (2018) | 10- Croatia | Daily GHI | CRS | $r = 0.967$ |
| This thesis Chapter 5 | 9-Iraq | Hourly | HC3v5 and CRS | See (Chapter 5, Section 5.5) |

## 2.4    Modelling Global Horizontal Irradiance

### 2.4.1    Introduction

Ground data are limited in many areas, they have technical and operational issues and SDDs have some rate of bias, while they cover the limitation of spatial resolution in the ground data generally. GHI is also required highly for several fields. Therefore, several studies have tried to estimate and model GHI, based on the limitation and availability of inputs to the model, spatial and temporal resolution and sky conditions and according to purpose of requirement. The approaches and techniques to model GHI are empirical, statistical, physical and soft computing (Chapter 1, Section 2.3). ANN models are machine learning methods which have been used broadly with a variety of purposes. The thesis will also use the ANN models to estimate GHI in Chapter 6. Therefore, ANN models are reviewed comprehensively as follows:

### 2.4.2 Using Satellite Images to Estimate GHI with Artificial Neural Network Models

ANN is a powerful soft computing approach to capture details and information in an image and has been broadly used for image classification, image data analysing, pattern recognition and analysing nonlinear relationships (Behrang et al., 2010; Yadav and Chandel, 2014a). Several studies have used ANN models to analyse satellite images for estimating GHI and other components of solar irradiance. For example, Meteosat-6 satellite images have been analysed to estimate monthly GHI over Turkey (Şenkal and Kuleli, 2009). Similarly, Meteosat-9 satellite images have also been used with extra data in a model in Andalusia, Spain (Linares-Rodriguez et al., 2013). ANNs have also been used with Heliosat-2 for converting multi-spectral MSG images to estimate hourly GHI (Quesada-Ruiz et al., 2015). In addition, images from MTSAT have been analysed and combined with other data in an ANN for predicting GHI (Lu et al., 2011). Other studies have analysed satellite images for obtaining climate variable data (such as land surface temperature), which were then paired with ground measurements in a model to estimate GHI (Şenkal, 2010; Fallahi et al., 2018). Another case is Qin et al. (2011), who inputted monthly precipitation calculated from The Tropical Rainfall Measuring Mission (TRMM) satellite data, with MODIS surface temperature data to model GHI as an output. Those papers have examined the use of ANNs with some climate parameters from satellite images to estimate GHI, but high temporal availability of those parameters is limited.

### 2.4.3 ANNs to Forecast GHI

ANNs have also been used to forecast GHI with various data and over various time intervals. For example, HC3, climate variables and other inputs (Bird Cs model and data of total cloud cover from numerical weather prediction models) have been analysed in ANN models to forecast GHI in intra-day and 1−6 hour ahead in Gran Canary, Spain in two different studies (Aguiar et al., 2016; Aguiar et al., 2015). They demonstrated that the SDD improved the forecasting results. Cloud properties and velocity metric data from satellite images with ground data have been used in an ANN model to forecast GHI at 30, 60, 90 and 120 min time scales (Marquez et al., 2013). Cs model and weather research data have been used to forecast GHI 24 hours ahead with an ANN model (Lima et al., 2016). Hybrid ANN models have also been used to forecast GHI one hour ahead (Crisosto

et al., 2018; Jadidi et al., 2018). Those papers demonstrated the role of ANNs in forecasting GHI at various time scales, and the role of SDDs and Cs as inputs to improve the model results.

### 2.4.4 Estimating Daily and Monthly GHI with ANNs

ANNs are considered to be one of the most powerful algorithms to find relationships between dependent and independent variables. They have been used broadly in literature (Table 2.3) to estimate GHI and other solar irradiance components with different types of data. For instance, geographical and meteorological parameters at different time scales as various inputs have been used with ANN models for a variety of climate regions and countries. For example, two cities in Oman (Al-Alawi and Al-Hinai, 1998), eight cities (Jiang, 2009) and nine cities  (Feng et al., 2018b) in China, one city (Chukwu, 2012) and 195 cities (Fadare, 2009)  in Nigeria, 27 stations (Ozgoren et al., 2012), seven cities (Kisi, 2014) and 30 cities (Kaba et al., 2018) in Turkey. In addition, they have been applied to five stations in Argentina (Jimenez et al., 2016), six cities in the Yucatan peninsula, Mexico (Quej et al., 2017), five cities in Italy (Renno et al., 2016), four cities in the USA and two cities (Sharifi et al., 2016), 19 cities (Kheradmanda et al., 2016), 10 cities (Jahani and Mohammadi, 2018) and 12 cities in Iran (Khosravi et al., 2018b) and Cairo city in Egypt (Hassan et al., 2017a). Generally, the results of those models in the literature show good agreement with ground data based on statistical indicators (Table 2.3). This indicates the importance of various types of ANN models and algorithms for estimating GHI. However, those studies mainly focused on daily time scales and in some cases on monthly time scales, but not at an hourly time scale, which is the focus of this study.

**Table 2.3:** The literature for estimating GHI with ANN models on daily and monthly time scales.

| Reference | Inputs | Temporal | Country | ANN type | Overall results |
|---|---|---|---|---|---|
| Kaba et al. (2018) | Latitude, longitude, TOA, SD, clod cover, min and max of AT. | Daily at 30 station | Turkey | Deep learning | $r$=0.980 |
| Jahani and Mohammadi (2018) | SD and min and max of AT. | Daily at ten stations | Iran | MLP coupled with genetic algorithm | $R^2$=0.92 |
| Khosravi et al. (2018b) | Latitude, Longitude, month, day, average, max and min of AT, air pressure, RH, WS and TOA. | Daily at 12 stains | Iran | multilayer feed-forward neural network | $R^2$=0.988 |

| Reference | Inputs | Temporal | Country | ANN type | Overall results |
|-----------|--------|----------|---------|----------|-----------------|
| Hassan et al. (2017a) | Day number, TOA, SD, AT, RH, air pressure and WS. | Daily at one city | Egypt | MLP | $R^2$=0.947 |
| Quej et al. (2017) | TOA, min and max AT and precipitation. | Daily at six cities | Mexico | feed forward neural network | $R^2$=0.545 to 0.705 |
| Sharifi et al. (2016) | Kt, TOA, AT and daylight hour. | Daily at two station | Iran | feed forward neural network | RMSE= 0.078–0.115 MJ/m$^{-2}$ day$^{-1}$ |
| Renno et al. (2016) | Latitude, longitude, daylight hour, SD, AT, precipitation and declination angle. | Daily at five station | Italy | MLP | MAPE = 4.57% |
| Jimenez et al. (2016) | Solar zenith angel, GHI, RH and min, max, average of AT. | Daily GHI for 5 stations | Argentina | feed forward - back propagation | RMSE 2.83% |
| Kisi (2014) | Latitude, longitude, altitude and month of the year. | Monthly at seven stations | Turkey | - | MAE ¼ 4.69 MJ/m$^2$ |
| Ozgoren et al. (2012) | Latitude, longitude, Altitude, month, min, max, average of AT, soil temperature, RH, WS, pressure, vapour pressure, cloudiness and SD. | Daily at 27 stations | Turkey | MNLR multilayer non-linear regression with ten models of ANN | MAPE 5.34% $R^2$ 0.993 |
| Kheradmanda et al. (2016) | Latitude, longitude, Altitude, AT, precipitation, RH, and SD. | Monthly at nine stations | Iran | Multilayer Perceptron | testing data MAPE 4.33% |
| Chukwu (2012) | SD, max AT, cloud cover and RH. | Monthly average at a city | Nigeria | MLP | r= 0.99 MPE 0.851 RMSE 0.003 |
| Fadare (2009) | Latitude, longitude, Altitude, month, AT, , RH, and SD. | Monthly at 195 stations | Nigeria | Standard multilayer feed- forward back propagation | r= 0.956 |
| Jiang (2009) | Latitude, longitude, Altitude, SD. | Monthly at three stations | China | feed- forward back propagation with 6 empirical models | MPE 4.65% |
| Al-Alawi and Al-Hinai (1998) | Location, month, air pressure, AT, vapour, RH, WS and SD. | Monthly for two cities | Oman | multilayer feed forward one model | MAPE 5.43% General accuracy is 95% |

## 2.4.5 Estimating GHI on Hourly Time Scales with ANNs

After an extensive review of recent literature, only four studies have been found that have used ANNs to estimate GHI at an hourly time step. The following studies each

focused on one city in Algeria (Dahmani et al., 2016), Malaysia (Ibrahim and Khatib, 2017) and Morocco (Loutfi et al., 2017), and on five cities in North Africa and Jordan (Hassan et al., 2017b). They are fully described and compared to this study in Chapter 6 (Table 2.4). On the other hand, other studies have estimated direct normal irradiance (Lopez et al., 2005; Renno et al., 2016), DHI (Alam et al., 2009; Soares et al., 2004) and have forecasted GHI, as mentioned in the previous paragraphs using ANN models on hourly time scales.

These studies (Table 2.4) also used other machine learning approaches with ANNs, estimated other solar components and estimated GHI at other time steps including daily and monthly, whereas the descriptions in the table are focused on the ANNs for estimating hourly GHI.

It is clear from the literature that studies using SDDs and combining them with observed climate variables and with TOA and Cs as several new input combinations in an ANN model to estimate hourly GHI, are quite limited.

**Table 2.4:** The literature for estimating GHI with ANN models on hourly time scales.

| Ref | Inputs | Neurons in the Hidden Layer | Training algorithm | Best RMSE |
|---|---|---|---|---|
| (Dahmani et al., 2016) | SD, AT, RH, WS, TOA, precipitation, Pressure, Declination, Zenith angle and Wind direction | 1–8 By 1 or 2 | feed-forward back propagation Levenberg-Marquardt | 13.3% |
| (Ibrahim and Khatib, 2017) | AT, RH, sunshine ratio, Day number, Month number, Number of an hour per day | 3 and 6 | Firefly algorithm | 18.9% |
| (Loutfi et al., 2017) | SD, AT, RH, WS, Declination angle, GHI daily, daylight hours, TOA, sunshine fraction | 10, 15 and 20 | feed-forward Levenberg- Marquardt | 13.1% |
| (Hassan et al., 2017b) | TOA, Solar time and Day number | 100, 180, 210 and 300 | - | 17% |
| This study | SD, At, RH, WS, Cs, TOA, HC3v5 and CRSv3 | 20–140 by 10 | feed-forward backpropagation Levenberg-Marquardt | See (Chapter 6, Section 6.5) |

## 2.5    Gaps in the Literature

After the review in the most relevant literature, the below gaps are identified, which need to be addressed and in new studies:

• There is a crucial need to demonstrate the difference between two broadly applied QC tests (TOACs and BSRN) for GHI data. This is because the tests have the same target but with various criteria, with which they test the GHI data for physical possible limits and extremely rare limits.

• To date, there are limited studies that identify the role of SD to test the quality of GHI data by using the upper and lower bounds of SD (high and low value of SD in a time interval such as 50 min and 0 min sunshine in one hour).

• To date, it is hard to find studies which have used the AT variable to check the quality of GHI.

• Some studies have used the lower bound of SD to test the GHI data, whereas using SD and AT to QC GHI have never been attempted.

• Very limited studies have utilised fingerprint plots and general plots by plotting SD, AT and GHI in one graph for a general QC and to detect any questionable data.

• Several studies have validated SDDs in various areas and regions. Therefore, there is no existing study to compare and validate SDDs (HC3v5 and CRSv3) in Iraq.

• There are no studies that investigate and evaluate two other pixel point data of SDD against GHI ground measurements.

• Up to now, a few studies have used SDDs to forecast GHI, whilst studies using SDDs with climate variables to estimate GHI were not found.

• Using the new combination of inputs from SDDs, climate variables, TOA and Cs to estimate hourly GHI was rare.

## 2.6 Aims of the Research

The overall aim of this study was to obtain good QC of hourly GHI data in an area with limited ground measurements and rare GHI modelling. This is by developing new QC tests, evaluating and comparing two SDDs for GHI data and correcting bias for SDDs under cloudy-sky conditions and modelling hourly GHI with new input combinations in ANN models.

## 2.7 Research Objectives

The following objectives are formed based on gaps in the literature and the aims of the study.

1. To determine the differences between (TOACs and BSRN) the QC set of tests for error detection in hourly GHI data, to develop a local test for error detection based on that difference and to evaluate the role of the local test.

2. To develop a new QC method for GHI by using the SD and AT variables and to evaluate it, if it works as a consistency test and to detect systematic errors and questionable data by graphics.

3. To validate the SDDs in new areas and to investigate the spatiotemporal features of SDDs to ground data, and to investigate the difference of two pixel point data from SDDs around a station compared with ground data of a station.

4. To evaluate the ability of a simple method for bias correction in SDDs under cloudy sky conditions.

5. To assess the role of new input combinations (SDDs. Cs, TOA and climate variables) in estimating and modelling hourly GHI data.

Chapter 3 :   **Study Area and Evaluation Criteria**

## 3.1 Study area

### 3.1.1 Location

The study area is located in the Kurdistan Region of northeast Iraq (latitudes [34°08'20"–37°22'36"], and longitudes [42°32'00"–46°14'29"]). The bordering countries in the east and the north are Iran and Turkey whereas to the west and south are other parts of Iraq (Figures 3.1 and 3.2).
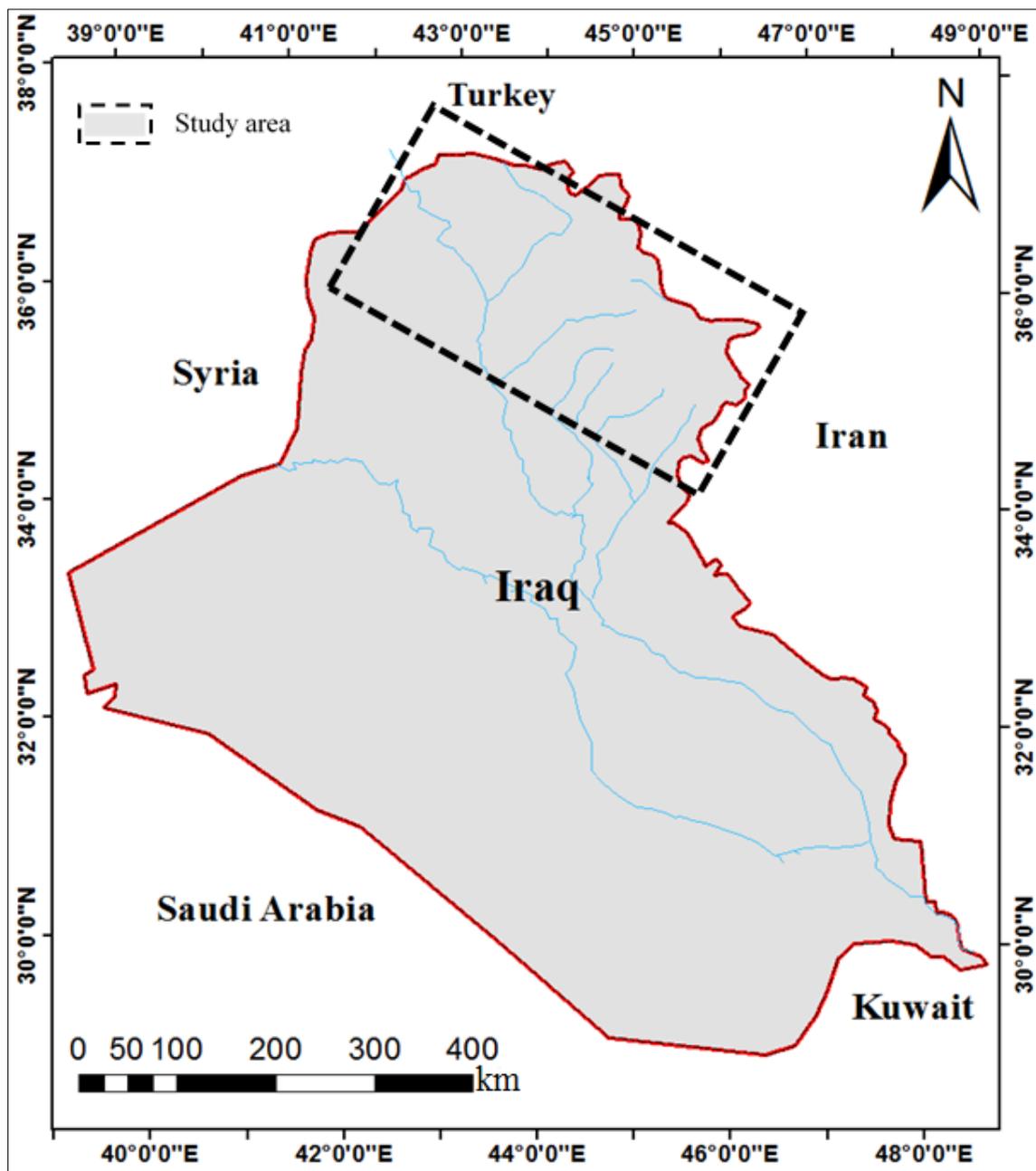


**Figure 3.1:** The case study location in Iraq with bordering countries.

Most of the stations are located in mountainous areas but a few of them are located in the plains (Figure 3.2). The twenty meteorological weather stations selected in that area (Figure 3.3), were used in Chapter 4 with six other stations worldwide for validation purpose of the new QC method. Hence, the meteorological stations were reduced to nine in Chapter 5 with using SDDs data in a station location and with two SDD data points around each station (Figure 3.2) and these same nine stations are used in Chapter 6.
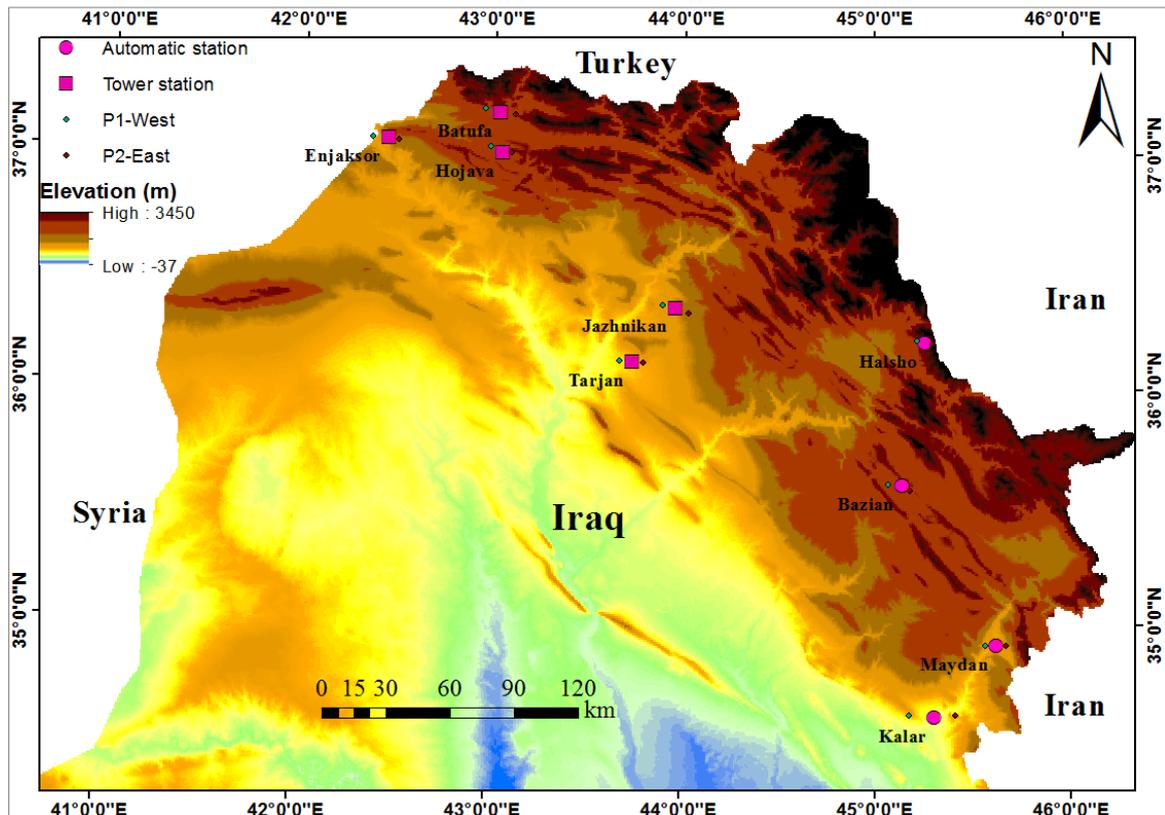


**Figure 3.2:** The topography of the study area with station locations and point locations around the stations used in Chapter 5 (USGS, 2018).

### 3.1.2 Climate

Three climate types according to the Köppen classification are seen in the study area where the selected stations are located (Figure 3.3). The climate types are hot semi-arid (BWh), warm-summer (Csb) and hot-summer (Csa) of Mediterranean Sea climate types (Kottek et al., 2006; Rasul, 2016). In general, the area has hot, warm dry summers and mild, cold wet winters, and spring and autumn are mild (Najmaddin, 2017). Climate variables are the outcome of the sun's irradiance with interactions of the atmosphere and the earth surface, and some of the variables were utilised in this study. Most of the climate

variables are described in detail; this is because there is no recorded data about other variables, which affect the GHI such as aerosol, cloud types and atmosphere contents. The monthly mean of hourly data of each variable are described to show a general description of each variable, this study utilises hourly data when the sun elevation angle is above 15°. The averages of data in each month were calculated from hourly data for each nine station. All the data presented in Figures 3.4–3.8 are calculated based on the availability of hourly data at nine stations for no more than four years of recorded data (Figure 3.2), some variables are lacking in certain years in some stations. Therefore, the figures are shown the averaged data from all nine stations based on the availability of climate variables at a station, except that GHI is shown for each station individually. For instance, climate variables (GHI, AT, SD, RH and WS) were recorded at four automatic stations (Halsho. Bazian, Maydan and Kalar), whereas the SD variable was not recorded at five tower stations (Batufa, Enjaksor, Jazhnikan, Hojava and Tarjan), which is the main difference between the automatic and the tower stations.
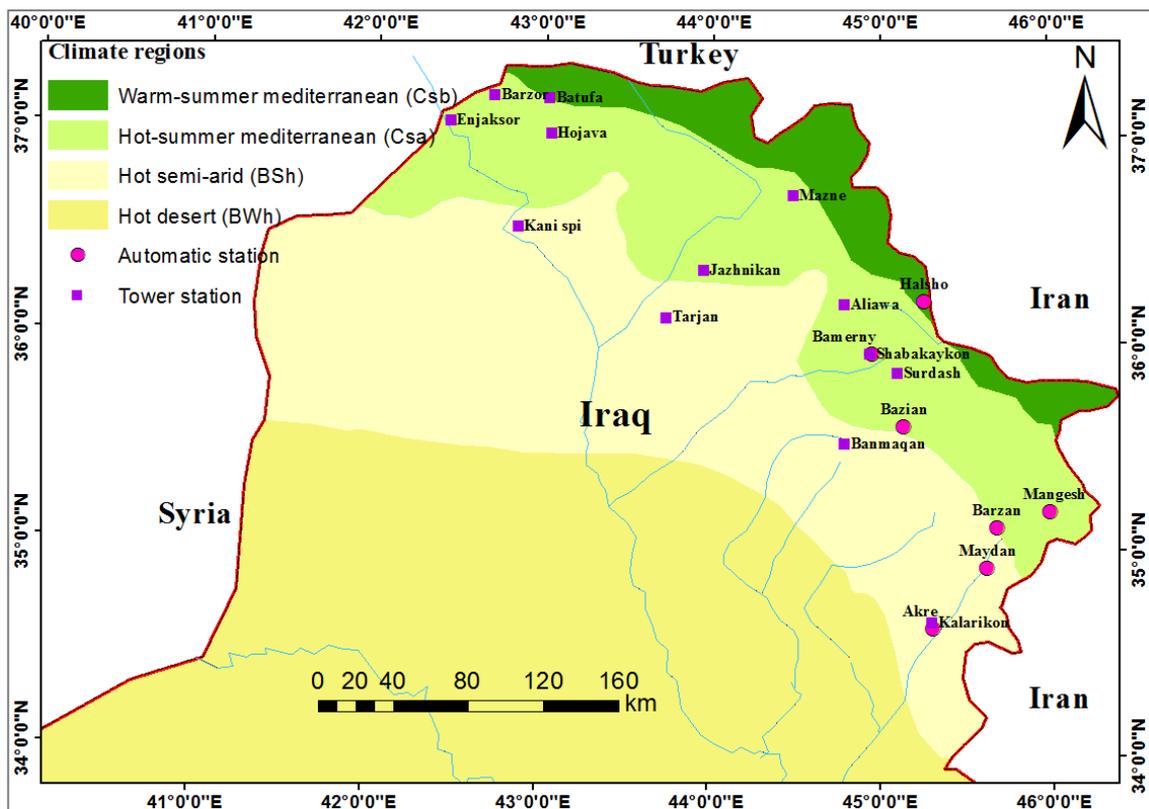


**Figure 3.3:** The case study area with Köppen climate regions and distribution of the stations.

### 3.1.2.1  GHI

The general description of monthly mean of hourly GHI and its value in W/m² on the
year in the study area can be seen by a close look at Figure (3.4). The annual mean value
of hourly GHI is between 450–500 W/m² among the stations. The GHI is low in winter
months and increases slightly to reach its peak in summer. Daytime lengthens and is
cloud–free sky in summer whilst daytime is shorter with cloudy and rainy skies in winter.
There is no vast difference from station to another, whereas some minor differences
between them are related to the local conditions.
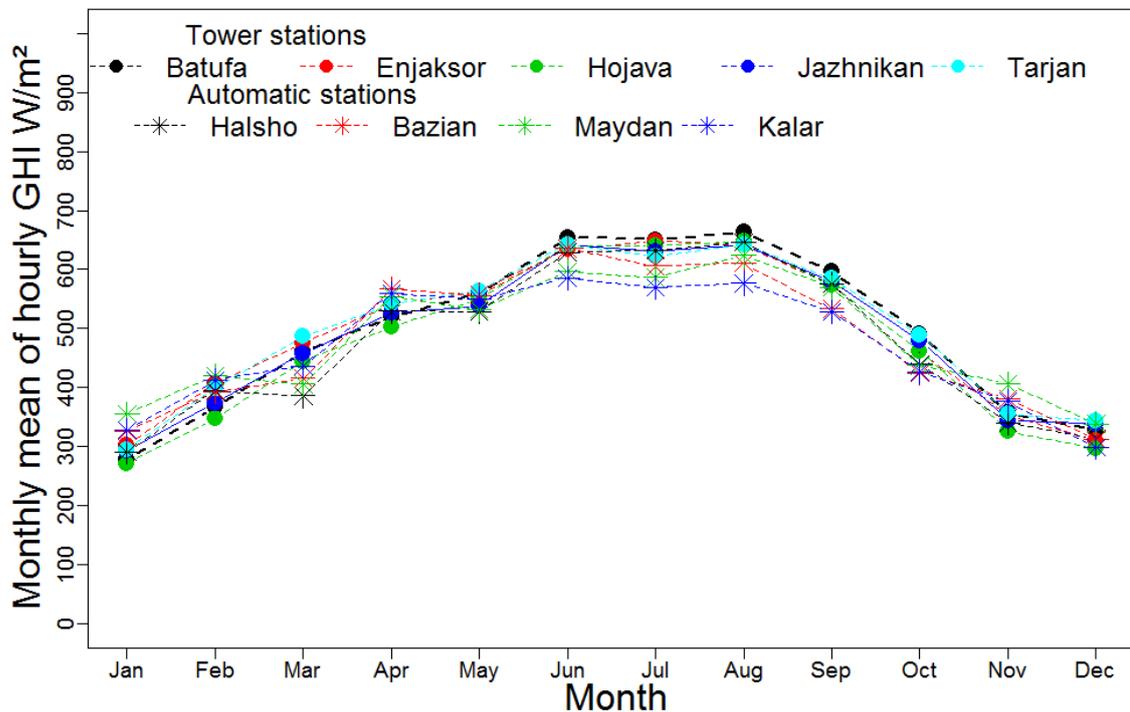


**Figure 3.4:** Mean of GHI at each station.

### 3.1.2.2  Sunshine Duration (SD) and Cloud Cover

There is a direct relationship between SD and GHI. This is obvious in Figure 3.5,
which shows the mean SD and GHI in each month. The SD is considered one of the most
vital variables for estimating GHI and it is used in literature broadly (Almorox and
Hontoria, 2004; Ampratwum and Dorvlo, 1999; Brabec et al., 2016; De Souza et al.,
2016; Fan et al., 2018; Fan et al., 2019; Nguyen and Pryor, 1997; Suehrcke et al., 2013).
Peak records of SD are in summer months, which near 13 hours and low in winter months,
nearing 7 hours. Figure 3.5 shows the mean SD at four automatic stations to describe SD

and compare it with GHI in the study area because the SD is recorded in those stations, unlike others.

SD records the direct irradiance, which describes the sky as cloudy or cloud free, which then gives a general idea about GHI. However, the type of clouds vary from thin and bright to accumulated and dark and have a clear effect on GHI (Betts et al., 2001). When the clouds are thin and high the diffuse irradiance is low, and vice versa for dark clouds. This information about clouds are unavailable in weather stations, but generally, dark accumulated clouds are common in winter, high and thin in spring and in autumn and summer, the sky is generally free from clouds. The various types of clouds affect the GHI, which reaches the earth's surface.
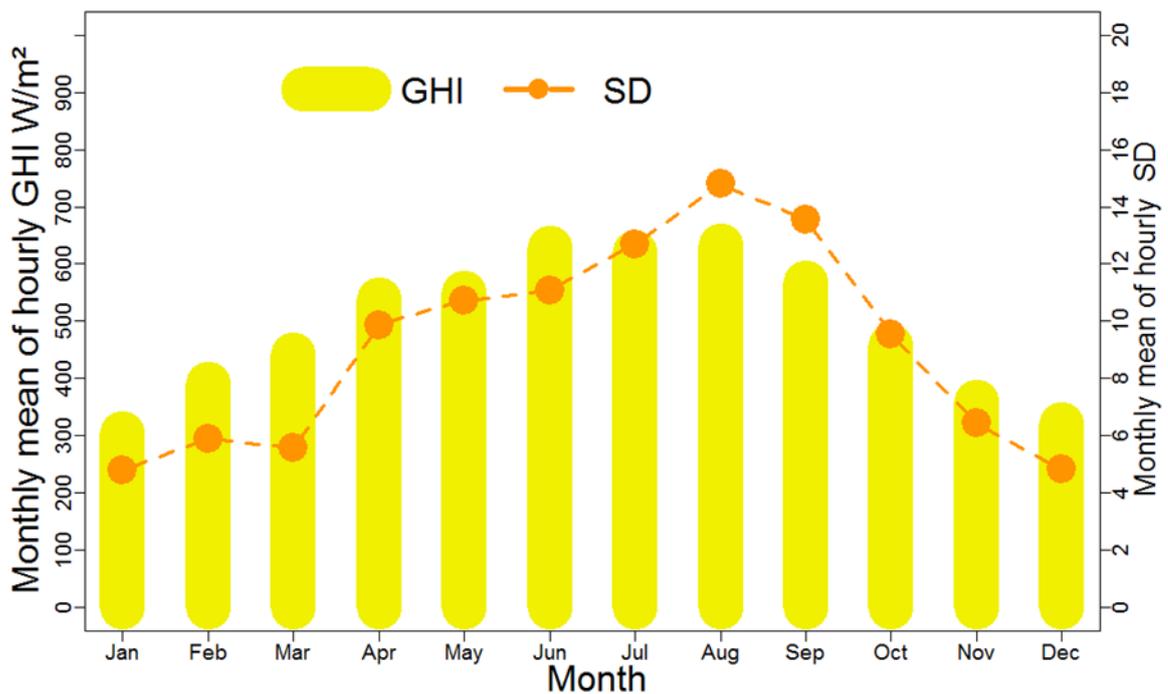


**Figure 3.5:** Mean SD at four automatic stations (Halsho, Bazian, Maydan and Kalar) and GHI for the nine stations in the study area as in Figure 3.4.

### 3.1.2.3  Air Temperature (AT)

AT is dependent on GHI but is affected by other factors. This can be noted in the comparison of AT to GHI in each month in Figure 3.6. AT is seasonally more variable than GHI (Figure 3.6). This is related to the late response of AT to the change of GHI owing to the process of conduction and convection after irradiance. Similar to SD and GHI, the overall lowest AT values are recorded in winter (8–15 °C) and the highest are

recorded in summer (32–39 °C), with moderate AT (15–25 °C) recorded in spring and autumn (Figure 3.6).
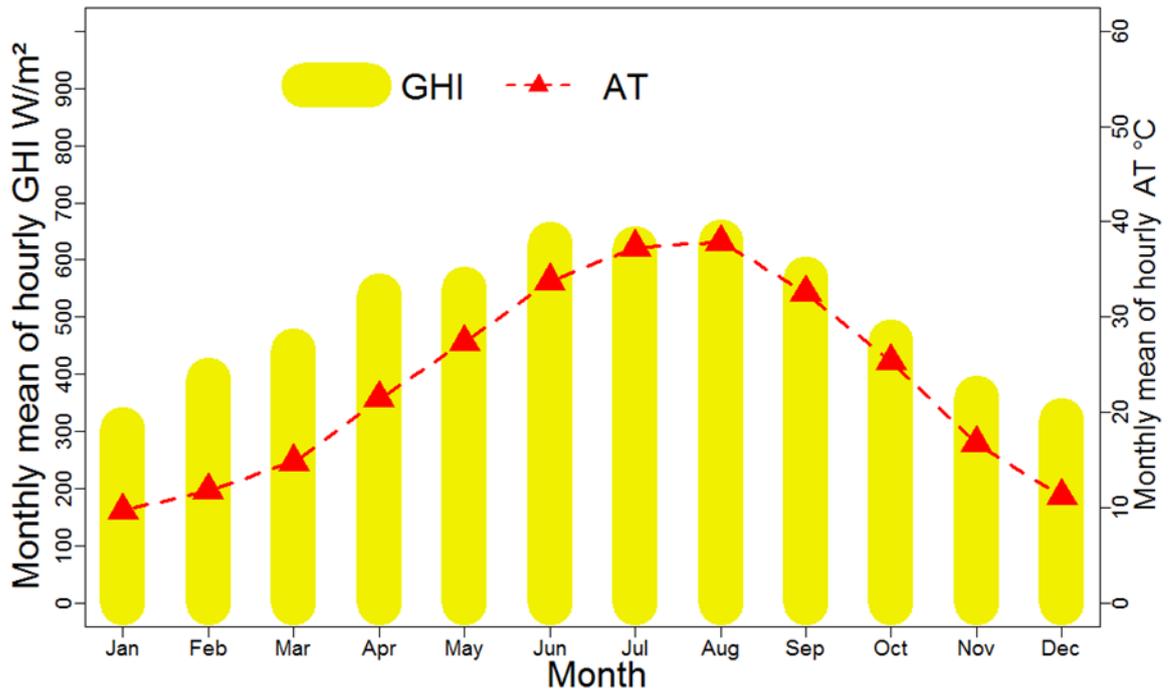


**Figure 3.6:** Mean AT and GHI for the nine stations in the study area as in Figure 3.4.

### 3.1.2.4   Precipitation

Precipitation in the study area falls mostly as rain with a few days of snow annually. Snow is higher in the mountain areas (stations; Halsho, Batufa and Hojava) compared to plain areas (stations; Maydan, Kalar and Tarjan). A similar situation is also true for rainfall. Precipitation monthly falls from October to May with high fluctuation from year to year. The annual average varies from 250–400 mm in the south east in low elevation land and increases in the north east in high elevation land to 700–900 mm year based on data from 1960–2015 (GDMS, 2017). The rain has two major effects on GHI. First, in rainy periods the rain clouds prevent the direct irradiance from reaching the earth's surface. This is limited to the rain time and this factor has a connection with the cloud types. Second, a positive side of rain is that it cleans the pyranometers in the rainy season. Regarding snow, this covers the dome of the pyranometers, but in the chosen stations this does not affect more than about ten days annually. However, in some stations snow occurrence is lower than 1 day in 20 years on average (GDMS, 2017).

### 3.1.2.5 Relative Humidity (RH)

There is a general relationship between RH and GHI (Figure 3.7). This reveals that the RH is considered one of the climate variables for modelling GHI. In addition, studies have demonstrated the effect of RH on estimating GHI (Soulayman, 2017). The high mean values of RH are recorded in winter, and low values are recorded in summer (Figure 3.7)
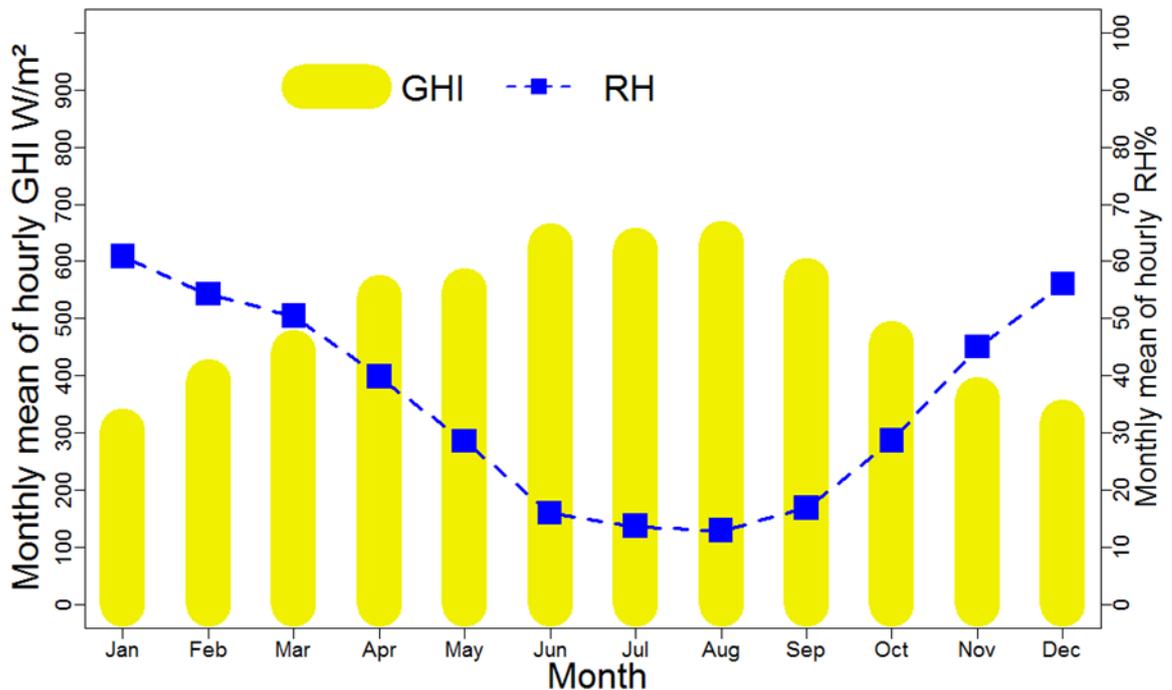


**Figure 3.7:** Mean RH and GHI for the nine stations in the study area as in Figure 3.4.

### 3.1.2.6 Wind Speed (WS)

The wind in the case study varies significantly locally. For example, the mountains behind the plain area has different air pressure, which causes local winds with various speeds from day to night and season to season (Ahmed and Mahammed, 2012). However, Figure 3.8 shows a stability of WS over the course of the year with a mean of nearly 4 m/s at all the nine stations. Some minimal fluctuations can be noted from season to season in Figure 3.8. The wind has a nonlinear relationship with GHI because the difference in GHI is one of the factors to move the wind indirectly. Figure 3.8 shows the WS at 10

meter elevation, whereas the WS at high elevations has a role in moving the clouds thus affecting the GHI. A study has shown that the wind stills aerosols which then affects the variability of GHI (Lin et al., 2015). The five to ten day average of dusty winds have been recorded in stations Kalar, Bazian and Maydan in the years 2003–2016 (GDMS, 2017). The dust causes a decrease in GHI, dust remaining on the pyranometers would affect the measurements if they are not cleaned directly.
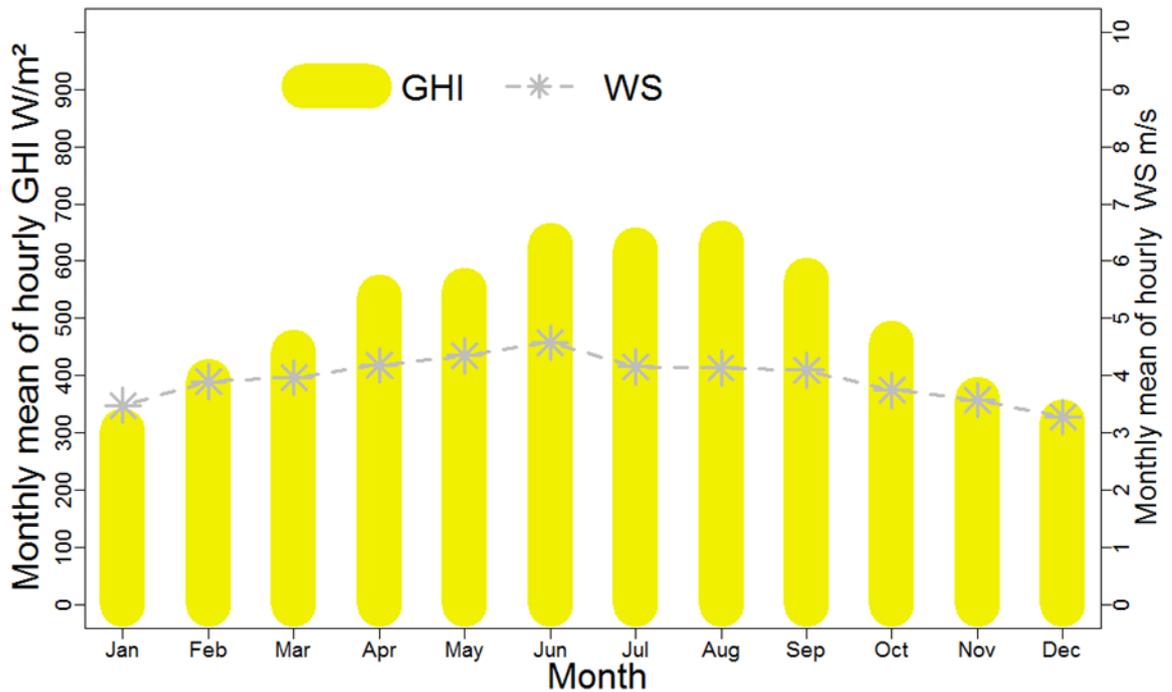


**Figure 3.8:** Mean WS and GHI for the nine stations in the study area as in Figure 3.4.

## 3.2    Evaluation Criteria

The validation performance between ground data and SDDs, namely HC3v5 and CRSv3 (Chapter 5), and between ground data with the ANN and regression model predictions for training, validation and test data (Chapter 6) have been evaluated using statistical indicators. These were correlation coefficient ($r$) in Equation (3.1), the bias in Equation (3.2), the relative bias (bias %) in Equation (3.3), the root mean square error (RMSE) in Equation (3.4), and the relative RMSE (RMSE %) in Equation (3.5) (Quej et al., 2017; Thomas et al., 2016b).

44

$$r = \frac{\sum_{i=1}^{n}(Xi - \bar{X})(Yi - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(Xi - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Yi - \bar{Y})^2}} \qquad (3.1)$$

$$Bias = \frac{\sum_{i=1}^{n}(Yi - Xi)}{n} \qquad (3.2)$$

$$relative\ bias = \frac{Bias}{Mean\ Xi} * 100 \qquad (3.3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Yi - Xi)^2}{n}} \qquad (3.4)$$

$$relative\ RMSE = \frac{RMSE}{Mean\ Xi} * 100 \qquad (3.5)$$

where $n$ = the number of observations, $Xi$ = the GHI of ground data and $Yi$ = the GHI of an SDD (Chapter 5), and Yi = the estimated GHI (Chapter 6). $\bar{X}$ and $\bar{Y}$ are the average of $Xi$ and $Yi$ respectively.

Chapter 4 :   **Quality Control of Global Horizontal Irradiance Estimates through BSRN, TOACs and Air Temperature/Sunshine Duration Test Procedures**

## 4.1   Introduction

This chapter addresses objectives:

1.  To determine the differences between (TOACs and BSRN) the QC set of tests for error detection in hourly GHI data, to develop a local test for error detection based on that difference and to evaluate the role of the local test.

2. To develop a new QC method for GHI by using the SD and AT variables and to evaluate it, if it works as a consistency test and to detect systematic errors and questionable data by graphics.

Pyranometers for recording solar irradiance data face several equipment and operational errors. Several studies have provided QC approaches, whereas most of them cannot be applied when only GHI is available and the comparison between different QC approaches is limited in the literature.

It is clear that further research into the QC of GHI observations is required. The aim of this chapter was, therefore, to compare and evaluate the results of two sets of tests (TOACs and BSRN) for hourly GHI data over 20 stations where data had not been quality assured and tested. The analysis assessed the reliability of each test where the criteria in each test are different, but they check the GHI data for the physically possible limit and extremely rare limit errors. The limits of the QC tests (BSRN and TOACs) were modified based on the local climate condition in the study area, and the new boundaries were named as a local test. This is to demonstrate the role of the local test compared to BSRN and TOACs QC tests. This study in this chapter also uses a simple new AT and SD tests, which are based on the relationship between GHI, AT and SD. This test is useful for stations that do not record diffuse and direct irradiance. It was validated with high quality data in which all of solar irradiance components were available at six stations worldwide with the same climate types as the study area. Finally, SD and AT are combined to

enhance the results and to detect the errors in those variables (SD and AT) too rather than GHI.

## 4.2    Materials and Methods

### 4.2.1    Dataset

#### 4.2.1.1    Study Sites

The study area is northeast Iraq, which was described in Chapter 3; Sections 3.1, where twenty stations have been selected. In addition, six stations worldwide, three from Australia and one from each of France, Brazil and Israel were used in this chapter for validation of the AT and SD tests.

#### 4.2.1.2    Ground Measurements

Hourly data of GHI (W/m$^2$), SD in minutes per hour and average AT (°C) were collected from seven automatic and thirteen tower meteorological stations in the Kurdistan region of northeast Iraq. SD is not recorded at the tower stations. Openly available one minute data of all solar irradiance components with SD from three Australian (BOM, 2018) and one minute data of all solar irradiance components with AT from three BSRN (König-Langlo et al., 2013) stations from France, Brazil and Israel were collected for validation purposes. The climate regions according to the Köppen classification and the station locations in the study area are shown in Figure (3.3). Tables (4.1, 4.2, and 4.3) show the geographical information, pyranometer types and timescales for each station. The data acquisition times were selected to be between sunrise and sunset when the sun elevation angle is above 15°. This is to avoid a high rate of errors due to the cosine effect for lower sun angles and the AT test when the elevation angle is low (Younes et al., 2005; Journée and Bertrand, 2011a; Pashiardis and Kalogirou, 2016), although some researchers suggested a 7° sun elevation angle (Lemos et al., 2017). The timescale of the data varies between the stations.

#### 4.2.1.3    Calculated Data

The hourly data for *Cs* from the McClear model and TOA hourly irradiance (W/m$^2$) were collected at SoDa (2018) for all stations as periods in the Tables (4.1, 4.2, and 4.3). Minute data of solar zenith angle for all station as periods in the Tables (4.1, 4.2, and 4.3)

were collected at NREL (2017). The minute data was aggregated to hourly data and the cosine solar zenith angle for each hour was also calculated.

**Table 4.1:** Description of processed data of hourly GHI measured by a CMP6 Kipp and Zonen Pyranometer with climate variables and calculated variables for the tower stations.

| Station | Latitude | Longitude | Elevation a.s.l (m) | Period (dd/mm/yy) | Number of data |
|---------|----------|-----------|---------------------|-------------------|----------------|
| Barzor | 37.1881 N | 42.6950 E | 509 | 01/07/2010 – 31/08/2011 | 4119 |
| Batufa | 37.1764 N | 43.0236 E | 947 | 01/01/2011 – 31/12/2013 | 10320 |
| Enjaksor | 37.0603 N | 42.4353 E | 509 | 01/01/2011 – 31/12/2014 | 13757 |
| Hojava | 37.0075 N | 43.0369 E | 933 | 01/01/2011 – 31/12/2013 | 10320 |
| Mazne | 36.7183 N | 44.4814 E | 677 | 01/06/2010 – 30/09/2011 | 4749 |
| Kani spi | 36.5556 N | 42.8483 E | 334 | 01/01/2011 – 31/12/2014 | 13757 |
| Jazhnikan | 36.3564 N | 43.9556 E | 430 | 01/01/2011 – 31/10/2013 | 9893 |
| Aliawa | 36.1933 N | 44.7908 E | 535 | 01/07/2010 – 30/11/2011 | 4877 |
| Tarjan | 36.1258 N | 43.7353 E | 276 | 01/01/2011 – 31/12/2013 | 10320 |
| Shabakaykon | 35.9536 N | 44.9422 E | 602 | 01/07/2010 – 30/11/2011 | 4877 |
| Surdash | 35.8625 N | 45.1036 E | 1040 | 01/01/2013 – 31/12/2013 | 3437 |
| Banmaqan | 35.5197 N | 44.7903 E | 887 | 01/06/2010 – 31/12/2010 | 1987 |
| Kalarikon | 34.6547 N | 45.3019 E | 254 | 01/06/2010 – 31/03/2011 | 2766 |

**Table 4.2:** Description of processed data of hourly GHI measured by a QMS101 Vaisala Pyranometer with climate variables and calculated variables for the automatic stations.

| Station | Latitude | Longitude | Elevation a.s.l (m) | Period (dd/mm/yy) | Number of data |
|---------|----------|-----------|---------------------|-------------------|----------------|
| Halsho | 36.2097 N | 45.2598 E | 1105 | 01/01/2013 – 31/12/2016 | 13757 |
| Dukan | 35.9541 N | 44.9505 E | 555 | 01/01/2015 – 26/09/2016 | 6138 |
| Bazian | 35.6021 N | 45.1376 E | 892 | 01/04/2014 – 30/12/2016 | 9534 |
| Halabja | 35.1889 N | 45.9928 E | 695 | 01/01/2013 – 31/12/2016 | 13757 |
| Darband | 35.1131 N | 45.6854 E | 513 | 01/01/2015 – 31/12/2016 | 6883 |
| Maydan | 34.9194 N | 45.6224 E | 330 | 01/01/2014 – 31/12/2016 | 10320 |
| Kalar | 34.6244 N | 45.3049 E | 218 | 01/01/2014 – 31/12/2016 | 10320 |

**Table 4.3:** Description of processed data of hourly GHI, DNI and DHI measured by Kipp and Zonen equipment with high quality and climate variables for six stations elsewhere in the world for validation.

| Station-Country | Latitude | Longitude | Elevation a.s.l (m) | Period (dd/mm/yy) | Number of data | Used for test | Köppen climate type* |
|---|---|---|---|---|---|---|---|
| Carpentras –France | 44.083 N | 5.059 E | 100 | 01/01/2015 –31/12/2016 | 6366 | AT | Csb |
| Sede Boqer - Israel | 30.905 N | 34.782 E | 477 | 01/01/2010 –31/12/2011 | 6899 | AT | BWh |
| Petrolina-Brazil | 34.624 S | 45.304 W | 387 | 01/01/2013 –31/12/2015 | 7318 | AT | Bsh |
| Geraldton - Australia | 28.795 S | 114.697 E | 33 | 01/01/2004 – 31/12/2005 | 5148 | SD | Bsh |
| Longreach - Australia | 23.4397S | 144.282 E | 192 | 01/01/2013 – 31/12/2013 | 3027 | SD | BSh |
| Broome – Australia | 17.9475S | 122.235 E | 7.4 | 01/01/2015 –31/12/2016 | 5778 | SD | BSh |

* Warm-summer Mediterranean (Csb), hot desert (BWh), hot semi-arid (BSh).

### 4.2.2   Method

The time series of hourly data for each station were plotted for variables GHI, AT and SD together for daytime recordings with sun elevation angle above 15°, and GHI alone for day and night data, in fingerprint plots, in which the x-axis represents a day in the year and the y-axis represents an hour in the day. These plots demonstrate a GHI value for the times with a colour scale from blue to red (R codes for creating a fingerprint plot are provided in Appendix A, Table A1). This is to check for any major problems with the data before testing every single observation. Secondly, in order to obtain high quality GHI data, the methods in the following sections were implemented and illustrated in Figure 4.1. The full QC method with R codes is provided in Appendix A (Table A2)

### 4.2.2.1   Missing value (NA) detection

Detecting missing hourly values in the time series and setting them to not applicable (NA) is essential to show missing observations which can later be used for different purposes, such as comparing two observations in the time series or comparing ground data with satellite data to avoid inappropriate comparisons (Ineichen, 2013; Schwandt et al., 2014; Pashiardis and Kalogirou, 2016).  All hourly time series data were checked

automatically for any gaps or unreliable filled values such as 999 and (///), and they were set to NA. The one-minute data of BSRN and Australian stations were aggregated to hourly data. For this aggregation, first, the time series was be gap-filled. The whole hour's data was set to NA for any hour in which there were missing data.

### 4.2.2.2 Comparison between BSRN and TOACs Tests

GHI data were checked for the physically possible limits for minimum and maximum observations using two tests. The first subtest uses the Baseline Surface Radiation Network (BSRN) subtest with two targets.

$$-4 \text{ W/m}^2 < \text{GHI}_{GD} < (\text{So/SE}^2)*(1.5(\text{Cos}\theta)^{1.2}) + 100 \text{ W/m}^2 \qquad (4.1)$$

With:

$_{GD}$: Ground Data

The tests were applied in several case studies (Chapter 2, Section 2.2). These can be obtained from a number of sources such as at (NREL, 2017). The first subtest of BSRN as Equation (4.1) is compared to the first subtest of TOACs Equation (4.2) for the same two targets, as shown below:

The second condition of Equation 4.1 (flag 2) for the BSRN subtest is compared to the second condition of Equation 4.2 (flag 3) for the TOACs subtest to detect the upper physically possible limits. The first condition of Equation (4.1) (flag 4) for the BSRN subtest is compared to the first condition of Equation 4.2 (flag 5) for the TOACs subtest to detect lower physically possible limits (Table 4.4). See Figure 4.1 for clarification.

$$0.03*\text{TOA} < \text{GHI}_{GD} < \text{TOA} \qquad (4.2)$$

The TOACs tests are described in detail in Geiger et al. (2002). The test requires the TOA, which is available from sources, namely SoDa (2018) and NREL (2017). It is calculated generally for any location and time by:

$$\text{TOA} = (\text{So/SE}^2)*(\text{Cos}\theta) \qquad (4.3)$$

Another comparison between BSRN and TOACs in other subtests is checking to detect extremely rare limit observations in the data; their borders for detection are lower than the previous subtests in Equations 4.1 and 4.2. The first one is related to the BSRN second subtest with the same requirements as Equation 4.1, and is calculated by Equation

(4.4). The target is the same for both BSRN and TOACs second subtests. The second subtest of BSRN has compared the condition of Equation 4.4 (flag 6) to the condition of Equation 4.5 (flag 7), which represents the second subtest of the TOACs. See Figure 4.1 for clarification.

$$GHI_{GD} < (So/SE^2)*(1.2(Cos\theta)^{1.2}) + 50 \ W/m^2 \qquad (4.4)$$

$$GHI_{GD} < 1.1*Cs \qquad (4.5)$$

The TOACs second subtest is based on comparing the ground data with 110% of the Cs irradiance value (Geiger et al., 2002). In reality, the ground data should be lower than the result of Cs irradiance (Lefèvre et al., 2013; Lemos et al., 2017). However, if it is higher than the Cs irradiance, the data value should be flagged for further checks. There are a number of models for estimating Cs irradiance (Reno et al., 2012). One simple model is based on daylight time by hours with a constant coefficient and TOA (FAO, 2017). Another model uses the Linke turbidity factor to demonstrate the clarity of the sky (Rigollier et al., 2000). For more detail about Cs irradiance, the reader is referred to Reno et al. (2012) and Chapter 1, Section 2.2. This study used the McClear model for Cs irradiance, which was based on a physical model and uses more than one input to the model, mostly from satellite images. Full details can be found in Lefèvre et al. (2013).

### 4.2.2.3  Comparison of Local Test with BSRN and TOACs Tests

The test was developed and was named as a local test to check GHI data for minimum and maximum physically possible limit (Equation 4.6) and extremely rare limit (Equation 4.7) observations, like BSRN and TOACs QC tests.

$$0.005*TOA < GHI_{GD} < 1.2*TOA \qquad (4.6)$$

$$GHI_{GD} < 1.25*Cs \qquad (4.7)$$

The local test boundaries were set to be between BSRN and TOACs. This is because the BSRN test boundaries were developed to be applicable for all the latitude. The TOACs boundaries are restricted whereas these have been applied in several countries (Chapter 2, Section 2.2). The local test boundaries were also set based on the local climate condition in the case study area (Chapter 3, Section 1.2) because the climate is seasonally changeable, which then it affects the minimum and maximum amount of the irradiance used for in the region.

The error detection for lower and upper physically possible limit observations of the local test, as in Equation 4.6 (flags 8 and 9), is compared to conditions in Equations 4.1 and 4.2 (flags 2, 3, 4, and 5) of BSRN and TOACs QC tests respectively. Similarly, the condition of Equation 4.7 (flag 10) of the local test is compared to Equations 4.4 and 4.5 (flags 6 and 7) of the BSRN and TOACs QC tests for extremely rare limit observations. For further clarification, see Table 4.4 and Figure 4.1.

#### 4.2.2.4   Sunshine Duration Test

The relationship between SD and GHI and its use as a consistency test of the GHI data have been described extensively in the literature (Muneer and Fairooz, 2002; Younes et al., 2005; Shi et al., 2008; Moradi, 2009; Journée and Bertrand, 2011a). The SD test is a good option for those areas where only GHI is available, and the irradiance components are not available. This was demonstrated by Moradi (2009) for checking daily data. The comparison test for checking the consistency of data cannot be applied because it depends on diffuse and direct (beam) irradiance in addition to global irradiance. Previous studies (Khaliliaqdam and Soltani, 2012; Moradi, 2009) used the lower limit of SD with a clearness index to test GHI data. Here, both the lower and upper bounds of SD are applied in the testing. First, for the lower bound of SD = 0, in a given time interval GHI should not exceed the maximum possible rate of diffuse irradiance. Otherwise, data values will be flagged as dubious quality according to Equation (4.8). This is because under cloudy conditions when SD is zero, pyranometers record the diffuse irradiance.

The maximum rate of diffuse irradiance is set to 35% of TOA, based on a satellite-derived data, which is available for the case study at SoDa (2018) because of the unavailability of measured diffuse irradiance in the case study. The max diffuse from CRSv3 of SDD was compared with various values of TOA in the case study, until it was set as 35% of TOA. Here, GHI was tested, based on the reality that direct (beam) irradiance should not exceed 120 W/m$^2$ if the sunshine duration is zero (WMO, 2008; Journée and Bertrand, 2011a). This means that the contribution of direct irradiance to the GHI is low when SD is zero, and most of the GHI in this situation is diffuse irradiance. Therefore, the condition of the test was set to explain why GHI is high, while SD is zero. The condition is set as in Equation 4.8 so that if the result is true, the data passes the test and vice versa. The data is checked for this type of error, which can occur because of miscalibration and operational related errors such as high reflected radiation from nearby

equipment. Hence, this situation may happen naturally by broken cloud or bright cloudy sky, but they are not regular. If the rate of the test does not reach a high percentage, it is considered acceptable. However, the hourly data are based on mean irradiance, which includes several situations.

Regarding the upper bound of SD, if the SD exceeds 83% in a given time interval, the solar irradiance should be above 35% of TOA. For SD between 50% and 83%, solar irradiance should be above 10% of TOA. This situation was checked with Equation 4.9. The test detects data values affected by partial shading of the sensor, semi-malfunction, bird droppings on the sensor and other forms of dirt. The test also checks data for systematic errors above 3% of TOA if they have not been detected by the lower limit of TOACs test.

$$SD = 0 \ \& \ GHI_{GD} < M\text{-diff} \tag{4.8}$$
where:
M-diff: Possible max diffuse irradiance equal to 35% of TOA in this study.

$$SD > 50 \min \text{ in } 60 \ \& \ GHI_{GD} > M\text{-diff}; \ 50 > SD < 30 \min \text{ in } 60 \ \& \ GHI_{GD} > 0.1*TOA \tag{4.9}$$

This test was validated with three Australian stations (Table 4.3). The validation was done by comparing SD tests in Equations 4.8 and 4.9 (flags 11 and 12) with flag 17, which is based on the consistency test with the availability of solar irradiance components (the conditions are shown in Table 4.4). The validation is based on comparing (1) how the SD tests passed data or flagged it as having errors with (2) the consistency test in high quality data at Australian stations. This is because the consistency test uses the data records of three pyranometers. Therefore, the rates of error detection in flags 11 and 12 on one side and flag 17 on the other side in that high quality data evaluate the SD tests based on their percentage of error detection compared to the consistency test.

### 4.2.2.5  Air Temperature Test

Using AT for quality checks of GHI data has not been widely described in the literature, although there is a significant relationship between both variables, especially during daytime. One of the main factors of temperature change is solar irradiance. When it is transmitted through the atmosphere, a fraction of the irradiance reaches the earth's

surface and is converted into thermal energy. However, there are several regional, local, and climatological factors affecting these processes.

Several models have used AT to estimate GHI (Hassan et al., 2016; Khaliliaqdam and Soltani, 2012). Therefore, the relationship between the two variables can be used to test GHI data, for example by utilising mean AT for each month for all data observations. Assume that if AT is higher than its monthly average, the GHI should be above 10% of TOA (Equation 4.10), and if the AT is lower than half its monthly average, GHI should be lower than the maximum possible diffuse ratio (Equation 4.11).

$$AT > MATM \; \& \; GHI_{GD} > 0.1* TOA \tag{4.10}$$
$$AT < MATM/2 \; \& \; GHI_{GD} < M\text{-diff} \tag{4.11}$$

where:

MATM: Monthly mean of daytime hourly AT in each month (from January to December)

These are the hypotheses for detecting possible errors in the data. However, the rate of flag errors can be due to local factors that affect the temperature change, whereas in the specific climate regions, the ratio of flag errors should not reach up to 3% of the data. The condition of the test is based on the connection between two variables, when the sun elevation angle is above 15°, and the limit of the test is set to a low level such as 10% and 35% of TOA. This is to decrease the effect of other factors discussed above, because the response of temperature to GHI is delayed slightly by absorption, conduction and transfer. Other QC tests of GHI data in the literature also have some conditionals, namely comparison and statistical tests (Muneer and Fairooz, 2002; Long and Shi, 2008; Shi et al., 2008; Journée and Bertrand, 2011a; Pashiardis and Kalogirou, 2016; Zo et al., 2017), which do not test all the data. This new AT test was checked in semi-arid and Mediterranean climate regions. Some modifications for other climate regions might be needed.

Hourly mean and its hourly half of the mean of AT is calculated in each month for sun elevation angle above 15°. For example, all hourly data of the AT in January are used to calculate mean and half of the mean to test the January GHI data, and so on for each month.

The AT test is useful because:

1. Other tests such as the upper and lower of physically possible limit and extremely rare limit cannot be used for detecting errors in the middle of the data. See Figures 4.6 and 4.7 green lines.

2. When the comparison test based on solar components cannot be applied because diffuse and direct beam irradiance have not been recorded.

3. The temperature is recorded for almost all stations, and the uncertainty of temperature recording equipment is minimal (WMO, 2008; Zahumenský, 2004).

4. When the sunshine record is unavailable at a station, and AT is available.

5. The test can be used for further checks to demonstrate the quality of solar irradiance data or to compare its result with other QC tests.

However, the test has some limitations such as the effect of some local factors and climate conditions on the result and also some natural situation of GHI might be detected as errors. This test will be checked with data from three BSRN stations (Table 4.3). This is quite similar to the validation SD test procedure, but Equations 4.10 and 4.11 AT tests (flags 13 and 14) were compared with the flag 17 consistency test.

### 4.2.2.6 Combining Air Temperature and Sunshine Duration Tests

To reduce uncertainty, to detect errors in both AT and GHI and to enhance the SD test, Equations 4.8 & 4.10 was combined in one new test (Table 4.4, flag 15) and Equations 4.9 & 4.11 was also combined in another new test (Table 4.4, flag 16). All arguments are written in one conditional. This is useful to check each variable against each other to see in which variable the errors are: GHI, AT or SD. The argument is based on conditions of three variables. In these new tests, data were checked for multiple possible errors, for instance, whether AT was above its mean, SD was above 50 min in 60 min, and GHI was lower than 10% of TOA. If there were any cases like this, the data value is flagged. Similarly, if AT was less than half MATM and SD was zero, and GHI was above 35% TOA, the data value was flagged.

All the input parameters for the equations to test the GHI data were calculated or downloaded from related sources for each hour according to the true local solar time for all stations in the case study area.

### 4.2.2.7 Quality Control Flags

As suggested by Maxwell et al. (1993) and applied by Younes et al. (2005) and Moradi (2009), none of the data were modified or deleted. Instead, they were flagged. The data were checked by all tests separately, and each subtest had a flag number. If the condition of an equation or a part of the equation for those equations with two conditions, is true, data were passed; otherwise, data were failed and flagged with the appropriate error flag number (Table 4.4, Figure 4.1). Some flags could be removed by aggregating hourly data to daily data, whereas others could not be removed, because the flaw in the data affects the quality of the aggregation. The flag procedure is considered an easy automatic way to count, check, delete, and interpolate any observation according to its flag number.

### 4.2.2.8 Counting All Tests

Flag 1 is set according to different sets of tests, from TOACs, BSRN and local test alone, by combining them with the SD tests, and subsequently with the AT tests separately, as well as by combining them with both SD and AT tests (Figure 4.1). Unlike previous studies (Moreno-Tejera et al., 2015; Zo et al., 2017) this procedure was used to assess the ratio of data values pass for each test separately. This is important, because in the previous sections all data have been tested with each subtest, whereas here data is assessed to select in which test or tests the data values passed.
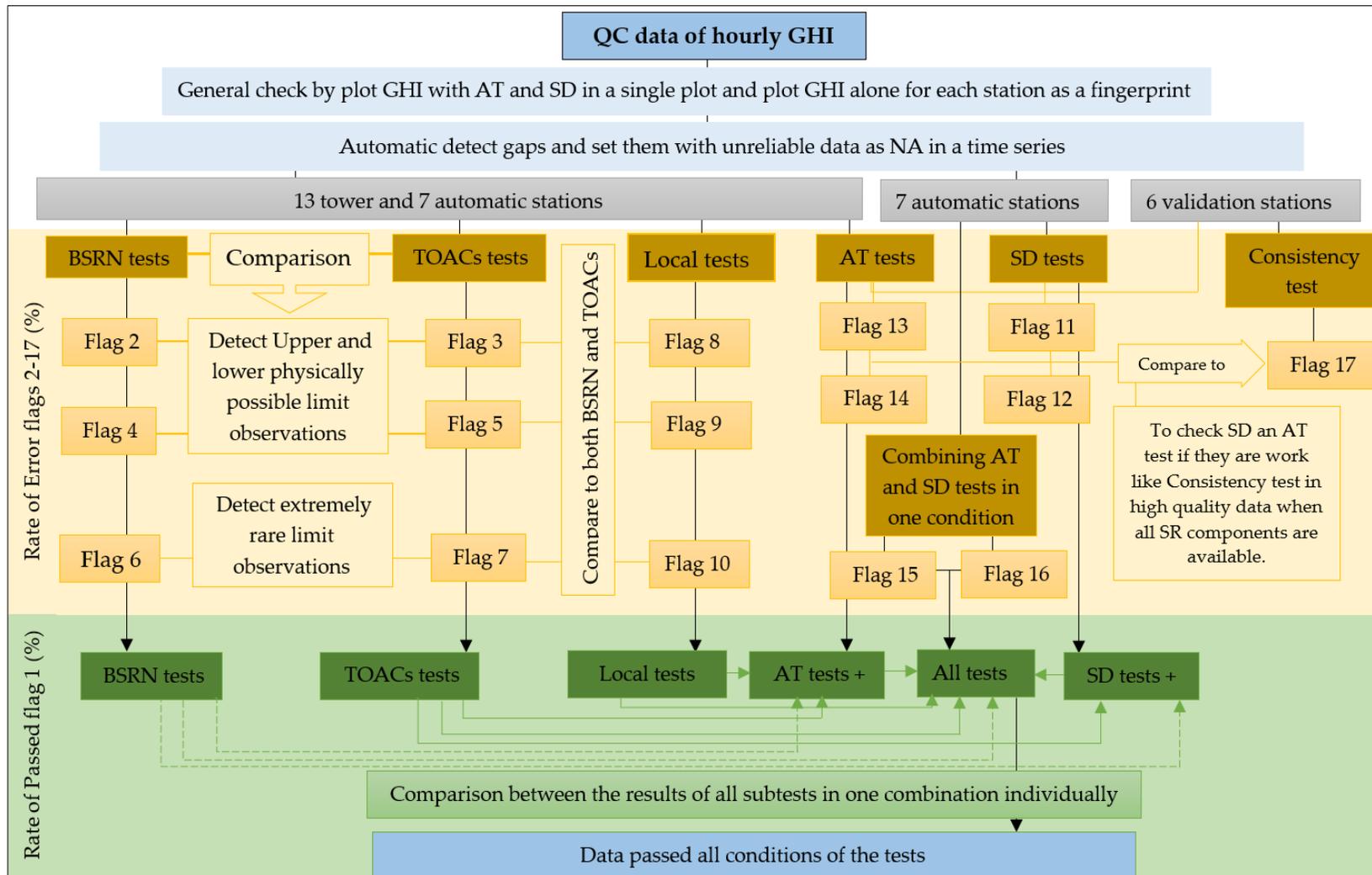
**Figure 4.1:** Flowchart of the proposed QC approach described in this chapter.

**Table 4.4:** Flag number and description of the quality control approach.

| Flag Number Pass | Flag Number Fail | Test condition or criteria | Test Description |
|---|---|---|---|
| 1 | 2 | Upper physically possible limit BSRN second part of Equation (4.1) | Comparison of GHI ground data against TOA and additional TOA following Equation (4.1) for the upper limit and with 3% and −4 for the lower limit. Checks for major errors and flags those as a fail flag. |
| 1 | 3 | Upper physically possible limit TOACs second part of Equation (4.2) | |
| 1 | 4 | Lower physically possible limit BSRN first part of Equation (4.1) | |
| 1 | 5 | Lower physical possibly limit TOACs first part of Equation (4.2) | |
| 1 | 6 | Extremely rare limit BSRN second part of Equation (4.4) | Comparison of GHI ground data against additional of TOA. |
| 1 | 7 | Extremely rare limit TOACs Equation (4.5) | Comparison of GHI ground data against the McClear model for Cs irradiance. |
| 1 | 8 | Upper physically possible limit local test Equation (4.6) | Same as flags 2–5 |
| 1 | 9 | Lower physically possible limit local test Equation (4.6) | |
| 1 | 10 | Extremely rare limit local test Equation (4.7) | Same as flag 7 |
| 1 | 11 | Sunshine 50 minutes in 60 and 35% of TOA: SD between 30 to 50 minutes in 60 and 10% of TOA Equation (4.9) | The test derived from the relation between GHI and SD as an argument when SD is high. Detects errors stemming from shaded or partly shaded conditions and partial malfunction of the sensor. |
| 1 | 12 | Sunshine zero and 35% of TOA Equation (4.8) | Same as flag 11, based on whether SD is zero. Persistency check of data and test for calibration errors of the sensor. |
| 1 | 13 | Temperature above its mean in the month and 10% of TOA Equation (4.10) | Tests the relation between GHI and AT similar to flag 11. Checks the plausibility of data. |
| 1 | 14 | A temperature lowers its half mean in the month and 35% of TOA Equation (4.11) | Similar to flag 12. Also checks the plausibility of the data. |
| 1 | 15 | Combine Equation (4.8) and Equation (4.10) | Based on the relationship among GHI, AT and SD. Tests all three variables against each other. |
| 1 | 16 | Combine Equation (4.9) and Equation (4.11) | |
| 1 | 17 | GHI/DNI*Cos θ +DHI =<1.08, GHI/DNI*Cos θ +DHI =>0.92 | Consistency check based on the combination of the solar irradiance components. Applied for stations in Table 4.7 |
| 1 | Specific number | To count one test or some tests together | Counts if the observation passes all tests or some tests. |

## 4.3   Results

The QC test procedures were applied to GHI data from 20 meteorological stations in northeastern Iraq. All results are presented in Tables 4.5, 4.6 and 4.7 and Figures 4.2–4.9. Table 4.5 shows the results of the tower stations for comparison between the BSRN (flags 2, 4, & 6) and TOACs (flags 3, 5, & 7) tests for each of their subtests, respectively, with AT tests as flags 13 and 14. Flag 2 is compared to flag 3, and flag 4 is compared to flag 5 for detecting observations above and below the upper and lower physically possible limits, respectively, for each subtest of the BSRN and TOACs tests. Similarly, flag 6 is compared to flag 7 for detecting extremely rare observations as subtests in the two tests. The conditions of the BSRN and TOACs tests were compared to the local test (flags 8, 9 and 10) for detecting observations above and below the upper and lower physically possible limit and extremely rare limit observations, respectively. Similar to Table 4.5, Table 4.6 shows the result of automatic stations for all flags in Table 4.5 with flags 11 and 12 of SD tests and flags 15 and 16 for combining AT and SD tests. This is because SD is available in automatic stations only. Table 4.7 shows the result of the validation for SD tests (flag 11 and 12) by consistency test (flag 17) at three Australian stations and evaluating AT tests (flag 13 and 14) by consistency test (flag 17) at three BSRN stations. The rate of passing data by flag 1 among the tests and combination of the tests are demonstrated by Figures 4.8 and 4.9 for the tower and the automatic stations, respectively. General checks for the time series data are shown for some examples in Figures 4.2–4.7. The borders and limits of the tests are shown for different hours of the day at one station as an example (Figures 4.6–4.7). The rates of NA detection are shown in Tables 4.5 and 4.6 for the 20 stations.

### 4.3.1   General check and NA Detection

All available GHI, SD and AT data are shown for a selection of stations in Figures 4.2 and 4.3 (all other stations are in Appendix A, Figures A1−A4). There are systematic errors in the GHI data for Maydan station (Figure 4.2-b) from January to March 2016. Other errors in the GHI data are found for Kalar station (Figure 4.3-b) from the end of 2015 until January 2016, and for Mazne station (Figure 4.3-a). Errors in SD data are present for Bazian (Figure 4.2-a) and Kalar (Figure 4.3-b) stations, especially in the hot summer months, whereas both GHI and AT are normal for the first two stations.

Data gaps (NA) are shown in Tables 4.5 and 4.6. A high rate of missing hourly values (11.3%) recorded by the automatic stations was found at Maydan, and the lowest rate at Halsho (3%). Missing values were not detected at the tower stations except at a negligible rate (0.3%) at Hojava station.

The fingerprint plot of stations, namely Halabja, Dukan and Kalar (Figures 4.4-b, 4.5-a and 4.5-b) show systematic errors in each year in April and September. Other stations are nearly normal with a few questionable data at some of them according to the fingerprint plots (Appendix A, Figures A5–A8). The example of Bazian is shown Figure 4.4-a.

### 4.3.2 Comparing BSRN and TOACs

The limitations and borders of the BSRN and TOACs tests are shown in Figure 4.6 and 4.7 as samples from different hours of the day. For the physically possible upper limit test (flags 2 and 3) most automatic and tower stations passed the flag 2 checks, whereas low rates were recorded in some stations for flag 3 (Tables 4.5 and 4.6).

All data values passed the flag 4 BSRN test, while flag 5 for the TOACs test was raised by 9.53%, 6.14%, 5.55% and 1.46%% of the data values recorded at Kalar, Banmqan, Mazne and Surdash, respectively. The error rate for the same flag is lower than 1.15% for other stations (Tables 4.5 and 4.6).

The data were checked to detect extremely rare limit observations by flags 6 and 7. The rate of flag 6 is zero in all stations except Banmqan, which recorded 0.75%. In contrast, high rate of flag 7 was recorded in most stations. The highest rates were 27.18%, 17.45%, 12.98% and 11.65% Surdash, Banmqan, Aliawa and Kalarikon at stations respectively. However, two low rates were recorded for the same flag, which are 0.03% and 0.46% at Halabja and Darband stations. Other values of flag 7 for remaining stations range from 1% to 9% (Tables 4.5 and 4.6).

### 4.3.3 Comparing BSRN and TOACs QC Tests to the Local Test

The limitations and borders of the local test compared to the BSRN and TOACs tests are shown in Figures 4.6 and 4.7. For the physically possible upper limit test (flags 2 and 3 of the BSRN and TOACs tests compared to flag 8), most stations passed those flags

with minor differences. For the physically possible lower limit test, error detected by flag 9 of the local test is quite similar to the flag 5 of TOACs test, and it is quite different compared to flag 4 of the BSRN test at most of the stations (Tables 4.5 and 4.6).

The results of error detection for extremely rare limit observations as flag 10 of the local test is quite different compared to the flags 6 and 7 of BSRN and TOACs respectively. The rates of flag 10 were under 2% at most of the stations, and zero rates were recorded at Dukan and Halabja stations. The high rates of flag 10 by 6.44% and 3.21% were recorded at Banmqan and Surdash stations (Tables 4.5 and 4.6).

### 4.3.4 Sunshine Duration Test

This test is applied only to automatic stations. The rate of flag 11 is near zero at three stations but reached 1.68% at Halabja and 7.36% at Kalar stations. Flag 12 registered high rates of 17.64%, 12.3%, 7.38% and 3.93% at Halsho, Bazian, Kalar and Maydan respectively, but the rates of the three other stations were lower than 1% (Table 4.6). In the validation of this test according to the Table 4.7, the rate of flag 11 was near 0%, but the flag 17 consistency test is zero in two Australian stations and near zero (0.02%) in the other one. Hence, flag 12 recorded low rates of 0.59%, 0.26% and 0.43% at the same three stations.
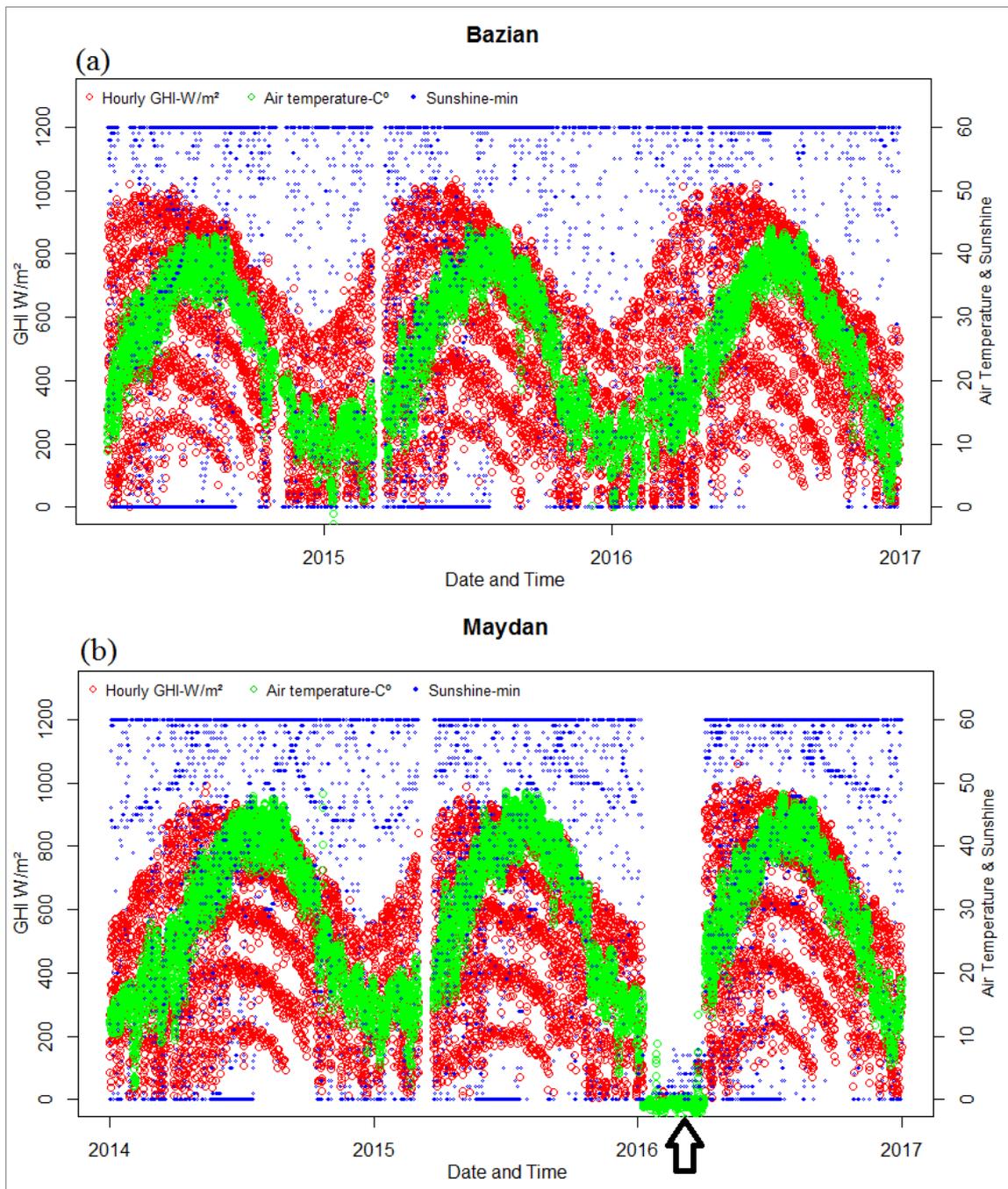
**Figure 4.2:** Scatter plots of GHI W/m$^2$ (first y-axis) and SD in minutes per hour with AT in °C (second y-axis) for all hourly time series data in each station.
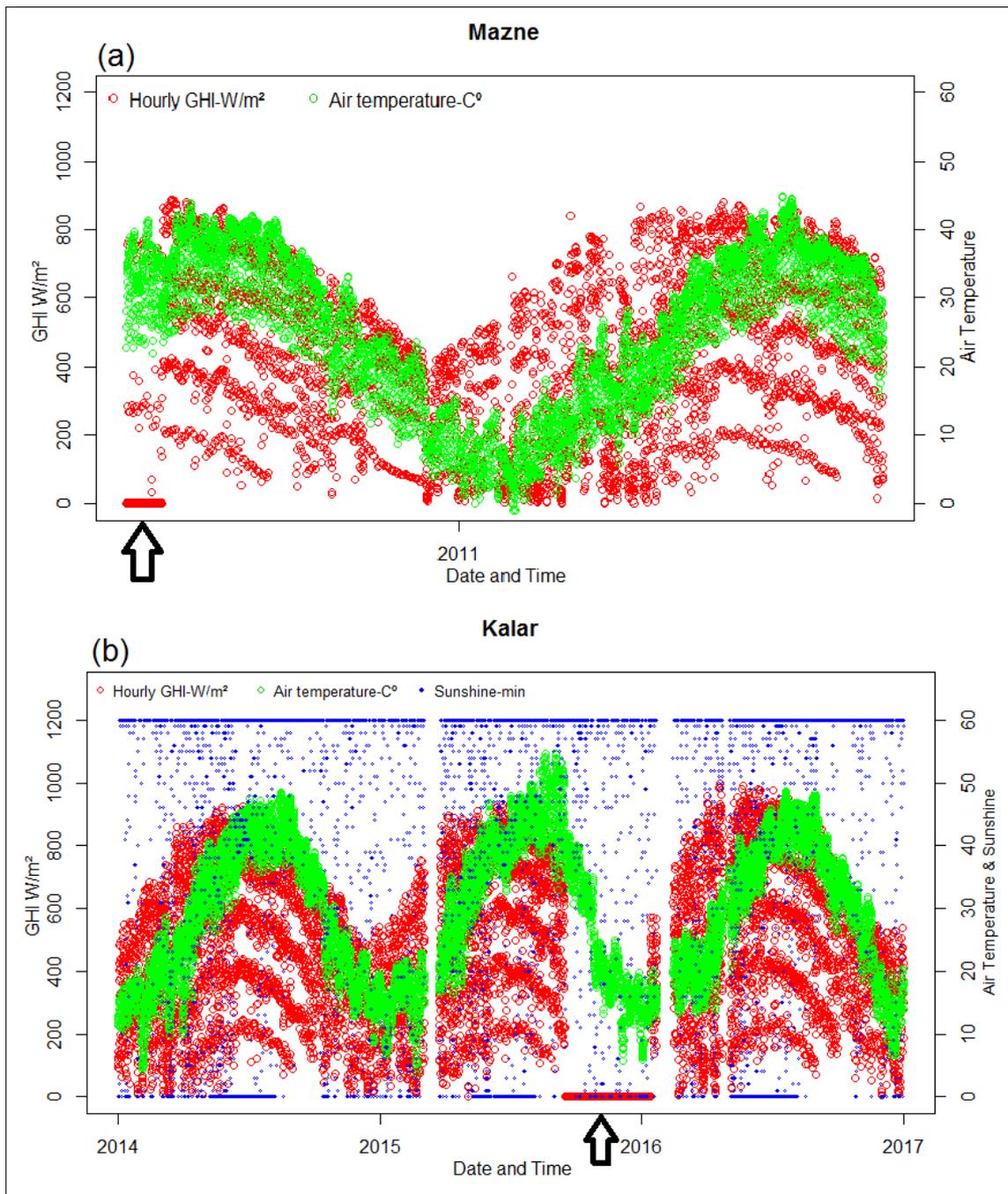
**Figure 4.3:** Scatter plots of GHI W/m$^2$ (first y-axis) and SD in minutes per hour with AT in °C (second y-axis) for all hourly time series data in each station.
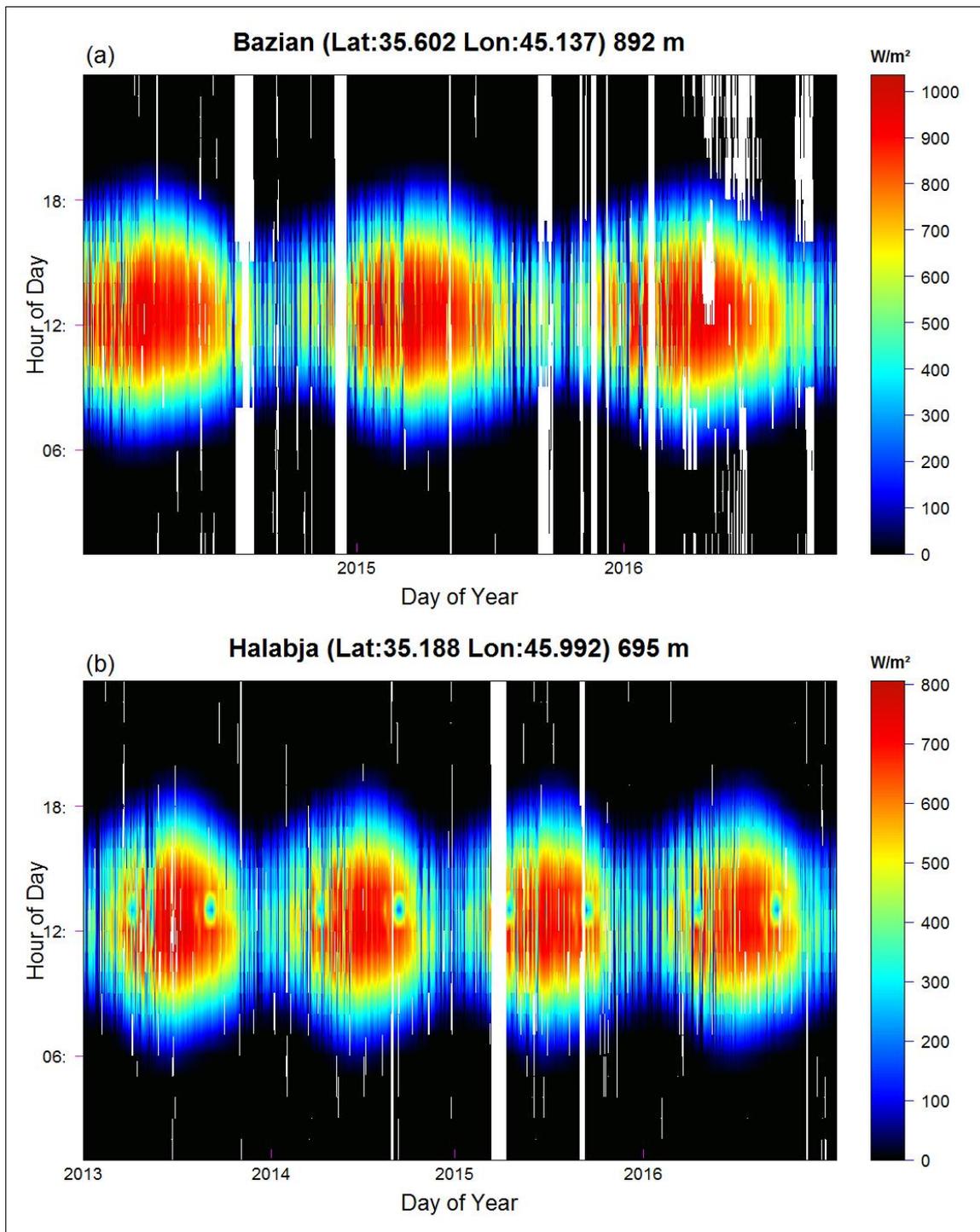
**Figure 4.4:** Fingerprint plot GHI time series of hourly data for four stations as an example, white colour shows NA values.
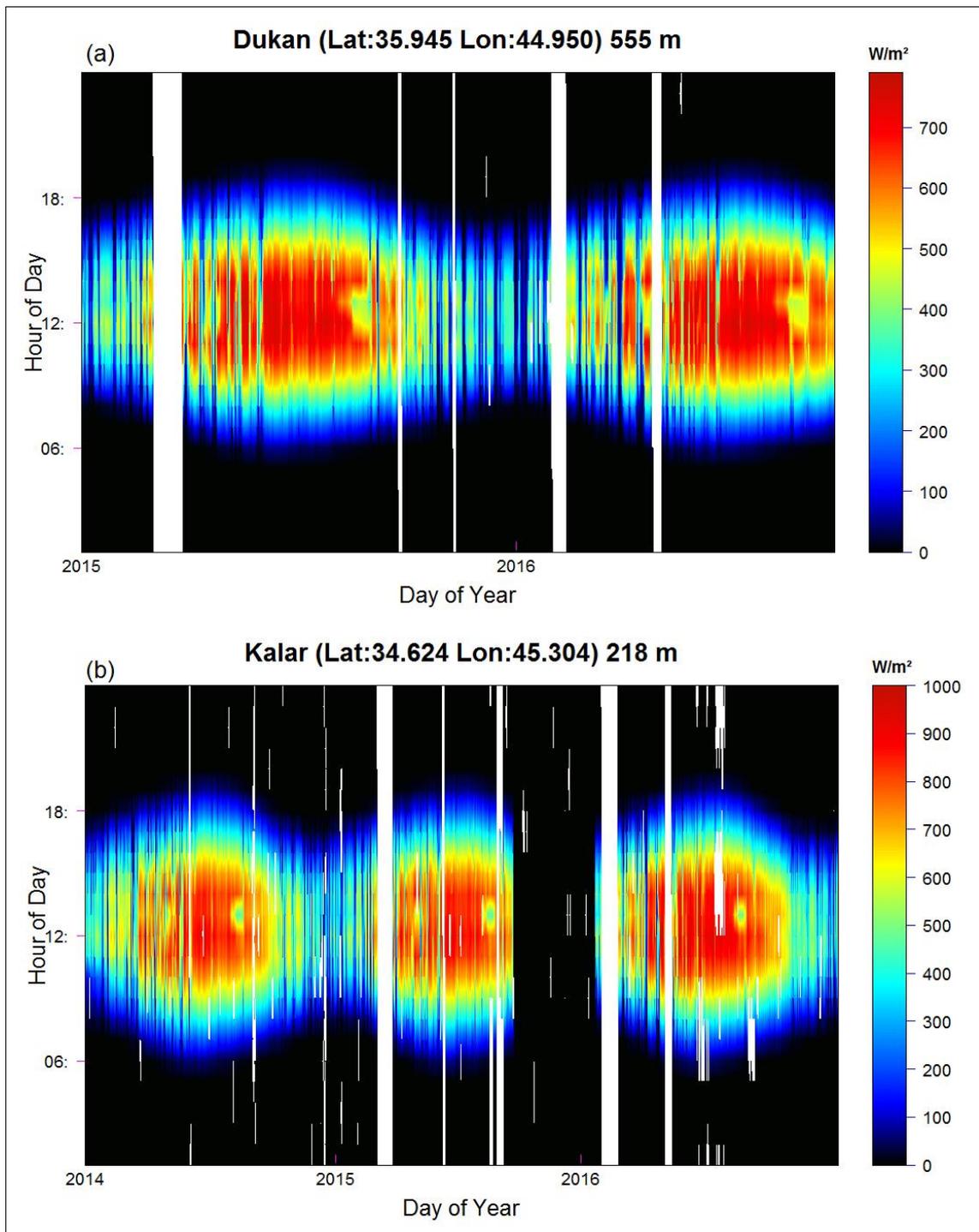
**Figure 4.5:** Fingerprint plot GHI time series of hourly data for four stations as an example, white colour shows NA values.
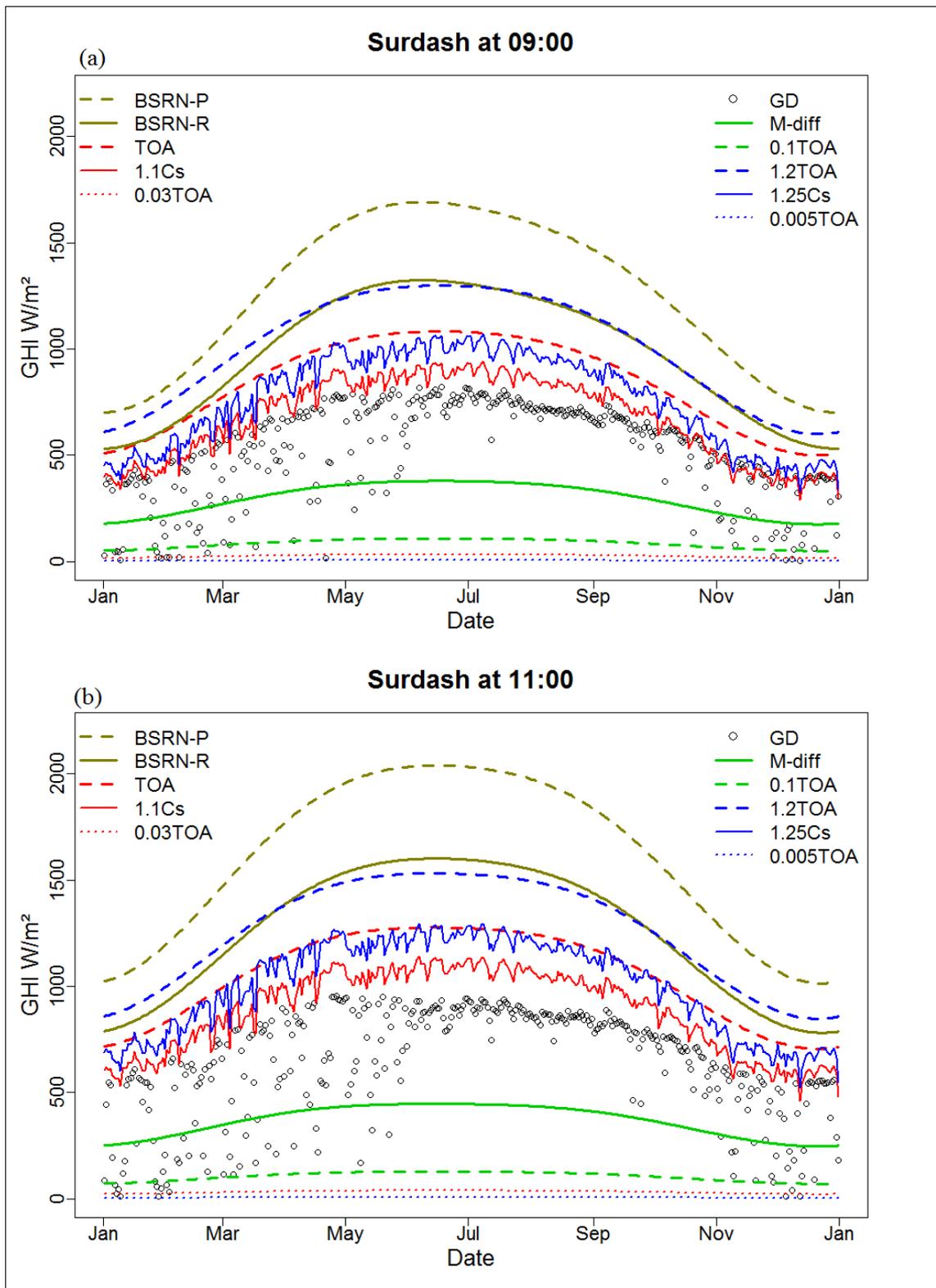
**Figure 4.6:** Scatter plots of the limits for tests, namely BSRN (BSRN-P & BSRN-R) in olive, TOACs in red and local test in blue for lower and upper physical possible limit and extremely

rare limit checking respectively, SD and AT are in green for consistency check and GD is GHI ground data for each hour at Surdash station as an example.
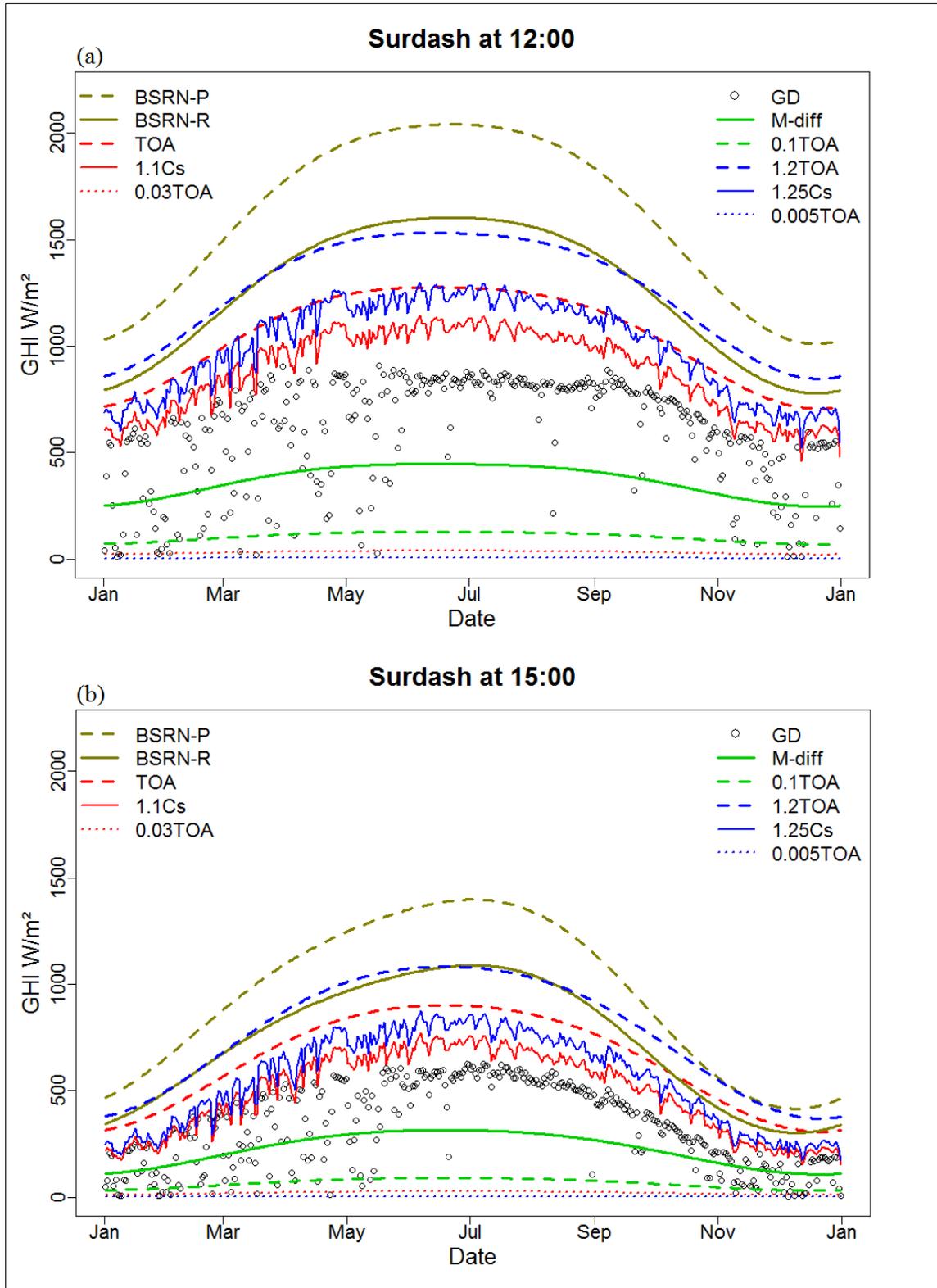


**Figure 4.7:** Same as Figure 4.6, but for hours at 12:00 and 15:00.

**Table 4.5:** The ratio of NA and the error flags in the tower station data. F= flag.

| Name | NA (%) | F2 (%) | F3 (%) | F4 (%) | F5 (%) | F6 (%) | F7 (%) | F8 (%) | F9 (%) | F10 (%) | F13 (%) | F14 (%) |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|
| Barzor | 0.0 | 0 | 0.0 | 0 | 0.61 | 0 | 8.79 | 0 | 0.02 | 1.22 | 0.68 | 0.68 |
| Batufa | 0 | 0 | 0.01 | 0 | 0.68 | 0 | 9.21 | 0 | 0.03 | 0.84 | 0.63 | 1.49 |
| Enjkasor | 0 | 0 | 0.04 | 0 | 0.43 | 0 | 8.12 | 0 | 0.01 | 1.58 | 0.81 | 1.40 |
| Hojava | 0.3 | 0 | 0.05 | 0 | 0.40 | 0 | 5.52 | 0.01 | 0.02 | 1.06 | 0.63 | 1.63 |
| Mazne | 0 | 0 | 0 | 0 | 5.55 | 0 | 4.36 | 0 | 5.01 | 0.67 | 3.31 | 1.49 |
| Kanispi | 0 | 0 | 0.01 | 0 | 0.46 | 0 | 8.03 | 0 | 0.01 | 1.67 | 0.71 | 1.05 |
| Jazhnikan | 0 | 0 | 0 | 0 | 0.31 | 0 | 4.45 | 0 | 0 | 0.66 | 0.53 | 1.02 |
| Aliawa | 0 | 0 | 0 | 0 | 0.24 | 0 | 12.98 | 0 | 0.01 | 2.29 | 0.33 | 0.76 |
| Tarjan | 0 | 0 | 0.08 | 0 | 0.29 | 0 | 9.18 | 0 | 0.01 | 1.67 | 0.65 | 1.77 |
| Shabakaykon | 0 | 0 | 0.0 | 0 | 0.86 | 0 | 2.21 | 0 | 0.31 | 0.29 | 0.69 | 0.63 |
| Surdash | 0 | 0 | 0.03 | 0 | 1.46 | 0 | 27.18 | 0 | 0.03 | 3.21 | 1.11 | 1.75 |
| Banmqan | 0 | 0 | 0.20 | 0 | 6.14 | 0.75 | 17.45 | 0 | 6.14 | 6.44 | 0.0 | 0.40 |
| Kalarikon | 0 | 0 | 0 | 0 | 0.07 | 0.0 | 11.65 | 0 | 0 | 1.08 | 0.04 | 0.36 |

Note: some flags were not applied because SD is not recorded at tower stations.

### 4.3.5    Air Temperature Test.

The occurrence rate of flag 13 is below 1% for most automatic and tower stations. The two highest rates were 4.54% and 3.31% at Kalar and Mazne. Similarly, the rate of flag 14 is below 1% at 11 stations, and the highest percentage of 1.81% was recorded at Halsho station. The rate of the remaining stations was lower than 1.8% (Tables 4.5 and 4.6). Table 4.7 compares this test with a consistency check, and reveals that the rate of flags 13 and 14 were 0%, but flag 17 reached 1.87% at Petrolina station. The same two flags recorded low rates of 0.04% and 0.58%, whereas flag 17 reached 2.5% at Sede Boqer. At Carpentras, flags 13 and 14 rose to 0.93% and 2.08% of the data while flag 17 had a very low rate of only 0.03%.

### 4.3.6    Combining Air Temperature and Sunshine Duration Test

Zero rates were recorded for flag 15 except at one station (Kalar) with 3.45%. The rate of flag 16 is also below 0.5% at two stations; it reached zero at the other five stations (Table 4.6).

### 4.3.7  Data Pass (Flag 1)

The result of flag 1, which indicates values of GHI pass tests or all tests, is represented by five combinations. First, according to all BSRN test, the GHI data showed a high percentage pass rate for flag 1 of all stations at 100%, except Banmqan station at 99.75% (Figures 4.8 and 4.9). Second, the results of flag 1 according to the TOACs test are quite different from the BSRN test. The two lowest pass rates were recorded at Surdash and Banmqan stations at 71% and 76%. The other stations, Aliawa, Kalar and Kalarikon, had pass rates less than 90%, and all other rates ranged from 90% to 99%. Third, the results of flag 1 according to the local test are different compared to the TOACs test by high rates. This is because data pass according to the local test are above 98% at most of the stations. Two low rates were also recorded for the local test at Kalar and Banmqan at 90% and 87% (the same seen as the TOACs test), but the rates of the local test are higher than TOACs test. The notable rate is that the rate of flag 1 according to the local test is 96% at Surdash stations, whereas it is 71% according to TOACs test. Flag 1 rates of the local test differs from the BSRN test by 2% at most of the stations while at Mazne, Kalar and Banmqan stations rates differ by 5%, 10% and 12% (Figures 4.8 and 4.9).

Next, when combining the AT tests with the previous two tests, the passed data percentage according to this test was generally lower by nearly 1–2 % of the TOACs test (Figures 4.8 and 4.9).

Finally, the combination of the SD tests combined with AT tests, local test, TOACs, and BSRN tests were applied only to the automatic stations. The results of SD with TOACs and BSRN are shown in Figure 4.9. The pass rate of flag 1 was lowest at Halsho, Kalar and Bazian (76%, 81% and 84%). In contrast, the pass rate of the other four stations was above 90%. The result of mixed tests of AT and SD with the local test, TOACs and BSRN were quite similar to the result of TOACs alone (Figure 4.8).

**Table 4.6:** The ratio of NA and error flags at the automatic stations.

| Name | NA (%) | F2 (%) | F3 (%) | F4 (%) | F5 (%) | F6 (%) | F7 (%) | F8 (%) | F9 (%) | F10 (%) | F11 (%) | F12 (%) | F13 (%) | F14 (%) | F15 (%) | F116 (%) |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Halsho | 3.0 | 0 | 0.01 | 0 | 1.15 | 0 | 5.61 | 0 | 0 | 0.86 | 0.01 | 17.64 | 0.80 | 1.81 | 0 | 0.36 |
| Dukan | 7.0 | 0 | 0 | 0 | 1.14 | 0 | 1.12 | 0 | 0.02 | 0 | 1.12 | 0.02 | 1.85 | 0.72 | 0 | 0 |
| Bazian | 5.6 | 0 | 0 | 0 | 0.36 | 0 | 3.17 | 0 | 0.02 | 0.36 | 0.07 | 12.3 | 0.29 | 0.72 | 0 | 0.04 |
| Halabja | 3.3 | 0 | 0 | 0 | 0.60 | 0 | 0.03 | 0 | 0.01 | 0 | 1.68 | 0.03 | 0.54 | 0.46 | 0 | 0.0 |
| Darband | 7.5 | 0 | 0 | 0 | 0.97 | 0 | 0.46 | 0 | 0.02 | 0.06 | 0.56 | 0.09 | 1.22 | 0.31 | 0 | 0 |
| Maydan | 11.3 | 0 | 0 | 0 | 0.45 | 0 | 2.83 | 0 | 0.16 | 0.55 | 0.02 | 3.93 | 0.07 | 0.20 | 0 | 0.0 |
| Kalar | 6.5 | 0 | 0 | 0 | 9.53 | 0 | 1.84 | 0 | 9.36 | 0.28 | 7.36 | 7.38 | 4.54 | 0.09 | 3.45 | 0.0 |

**Table 4.7:** The ratio of error flags at the validation stations.

| Name | F11 (%) | F12 (%) | F113 (%) | F14 (%) | F117 (%) |
|------|------|------|------|------|------|
| Carpentras –France | - | - | 0.93 | 2.08 | 0.03 |
| Sede Boqer - Israel | - | - | 0.04 | 0.58 | 2.5 |
| Petrolina- Brazil | - | - | 0 | 0 | 1.87 |
| Geraldton -  Australia | 0 | 0.59 | - | - | 0.02 |
| Longreach - Australia | 0.03 | 0.26 | - | - | 0.0 |
| Broome – Australia | 0.02 | 0.43 | - | - | 0.0 |
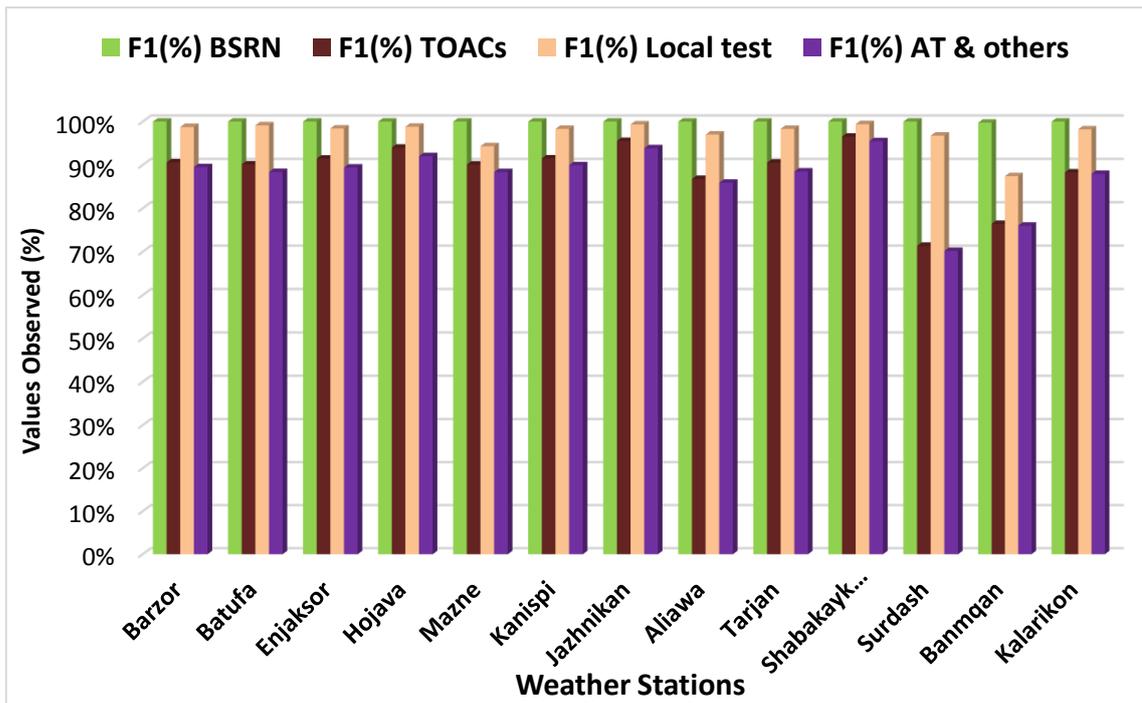
**Figure 4.8:** Bar chart of flag 1 passed data according to each set of tests in tower stations. 'Others' refers to previous tests such as BSRN and TOACs.
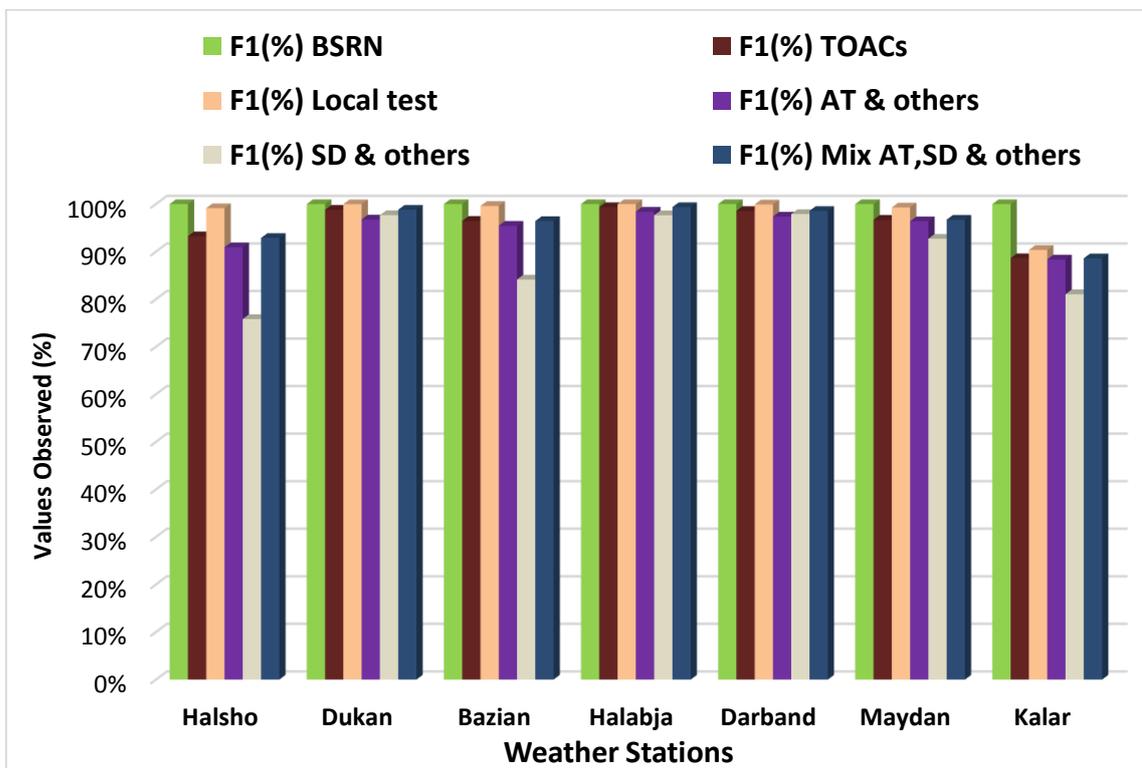


**Figure 4.9:** Bar chart of flag 1 passed data according to each set of tests in automatic stations.

## 4.4 Discussion

The application of the sets of QC tests to the GHI station data in northeastern Iraq showed that both station types generally have high data quality. Data gaps are generally very limited at the tower stations, whereas all automatic stations have a rate of missing values, which are similar to previously published studies (Roesch et al., 2011; Schwandt et al., 2014). The results of QC of GHI show slight differences between the BSRN and TOACs tests on one side and the difference between the local test compared to BSRN and TOACs on other side. The results of quality checks of GHI based on SD vary compared with those of the station and with the lower and upper limits of SD. Small error rates are detected by the AT tests for all stations. The errors flagged up by a combination of SD and AT are generally low. The results of evaluating SD and AT by consistency test at validation stations support the SD test but not the AT tests. The rate of AT test errors at the validation stations also is low.

The general reliability and error rates for GHI, SD and AT can be highlighted by comparing them to each other (Figures 4.2 and 4.3) or for GHI by fingerprint plots (Figures 4.4 and 4.5). Those plots are important to check all data, whereas their results are more obvious if they are used to check one to three years data rather than above three years. The figures represent hourly data. For instance, when minute data is aggregated as hourly values and the errors are therefore difficult to detect. Figure 4.2 identifies some equipment errors in the GHI data when compared to a time series of AT. Figure 4.2-b shows that both GHI and AT have systematic errors. This type of error is not easily detected by comparing both variables. However, for the fingerprint plot some errors can be seen easily if the plot represents one- or ten-minutes data while some errors are seen for hourly data (Figures 4.4 and 4.5).

The two tests showed different results (Tables 4.5 and 4.6). In the past, several studies have applied either BSRN or TOACs. The difference between the two tests is evident from Tables 4.5 and 4.6 and Figures 4.8 and 4.9. This is because the limits of the BSRN tests are higher than the TOACs tests (Figures 4.6 and 4.7).

The detection errors of the upper physically possible limit as (flag 2) and (flag 3) are relevant in the two tests, whereas more observations are flagged as errors by the TOACs test than by the BSRN test (Tables 4.5 and 4.6). This is because the TOACs test depends

on TOA and BSRN depends on increased TOA (Equation 4.1, Figures 4.6 and 4.7). This result of each test is in agreement in term of error detection rates with published studies (Schwandt et al., 2014; Khaliliaqdam and Soltani, 2012; Urraca et al., 2017a; Moradi, 2009; Pashiardis and Kalogirou, 2016; Geiger et al., 2002) that have applied each test separately.

The most important feature is that the lower limit of the BSRN test (flag 4) leaves errors, even significant systematic errors, undetected (Tables 4.5 and 4.6, Figures 4.2–4.5). This is due to setting the lower limit to ($-4$ W/m$^2$). In contrast, many data values are flagged by the TOACs test (flag 5) (Tables 4.5 and 4.6, Figures 4.2 and 4.3). This might be due to the full or partial shading of the sensor, dirt on the sensor or malfunction of the sensor (Younes et al., 2005; Schwandt et al., 2014; Lemos et al., 2017). The lower limit of the BSRN test can be useful when checking day- and night-time data in cold regions. It is clear that there is no negative value of GHI, but this situation happens when the calibration calculation depends on the temperature difference between the dark and bright area in the active part of the sensor, which occurs at night (Roesch et al., 2011; Zo et al., 2017).

Another interesting aspect is that the error rates of the BSRN for the extremely rare limit test observations (flag 6) are zero; while flag 7 of the TOACs test for the same target recorded high rates at most stations. The high difference between the two tests is related to the border limits (Figures 4.6 and 4.7). However, high error rates of flag 7 are detected in this study. Similar error rates are reported in the literature (Journée and Bertrand, 2011a; Lemos et al., 2017; Moradi et al., 2009). Those studies explained that this situation happens in mid-high latitudes when clouds reflected radiation is received by the sensor, resulting GHI greater than Cs irradiance. In the case study, the high errors are also explained in the case of Surdash and Banmqan stations, by operational issues such as miscalibration and dirty sensor because most other tests detect the errors in those two stations or by the Cs model, which can under estimate GHI (Lefèvre and Wald, 2016) especially when the sun elevation angle is between 16–20°.

The results of the local test are interesting compared to BSRN and TOACs because the systematic errors were detected by flag 9 of the local test similar to flag 5 of the TOACs test and unlike flag 4 of the BSRN test, and the high rates of values passed flag 10 of the local test similar to flag 6 of the BSRN test and unlike flag 7 of the TOACs test

(Tables 4.5 and 4.6). This is related to the modifications of boundaries, which have been edited in the local test (Equations 4.6 and 4.7, Figures 4.6 and 4.7). This reveals the vital of modification of those QC tests (BSRN and TOACs) when they are needed to be applied in a new study area as in this study.

The result of the SD test for flag 11 recorded two high error rates, which are partly related to errors in the SD recorder. Other error rates indicate that the pyranometer needed to be checked for partial shading or dirt contamination especially in April and September (Figures 4.4 and 4.5) when most of the errors are detected by flag 11 (Table 4.6). For the lower limits of the SD test (flag 12) the high rate of errors at Bazian and Kalar stations during particular times are related to systematic errors in SD itself, not in the GHI data (Figures 4.2-a and 4.3-b), as seen in the flag 16 result. This indicates that when another variable is used for QC of GHI data, it should be checked prior to the analysis. Studies mentioned some ways for testing an SD recorder (Zahumenský, 2004; Journée and Bertrand, 2011a). Some studies have used SD for testing the GHI data without checking it against another variable such as AT for more accurate results (Moradi, 2009; Khaliliaqdam and Soltani, 2012). Therefore, SD were compared with the possible maximum diffuse TOA irradiance and with AT, and the results are acceptable (flag 15 and flag 16). In this way, both SD and GHI are tested (Table 4.6). From the comparison of flags 11 and 12 (Tables 4.6 and 4.7) at the case study stations and the validation stations, the results show good agreement. The rates of errors are relevant except where the errors are related to SD and not to the GHI. However, flag 12 recorded low rates (under 0.5%) and the incidence of flag 17 was nearly zero in all validation cases. These findings generally support the use of SD as a consistency check.

Published uses of AT for QC of GHI are limited in the literature. The results here demonstrate good agreement of the AT test for its lower and upper limits with other tests such as in the case of Kalar, Dukan, Surdash and Mazne stations (Tables 4.5 and 4.6). This is for its upper limit, which supports the results of other tests, and the rates of error are because of the same reason as flag 5. For low AT and high GHI, the result of AT also supports, other tests at most tower stations and Kalar and Halabja automatic stations. In the case of comparing AT (flag 13 and flag 14) with the consistency test (flag 17). The validation station results did not support using AT as a consistency test because error rates was detected by the consistency test unlike AT test (Table 4.7). This is because the error

rates detected by the AT tests at Carpentras and Petrolina are related mainly to the local conditions, whereas comparison rates at Sede Boqer tend to be relevant. The low error rates of nearly 2% of dubious data values according to the AT tests in high quality data tend to support the use of AT a plausible test (Table 4.7). This supports our previous argument with the AT test, which is based on no more than 3% errors in a station dataset. Other studies have used the other components of solar irradiance (Moreno-Tejera et al., 2015; Zo et al., 2017) and using only SD (Moradi, 2009; Khaliliaqdam and Soltani, 2012) whereas our model is based on AT which is available at most stations.

Another interesting point of this chapter is related to the error rates identified by flags 15 and 16, which indicate that the errors are not generally in the GHI data but are instead related to the SD and AT variables.

The rate of flag 1, which means data pass for various types of tests, reveals a high difference between BSRN and TOACs tests. This indicates that the BSRN test is not acceptable if only GHI is available because even systematic errors are not detected (Figures 4.2–4.9). The TOACs test detected several systematic errors, and other errors, whereas some other observations have also been flagged due to the using of 110% of Cs model as a limit. The rates of flag 1 based on the local test are quite acceptable because not all data passed all tests like BSRN and the rates of passing values were not low like TOACs. This indicates how the modification of those tests is crucial, if they are applied in an area when only GHI available. However, unlike this study several studies applied those QC tests such as BSRN (García Cabrera et al., 2018; Long and Shi, 2008; Perez-Astudillo et al., 2018; Roesch et al., 2011; Zo et al., 2017) and TOACs (Khaliliaqdam and Soltani, 2012; Moradi, 2009; Tang et al., 2010) in regions without any modifications. The rates of data pass based on AT tests with others are lower by nearly 2% based on TOACs, which indicates that some data values are flagged as dubious only by the AT test. The rates of data passing the QC according to the SD test are low because errors could be in the SD recorder, which were detected by combining it with the AT test (Table 4.6, Figures 4.8 and 4.9). The results reveal that SD and AT tests have detected a rate of dubious data which had not been detected by any other tests due to the chosen limits of those tests (Figures 4.6 and 4.7).

Limitations of the tests are as follows. Firstly, the general plot (Figures 4.2 and 4.3) is not always reliable especially for a large number of observations. Secondly, the chosen

limits for the SD and AT tests might not be perfect. For example, the upper limit of SD was set to 50 and from 30–50 minutes in one hour, which is based on an assumption about why SD is high and GHI is low; there might be some questionable data under that limit. Previous studies have used only one argument as a lower limit (Khaliliaqdam and Soltani, 2012; Moradi, 2009). Our set limits for these two tests are near the middle of the data for the upper limit and far from the TOACs by 7% for the lower limit (Figures 4.6 and 4.7). This is important to identify errors in that border and contribute to other tests. Thirdly, the mean and the half of mean AT in each month are used to test GHI with 10% and 35% of TOA, which also tends not to be perfect. This is mainly because there are some times when the arguments of AT test may happen naturally, especially when the AT is lower than its half of the mean in the month and GHI is above 35% of TOA. The arguments need to be modified by using the AT test in other climate regions.

Owing to the limitations of recording minimum and maximum AT in many stations in the study area and the fact that the increase in AT from one hour to the next is not high (Zahumenský, 2004), the mean AT was used for the test. In contrast, some studies have estimated GHI from minimum and maximum AT (Hassan et al., 2016; Besharat et al., 2013). Unless SD and AT have been validated with the consistency test, which tests all single observations, AT and SD have limited borders for tests according to conditional arguments (Table 4.4, Figures 4.6 and 4.7), which means that they do not test every single observation in the time series. Generally, the results of AT for testing GHI data show good agreement with other tests and AT is useful to enhance the SD test.

## 4.5   Conclusions

The study in this chapter has applied QC approaches for flagging data values of hourly GHI, which are of dubious quality, using the BSRN and TOACs tests for the upper and lower limits of physically possible and extremely rare observations, modifying those QC tests as a local test for the same target. New quality checks were developed based on SD and AT for stations where solar irradiance components were unavailable; these tests detected further errors in the data at seven automatic stations and thirteen tower stations in semi-arid and Mediterranean Sea climate regions. The new tests were validated with high quality meteorological data from six stations in various regions around the globe with similar climate types. The results demonstrate the high percentage difference between BSRN and TOACs for each subtest due to the different limits. This indicates that

BSRN cannot be used when only GHI is available because most errors will not be detected. Hence, the extremely rare limit of TOACs detected high rates of errors. The results of the local test covered the limitations in the BSRN and TOACs QC tests by detecting systematic errors and not flagging high rates of observations as errors for the extremely rare limit test. SD can be used as a partial consistency test, which has been supported by the validation results. Contrary to that, AT has not been supported as a suitable test. However, it is possible that AT can be used to generally check GHI data, especially when the components of solar irradiance data are unavailable. The AT test detected very low error rates in high quality data at the validation stations. Further research is required to compare BSRN and TOACs tests in other areas. Using several arguments with mean AT or with minimum and maximum AT, and also using other climate variables to check the quality of solar irradiance data will be useful.

Chapter 5 :   **Validation of Hourly Global Horizontal Irradiance for Two Satellite-Derived Datasets in Northeast Iraq**

## 5.1   Introduction

This chapter addresses objectives:

3. To validate the SDDs in new areas and to investigate the spatiotemporal features of SDDs to ground data, and to investigate the difference of two pixel point data from SDDs around a station compared with ground data of a station.

4. To evaluate the ability of a simple method for bias correction in SDDs under cloudy sky conditions.

Validation of SDD is considered to be crucial owing to the limited recording and limited spatial coverage of ground measurements and a high cost of instruments for GHI. This chapter deals with the validation of SDDs with GHI quality controlled data from chapter four. Nine stations are selected out of twenty stations owing to the low rates of data passing the QC tests and issues related to calibration of the pyranometers.

Therefore, this chapter aimed to compare the hourly GHI from HC3v5 and CRSv3 with ground measurements at those nine stations in northeast Iraq. It is first study to validating those SDDs in that region. One intention of this chapter is to evaluate the spatial-temporal performance of those datasets in all-sky, clear-sky, and cloudy-sky conditions and with the clearness index. Another goal is to use a new approach for validation, which is limited in the literature, comparing the GHI from ground measurements at a station against the GHI from SDDs at each station location and the two points around it, at a spatial resolution of 5 km (corresponding to that of MSG imagery), and with each point collected from a different pixel. The chapter was also intended to apply and to test a method for bias correction of GHI in SDDs under cloudy-sky conditions in the case study area.

## 5.2    Materials and Methods

### 5.2.1    Study Site

The study area is in northeast Iraq, and was described in Chapter 3, Section 3.1. Based on the QC approaches (Chapter 4) and some calibration issues of GHI data, nine stations were used in this chapter from the twenty stations considered in Chapter 4.

### 5.2.2    Ground Measurements

The hourly GHI data with some other climate variables were collected from two station types. First, the data are from tower stations. The pyranometer used for recording data in these stations is the Kipp and Zonen CMP6 Pyranometer. The data were collected for the period 2011–2014 from five stations, which lacked some years, from the Ministry of Electricity-Kurdistan Regional Government (KRG) (Table 5.1). Others are automatic stations equipped with an Vaisala QMS101 Pyranometer. The data were collected from the General Directorate of Meteorology and Seismology-KRG from four stations (2013–2016), which lacked some years (Table 5.2).

**Table 5.1:** Tower stations with hourly GHI from Kipp and Zonen CMP6 Pyranometer.

| Station | Coordinates (Degrees) | | Elevation a.s.l (m) | Period (dd/mm/yy) |
|---|---|---|---|---|
| Batufa | 37.1764 N | 43.0236 E | 947 | 01/01/2011–31/12/2013 |
| P1-Batufa | 37.1952 N | 42.9478 E | 854 | |
| P2-Batufa | 37.1689 N | 43.1042 E | 885 | |
| Enjaksor | 37.0603 N | 42.4353 E | 509 | 01/01/2011–31/12/2014 |
| P1-Enjaksor | 37.0642 N | 42.3544 E | 433 | |
| P2-Enjaksor | 37.0533 N | 42.4936 E | 520 | |
| Hojava | 37.0075 N | 43.0369 E | 933 | 01/01/2011–31/12/2013 |
| P1-Hojava | 37.0331 N | 42.9803 E | 856 | |
| P2-Hojava | 37.0061 N | 43.0883 E | 940 | |
| Jazhnikan | 36.3564 N | 43.9556 E | 430 | 01/01/2011–31/10/2013 |
| P1-Jazhnikan | 36.3672 N | 43.8936 E | 376 | |
| P2-Jazhnikan | 36.3347 N | 44.0294 E | 467 | |
| Tarjan | 36.1258 N | 43.7353 E | 276 | 01/01/2011–31/12/2013 |
| P1W-Tarjan | 36.1297 N | 43.6686 E | 263 | |
| P2-Tarjan | 36.1208 N | 43.7931 E | 308 | |

**Note:** The periods in the table are available for the ground measurements; the SDDs and TOA for the same periods have been collected for each station location and points around the stations, which were used for validation.

### 5.2.3 Calculated Data

These were described in (Chapter 4, Section 2.1.2).

### 5.2.4 Satellite-Derived Datasets

The SoDa portal (SoDa, 2018) is owned by MINES ParisTech and Transvalor-France. It provides a dataset of solar irradiance components, which are based on converting satellite images of MSG in the field view of the SEVIRI instrument covering Europe, Africa, the Middle East and part of South America (Figure 5.1) by the HC3 and CRS datasets. The hourly GHI data from HC3v5 and CRSv3 SDDs for each station location and for points around each station have been collected from the SoDa website (SoDa, 2018), based on the available period of ground data.

**Table 5.2:** Automatic stations with hourly GHI Vaisala QMS101 Pyranometer.

| Station | Coordinates (Degrees) | | Elevation a.s.l (m) | Period (dd/mm/yy) |
|---------|-----------|-----------|---------------------|-------------------|
| Halsho | 36.2097 N | 45.2598 E | 1105 | 01/01/2013–31/12/2016 |
| P1-Halsho | 36.2201 N | 45.2235 E | 1119 | |
| P2-Halsho | 36.2058 N | 45.3000 E | 1395 | |
| Bazian | 35.6021 N | 45.1376 E | 892 | 01/04/2014–30/12/2016 |
| P1-W Bazian | 35.6059 N | 45.0689 E | 872 | |
| P2-E Bazian | 35.5796 N | 45.1817 E | 828 | |
| Maydan | 34.9194 N | 45.6224 E | 330 | 01/01/2014–31/12/2016 |
| P1-Maydan | 34.9203 N | 45.5656 E | 388 | |
| P2-Maydan | 34.9182 N | 45.6716 E | 396 | |
| Kalar | 34.6244 N | 45.3049 E | 218 | 01/01/2014–31/12/2016 |
| P1-Kalar | 34.6220 N | 45.1768 E | 230 | |
| P2-Kalar | 34.6237 N | 45.4103 E | 210 | |

**Note:** The periods in the table are available for the ground measurements; the SDDs and TOA for the same periods have been collected for each station location and points around the stations, which were used for validation.
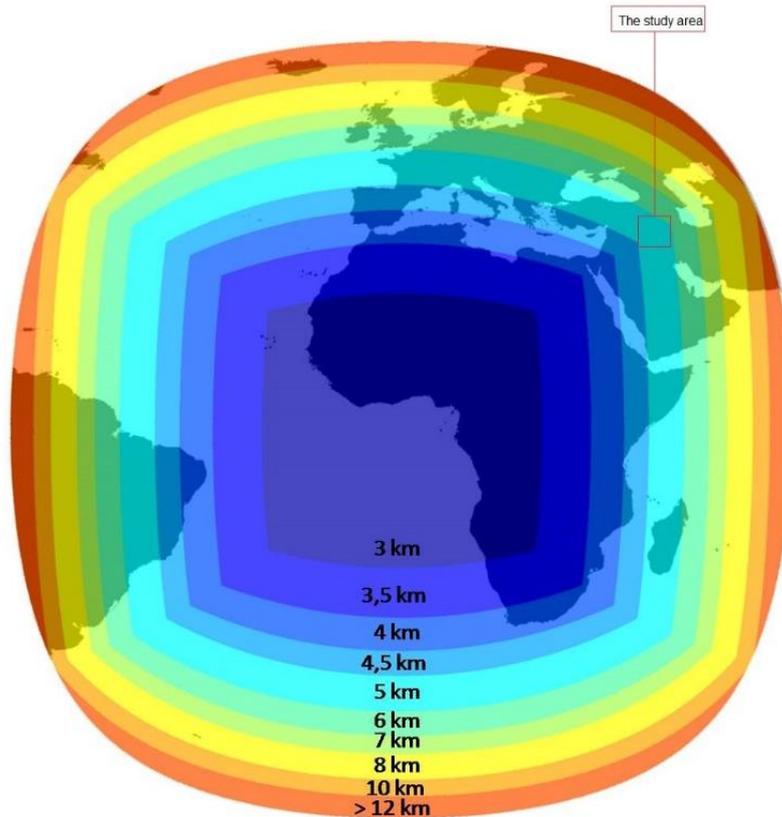
**Figure 5.1:** Spatial coverage of the SEVIRI instrument for MSG images (SoDa, 2018).

### 5.2.4.1 HelioClim-3 (HC3)

The HC3 dataset was created by converting Meteosat images to estimate the GHI for every 15 min since February 2001 using the original H2 method. The principle of H2 is to calculate solar irradiance statistically by the cloud cover index, which is created using the reflectance in the visible image of MSG and ground albedo (Moradi et al., 2009). The method has been modified several times by various inputs. It initially refers to Cano et al. (1986), and a new method was published in Rigollier et al. (2004). The MSG image processing in this model gives GHI. Then, DNI and DHI are estimated (Eissa et al., 2015a).

The most common versions of HC3 are v4 and v5. V4 inputs are the *Cs* model of the ESRA and the Linke turbidity factor (Thomas et al., 2016c). One limitation of this release is that it does not detect a local effect on the Linke turbidity factor (Thomas et al., 2016b). The *Cs* model gives solar irradiance globally as in the free cloudy-sky (Eissa et al., 2015b). HC3v5 works largely on the same principle as HC3v4, but is different because it uses the McClear Cs model (Marchand et al., 2017). McClear is a model for providing

solar irradiance under clear-sky conditions. It counts the optical depth of the atmosphere as a column, which contains aerosol, water vapour and ozone. It is provided by the Copernicus atmosphere monitoring service (Lefèvre and Wald, 2016). The data within those datasets are available online (SoDa, 2018) for MSG coverage for free from 2004 to 2006, and with payment from 2007 onwards.

### 5.2.4.2 Copernicus Atmosphere Monitoring Service (CAMS), Radiation Service (CRS)

CRS is a dataset of solar irradiance components, which provides H4 data using the satellite images of MSG. H4 is a modified method of their previous version. Cloud cover from APOLLO and ground albedo from MODIS and the McClear Cs model are used in this method (Eissa et al., 2015b; Thomas et al., 2016a; Thomas et al., 2016b). The data are available for free from 2004 until two days before for the areas covered by MSG images. The third version of CRS is available after bias correction (SoDa, 2018). This study has used CRS version 3 (CRSv3). Further information about the HC3v4-5 and CRSv3 projects can be found at (SoDa, 2018) and (Blanc et al., 2011b; Eissa et al., 2015a; Eissa et al., 2015b; Marchand et al., 2017; Thomas et al., 2016c; Thomas et al., 2016a; Thomas et al., 2016b).

### 5.2.5 Quality Control of GHI Measurements

Full information about the QC of the GHI ground data is described in Chapter (4). In addition, data cleaning were done for both SDDs and ground data by setting the solar elevation angle above 15°. Missing values were found and set as NA. The two datasets were harmonised for true local solar time. Systematic errors were removed from the data, and some questionable values of data according to various tests were not used in the validation process.

### 5.2.6 Validation Approach

The validation approach is illustrated in Figure 5.2. Most of the previous studies for validation of GHI SDD against ground data have separated data into all-sky and clear-sky conditions (Amillo et al., 2014; Eissa et al., 2015a; Schillings et al., 2004). The division also depends on the clearness index (Kt). The Kt is calculated by dividing hourly GHI ground data by TOA (Eke et al., 2017). The TOA was collected from SoDa (2018). For calculating the Kt of SDDs see Figure 5.2. The Kt was used for validation and setting

limits among the various sky conditions (Mueller et al., 2011; Amillo et al., 2014) as below:

- Clear-sky conditions: $0.65 < Kt \leq 1$
- Intermediate sky conditions: $0.3 < Kt \leq 0.65$
- Cloudy-sky conditions: $0 < Kt \leq 0.3$

This study separates the ground data into all-sky, clear-sky and cloudy-sky conditions based on the above Kt limits. This is to test the SDDs in various situations and to demonstrate under which situations the SDDs are the most accurate.

The approach uses the ground data of a station to assess the SDDs with data from the station location pixel and with another two points of SDD pixel data. One pixel data point is selected to the east and another is selected to the west of a station at a distance of 6–10 km, (Tables 5.1 and 5.2, Figure 5.3). This is to select a different pixel from the station location pixel; given the spatial resolution of MSG imagery is 5 km in the case study region (Figure 5.1). Hereafter, P1 is called the west point for each station, and P2 is called the east point. This is for further investigation into the validation of SDD for more than one-pixel around the station and to address whether the SDD values from neighbouring pixels are the same or different. This is because the solar irradiance intensity may be the same in an area of 25 km$^2$ (Bojanowski et al., 2013; Bouchouicha et al., 2016).

The validation performance between ground data and the SDDs, were evaluated for the all-sky conditions for hourly GHI for the stations and points around the stations and clearness index for all-sky conditions at stations, and the GHI for the clear-sky and cloudy conditions at the stations (Chapter 3, Section 3.2).

The performance of two SDDs against the ground data have also been assessed in all-sky conditions to demonstrate the variability within reproducing the ground data by SDDs by using the hourly mean and standard deviation of GHI in a month. The monthly mean and standard deviation of hourly GHI were calculated for ground data and SDDs for each month in the selected period of a station.

**Validation of hourly GHI for two Satellite-Derived Databases (SDDs)**
With **hourly GHI** quality controlled ground data of **nine** stations in **northeast Iraq**

**5 tower** stations (1-5) with ground data recorded by **Kipp and Zonen** Pyranometer.
**4 automatic** stations (6-9) with ground data recorded by **Vaisala** Pyranometer.

**HC3v5** SDD from MSG by Heliosat-2     **CRSv3** SDD from MSG by Heliosat-4

Taking a point pixel data at a location in the west of a station as P1

Taking a point pixel data at a location in the east of a station as P2

Taking the exact pixel point location data of a station

All-sky conditions     Clear-sky conditions     Cloudy-sky conditions

Cleanness index (Kt) All sky conditions

Comparison of hourly mean and standard deviation in a month between each SDD against available period of ground data for each station

The results are presented in tables by r, bias and RMSE

$$Kt(SDD1) = \frac{HC3v5}{TOA} \leftrightarrow Kt(GD) = \frac{GHI}{TOA}$$

$$Kt(SDD2) = \frac{CRSv3}{TOA} \leftrightarrow Kt(GD) = \frac{GHI}{TOA}$$

Further illustration for validation results with some scatterplots

Results are presented in figures

Bias corrections of SDDs (HC3v5 and CRSv3) in Cloudy-sky conditions

**Figure 5.2:** The flowchart of the approach in this chapter.
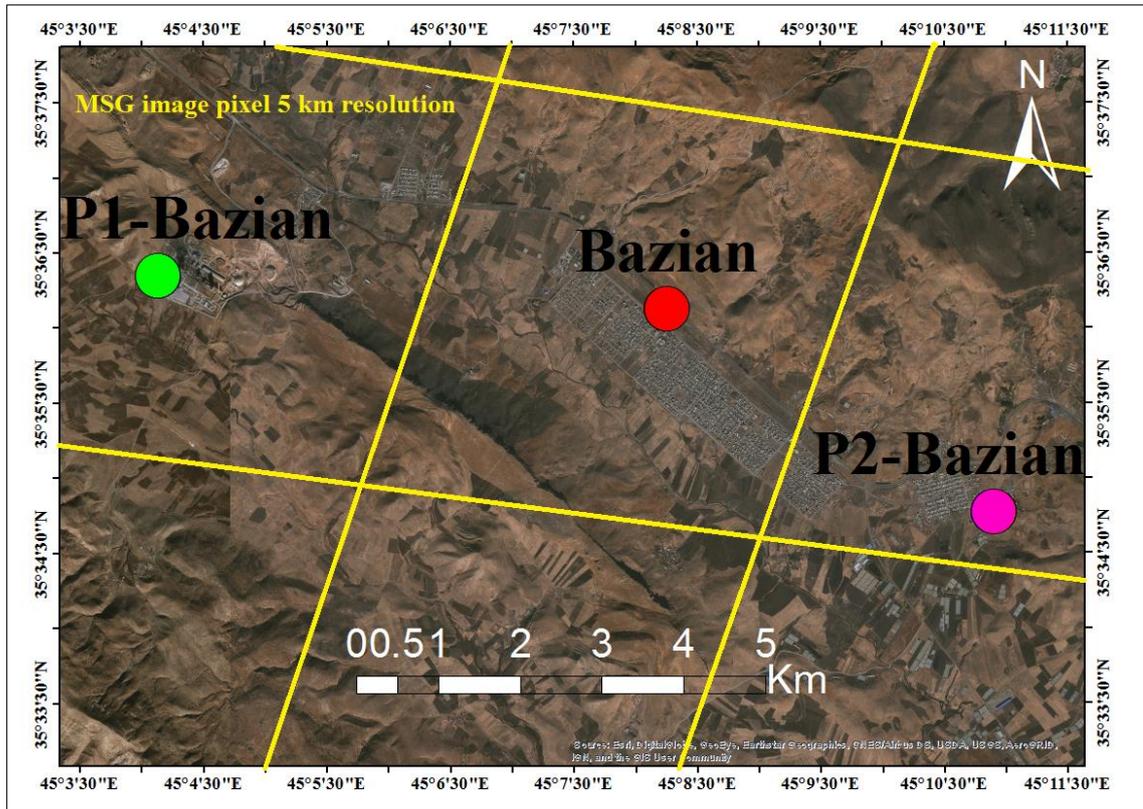
**Figure 5.3:** Example of point pixel selection of SDD around Bazian station.

### 5.2.7 Bias Correction for SDD

Several studies have shown an overestimation of bias for SDDs under cloudy-sky conditions (Ameen et al., 2018b; Amillo et al., 2014; Eissa et al., 2015b; Xia et al., 2017). Several other studies have investigated various methods to correct bias in SDDs (Cebecauer and Suri, 2010; Davy et al., 2016; Frank et al., 2018; Gueymard et al., 2012; Lange, 2018; Qu et al., 2014; Rincón et al., 2018; Vernay et al., 2013). For further detail, see Polo et al. (2016). A simple method for bias correction uses a short-term of ground measurements to correct long-term of SDDs data by identifying seasonal and systematic errors in SDDs. The simple method was applied by Polo et al. (2015) for bias correction of DNI from SDD under clear sky conditions in India. However, the simple method was used in this study for bias correction of GHI from SDD under cloudy-sky conditions in which uses a new test for the method in different climate region. The simple method is fully described by Polo et al. (2015).

The simple method utilises a linear regression equation to remove bias in the dataset. First, estimate SDD from the ground using Equation 5.1 to calculate coefficients in the

short-term dataset, such as one-year data or some months data. Second, it estimates new SDD in Equation 5.2 with bias removal from these data by fitted line to the $y=x$ line; it uses the coefficients calculated from Equation 5.1. Finally, it uses the newly created SDD with SDD in the short-term dataset to calculate the coefficients of the regression equation (Equation 5.3). The coefficients of Equation 5.3 will be used in Equation 5.4 to correct long-term SDD datasets that are available in a station or in nearby areas. An example of R codes applied is shown in Table 1B, Appendix B, which shows how corrected SDD is calculated based on the simple method with further explanation for steps. The coefficients and regression equations are also provided in Table 2B, Appendix B for each station for both SDD.

$$SDD1 = (a*GHI_{GD}) + b \tag{5.1}$$
$$New\ SDD = SDD1 - [(a - 1)\ GHI_{GD} + b] \tag{5.2}$$
$$Corrected\ SDD1 = (a*\ New\ SDD) + b \tag{5.3}$$
$$Corrected\ SDD2 = (a*SDD2) + b \tag{5.4}$$

where:

SDD1 is SDD in a short term

SDD2 is SDD in a long term

a and b are the coefficients of the regression are calculated in each step.

## 5.3    Results

Table 5.3 represents the results of the hourly GHI in all-sky conditions for the stations and the points around them; Table 5.4 represents the results of the clearness index in all-sky conditions for the stations; and Table 5.5 represents the results of the hourly GHI in clear-sky and cloudy-sky conditions for the stations.

Figures 5.4 and 5.5 show the results of the Monthly mean and standard deviation for each SDD with ground data. Figures 5.6 and 5.7 give further results between the stations with the points around them, and between SDDs in all-sky, clear-sky and cloudy-sky conditions for the results of the validation percentages of the bias and the RMSE, respectively. The results of some stations, as examples, with scatter plots are shown in Figures 5.8–5.10 (all other statins are in Appendix B, Figures B1–B6) for the GHI and

clearness index in all-sky conditions, and the GHI in clear-sky and cloudy-sky conditions. Table 5.6, Figures 5.11, and 5.12 show the results of bias correction for SDDs.

Overall, the study is focused on all-sky conditions to show the results in different ways such as within the clearness index, the mean and standard deviation in a month and other statistical indicators, which have been used in all three sky conditions. This is to avoid complex results when presenting all of the above data in a variety of sky conditions.

**Table 5.3:** Validation of hourly GHI under all-sky conditions for stations and points around them. Mean, Bias and RMSE units are W/m$^2$.

| Stations | Number of data | Mean | HC3v5 | | | | | CRSv3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | r | Bias | Bias % | RMSE | RMSE % | r | Bias | Bias % | RMSE | RMSE % |
| Batufa | 10,218 | 511 | 0.96 | −6 | −1.2 | 74 | 14 | 0.95 | −27 | −5.3 | 92 | 18 |
| P1 | 10,218 | 511 | 0.95 | −15 | −2.9 | 72 | 14 | 0.94 | −28 | −5.5 | 100 | 20 |
| P2 | 10,218 | 511 | 0.96 | −16 | −3.1 | 77 | 15 | 0.95 | −29 | −5.7 | 93 | 18 |
| Enjaksor | 13,622 | 518 | 0.97 | −4 | −0.8 | 64 | 12 | 0.95 | −20 | −3.9 | 85 | 16 |
| P1 | 13,622 | 518 | 0.96 | −9 | −1.7 | 75 | 14 | 0.94 | −19 | −3.7 | 90 | 17 |
| P2 | 13,622 | 518 | 0.97 | −6 | −1.2 | 64 | 12 | 0.95 | −21 | −4.1 | 86 | 17 |
| Hojava | 10,195 | 503 | 0.96 | 0 | 0 | 74 | 15 | 0.95 | −19 | −3.8 | 89 | 18 |
| P1 | 10,195 | 503 | 0.95 | 0 | 0 | 83 | 17 | 0.94 | −20 | −4 | 96 | 19 |
| P2 | 10,195 | 503 | 0.96 | 3 | 0.6 | 78 | 16 | 0.94 | −16 | −3.2 | 91 | 18 |
| Jazhnikan | 9856 | 518 | 0.96 | −13 | −2.5 | 73 | 14 | 0.95 | −20 | −3.9 | 82 | 16 |
| P1 | 9856 | 518 | 0.96 | −14 | −2.7 | 77 | 15 | 0.95 | −21 | −4.1 | 84 | 16 |
| P2 | 9856 | 518 | 0.96 | −12 | −2.3 | 73 | 14 | 0.95 | −21 | −4.1 | 83 | 16 |
| Tarjan | 10,261 | 521 | 0.96 | −20 | −3.8 | 74 | 14 | 0.95 | −26 | −5 | 81 | 16 |
| P1 | 10,261 | 521 | 0.96 | −21 | −4 | 78 | 15 | 0.95 | −26 | −5 | 83 | 16 |
| P2 | 10,261 | 521 | 0.96 | −20 | −3.8 | 73 | 14 | 0.95 | −26 | −5 | 80 | 15 |
| Halsho | 13,183 | 503 | 0.96 | 1 | 0.2 | 81 | 16 | 0.95 | 7 | 1.4 | 89 | 18 |
| P1 | 13,183 | 503 | 0.95 | 6 | 1.2 | 84 | 17 | 0.95 | 7 | 1.4 | 90 | 18 |
| P2 | 13,183 | 503 | 0.94 | 1 | 0.2 | 93 | 18 | 0.95 | 5 | 1 | 90 | 18 |
| Bazian | 8884 | 515 | 0.96 | −8 | −1.6 | 69 | 13 | 0.96 | −2 | −0.4 | 76 | 15 |
| P1 | 8884 | 515 | 0.96 | −6 | −1.2 | 74 | 14 | 0.95 | −4 | −0.8 | 79 | 15 |
| P2 | 8884 | 515 | 0.97 | −8 | −1.6 | 68 | 13 | 0.96 | −2 | −0.4 | 76 | 15 |
| Maydan | 9089 | 514 | 0.97 | −5 | −1 | 68 | 13 | 0.96 | 3 | 0.6 | 73 | 14 |
| P1 | 9089 | 514 | 0.96 | −6 | −1.2 | 72 | 14 | 0.96 | 1 | 0.2 | 75 | 15 |
| P2 | 9089 | 514 | 0.97 | −3 | −0.6 | 65 | 13 | 0.96 | 6 | 1.2 | 71 | 14 |
| Kalar | 7979 | 474 | 0.95 | 20 | 4.2 | 84 | 18 | 0.94 | 19 | 4 | 84 | 18 |
| P1 | 7979 | 474 | 0.94 | 19 | 4 | 88 | 19 | 0.93 | 19 | 4 | 88 | 19 |
| P2 | 7979 | 474 | 0.95 | 21 | 4.4 | 84 | 18 | 0.92 | 25 | 5.3 | 97 | 20 |

**Table 5.4**: Validation of hourly GHI under all-sky conditions for the clearness index.

| Stations | Number of data | Mean | HC3v5 Clearness Index | | | | | CRSv3 Clearness Index | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | r | Bias | Bias % | RMSE | RMSE % | r | Bias | Bias % | RMSE | RMSE % |
| Batufa | 10,218 | 0.602 | 0.89 | −0.009 | −1.5 | 0.102 | 16.94 | 0.85 | −0.038 | −6.3 | 0.121 | 20.1 |
| Enjaksor | 13,622 | 0.611 | 0.89 | −0.01 | −1.64 | 0.089 | 14.57 | 0.83 | −0.032 | −5.2 | 0.117 | 19.1 |
| Hojava | 10,195 | 0.592 | 0.87 | −0.004 | −0.68 | 0.1 | 16.89 | 0.85 | −0.03 | −5.0 | 0.116 | 19.5 |
| Jazhnikan | 9856 | 0.602 | 0.86 | −0.024 | −3.99 | 0.1 | 16.61 | 0.85 | −0.031 | −5.1 | 0.107 | 17.7 |
| Tarjan | 10,261 | 0.612 | 0.85 | −0.034 | −5.56 | 0.103 | 16.83 | 0.84 | −0.04 | −6.5 | 0.109 | 17.8 |
| Halsho | 13,183 | 0.584 | 0.87 | −0.003 | −0.51 | 0.109 | 18.66 | 0.84 | 0.008 | 1.3 | 0.12 | 20.5 |
| Bazian | 8884 | 0.593 | 0.87 | −0.014 | −2.36 | 0.093 | 15.68 | 0.85 | −0.006 | −1.0 | 0.098 | 16.5 |
| Maydan | 9089 | 0.594 | 0.87 | −0.016 | −2.69 | 0.091 | 15.32 | 0.86 | −0.003 | −0.5 | 0.092 | 15.4 |
| Kalar | 7979 | 0.565 | 0.83 | 0.013 | 2.3 | 0.099 | 17.52 | 0.82 | 0.017 | 3.0 | 0.097 | 17.1 |
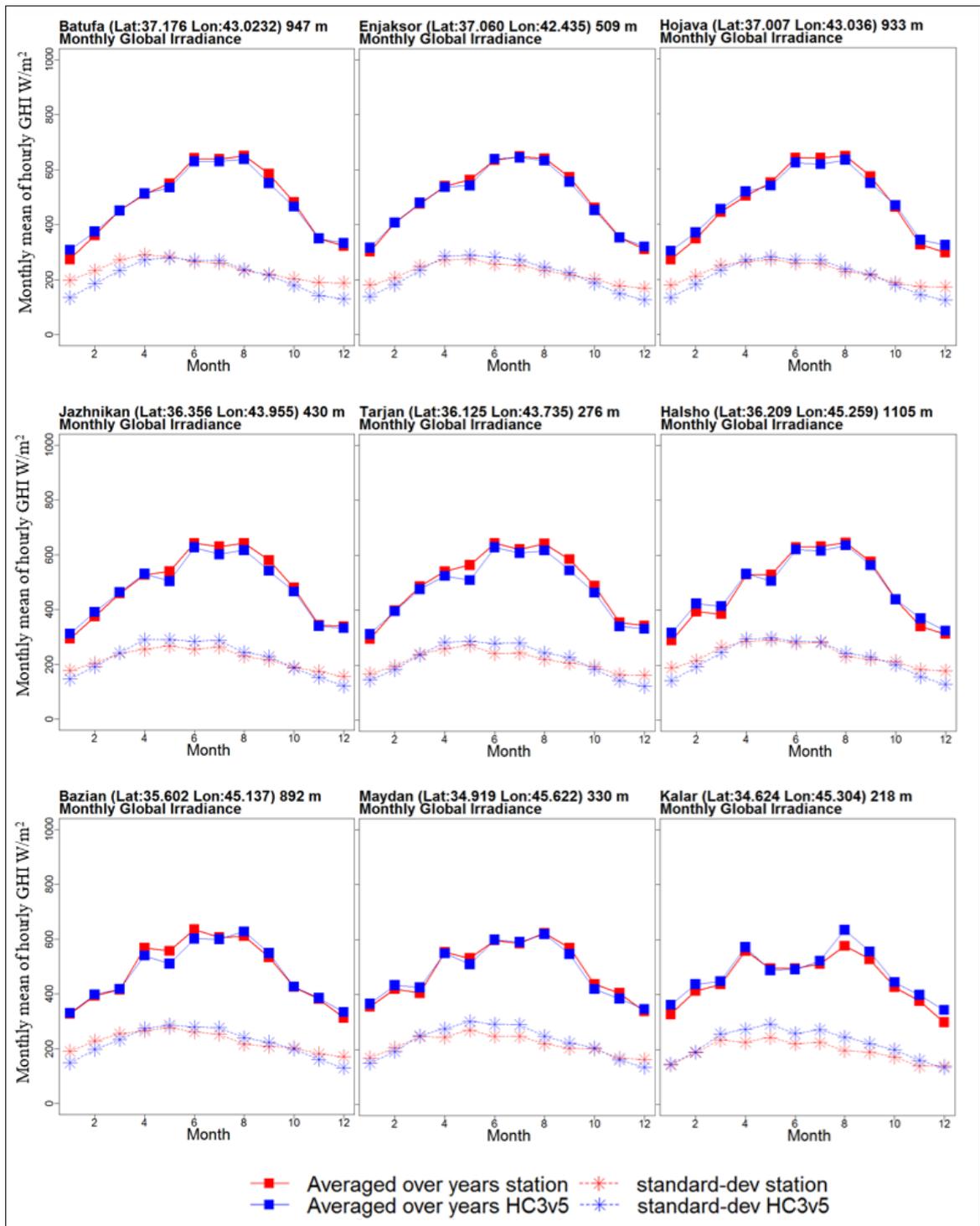
**Figure 5.4:** Monthly mean and standard deviation of hourly GHI data in each month aggregated over the data availability for each station with HC3v5. The difference between dots reveals the errors in a month and vice versa. If the dot of the SDD in a month is above the dot of the ground data, it denotes overestimation; otherwise, it denotes underestimation.
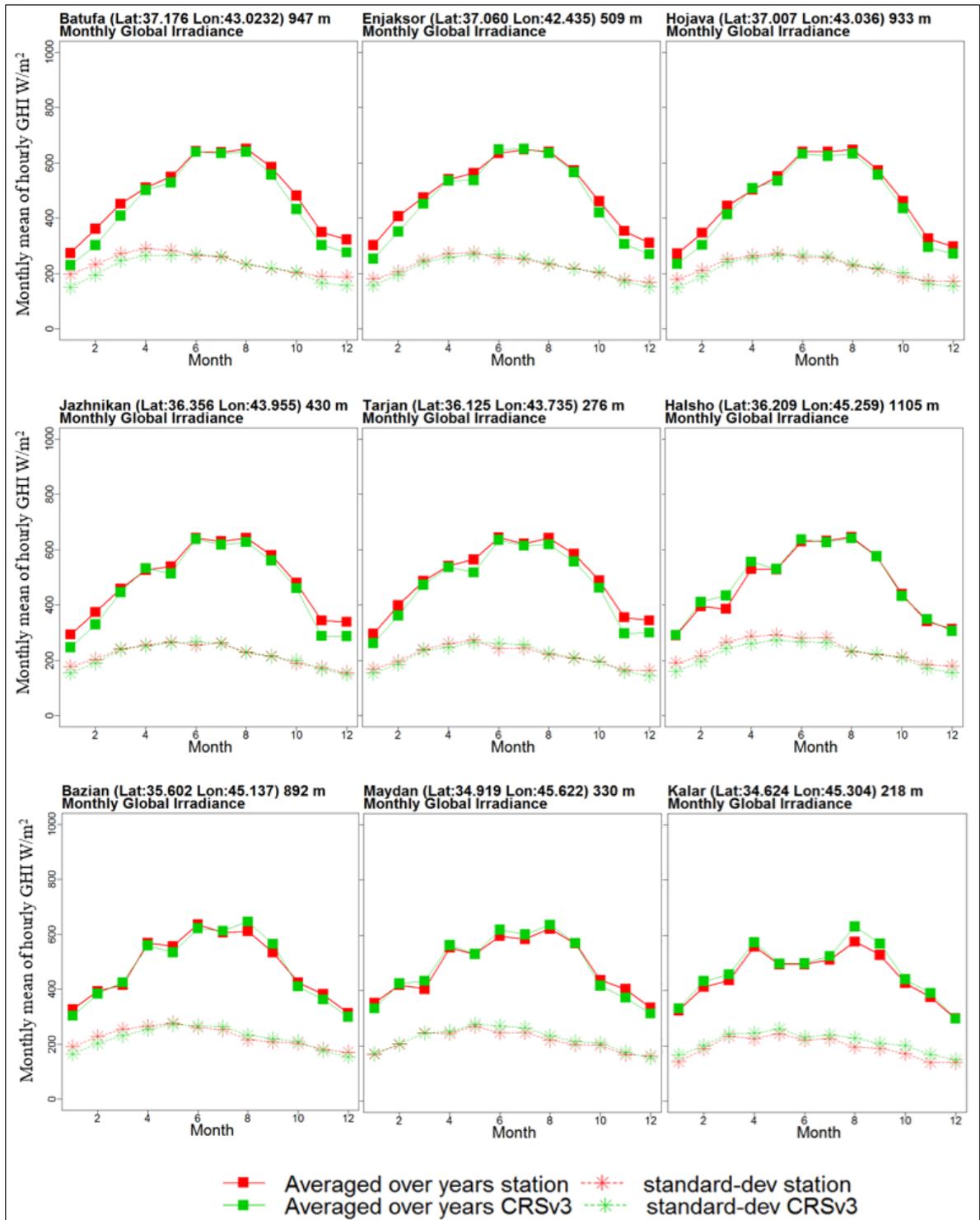
**Figure 5.5:** Monthly mean and standard deviation of hourly GHI data in each month aggregated over data availability for each station with CRSv5. The difference between dots reveals errors in a month and vice versa. If the dot of the SDD in a month is above the dot of the ground data, it denotes overestimation; otherwise, it denotes underestimation.
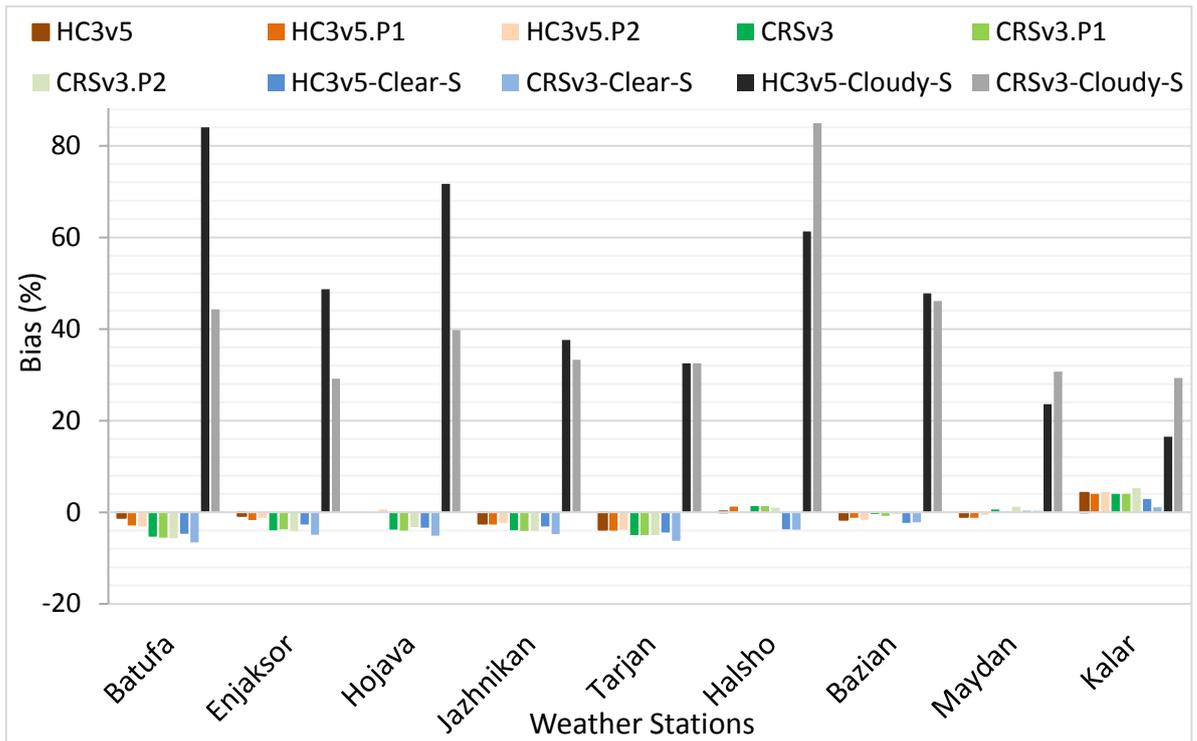
**Figure 5.6:** Comparison of bias (%) for the hourly GHI for all-sky conditions among stations with points around them for HC3v5 and CRSv3. Clear-skies and cloudy-skies at stations are represented by blue, light blue, black and grey colours, respectively.
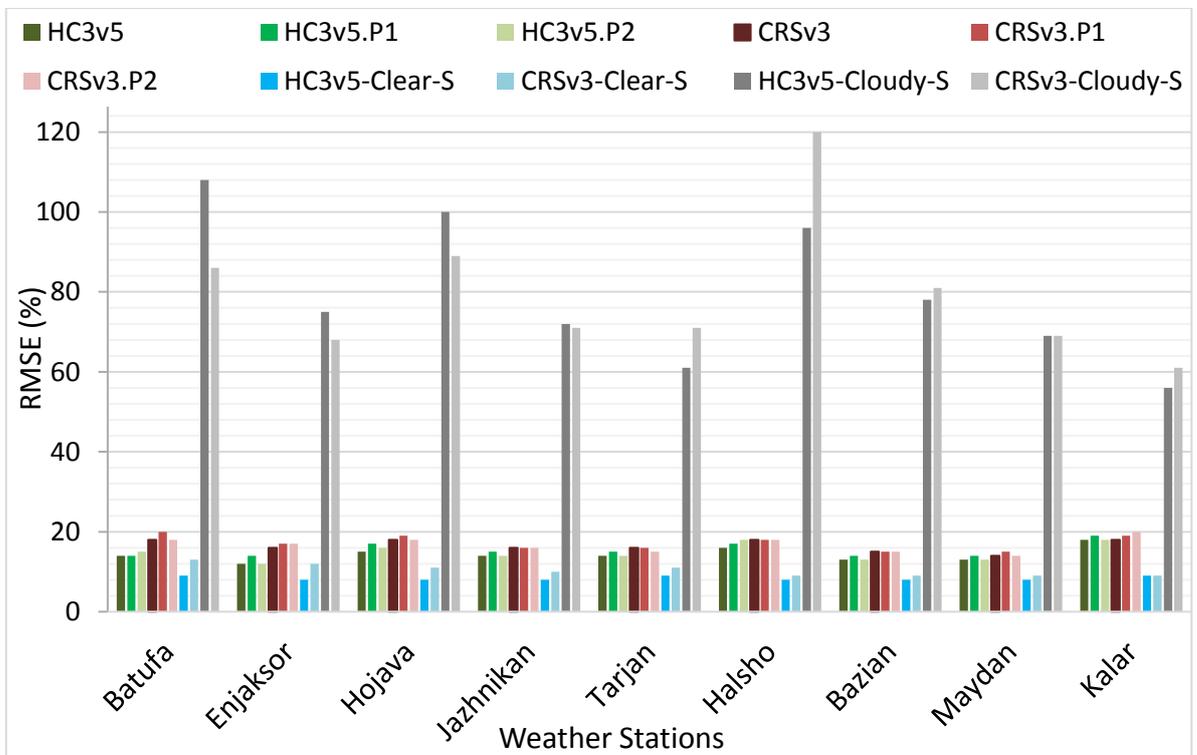


**Figure 5.7:** As in Figure 5.6, but for RMSE (%).

**Table 5.5:** Validation of hourly GHI under clear-sky and cloudy-sky conditions. Mean, Bias and RMSE units are W/m$^2$.

| Station | Condition | Number of Data | Mean | HC3v5 | | | | | CRSv3 | | | | |
|---------|-----------|----------------|------|-------|------|-----------|------|-----------|-------|------|-----------|------|-----------|
| | | | | $r$ | Bias | Bias % | RMSE | RMSE % | $r$ | Bias | Bias % | RMSE | RMSE % |
| Batufa | Clear-sky | 5937 | 679 | 0.97 | −32 | −4.7 | 58 | 9 | 0.95 | −45 | −6.6 | 86 | 13 |
| | Cloudy-sky | 1448 | 106 | 0.68 | 89 | 84 | 114 | 108 | 0.63 | 47 | 44.3 | 91 | 86 |
| Enjaksor | Clear-sky | 7955 | 672 | 0.97 | −18 | −2.7 | 53 | 8 | 0.95 | −33 | −4.9 | 78 | 12 |
| | Cloudy-sky | 1507 | 113 | 0.68 | 55 | 48.7 | 85 | 75 | 0.64 | 33 | 29.2 | 77 | 68 |
| Hojava | Clear-sky | 5669 | 672 | 0.97 | −23 | −3.4 | 56 | 8 | 0.95 | −34 | −5.1 | 77 | 11 |
| | Cloudy-sky | 1365 | 113 | 0.64 | 81 | 71.7 | 113 | 99 | 0.59 | 45 | 39.8 | 101 | 89 |
| Jazhnikan | Clear-sky | 5513 | 681 | 0.97 | −21 | −3.1 | 54 | 8 | 0.96 | −33 | −4.8 | 69 | 10 |
| | Cloudy-sky | 1018 | 117 | 0.63 | 44 | 37.6 | 84 | 72 | 0.69 | 39 | 33.3 | 83 | 71 |
| Tarjan | Clear-sky | 5983 | 666 | 0.97 | −29 | −4.4 | 62 | 9 | 0.96 | −41 | −6.2 | 72 | 11 |
| | Cloudy-sky | 965 | 120 | 0.68 | 39 | 32.5 | 73 | 61 | 0.68 | 39 | 32.5 | 85 | 71 |
| Halsho | Clear-sky | 7498 | 677 | 0.97 | −25 | −3.7 | 56 | 8 | 0.97 | −26 | −3.8 | 61 | 9 |
| | Cloudy-sky | 2083 | 106 | 0.61 | 65 | 61.3 | 102 | 96 | 0.6 | 90 | 84.9 | 127 | 120 |
| Bazian | Clear-sky | 4673 | 690 | 0.97 | −16 | −2.3 | 54 | 8 | 0.96 | −15 | −2.2 | 64 | 9 |
| | Cloudy-sky | 937 | 115 | 0.64 | 55 | 47.8 | 90 | 78 | 0.65 | 53 | 46.1 | 93 | 81 |
| Maydan | Clear-sky | 4729 | 678 | 0.97 | 2 | 0.3 | 55 | 8 | 0.96 | 2 | 0.3 | 62 | 9 |
| | Cloudy-sky | 813 | 127 | 0.51 | 30 | 23.6 | 88 | 69 | 0.65 | 39 | 30.7 | 88 | 69 |
| Kalar | Clear-sky | 2755 | 659 | 0.96 | 19 | 2.9 | 57 | 9 | 0.95 | 7 | 1.1 | 60 | 9 |
| | Cloudy-sky | 686 | 133 | 0.64 | 22 | 16.5 | 74 | 56 | 0.72 | 39 | 29.3 | 81 | 61 |
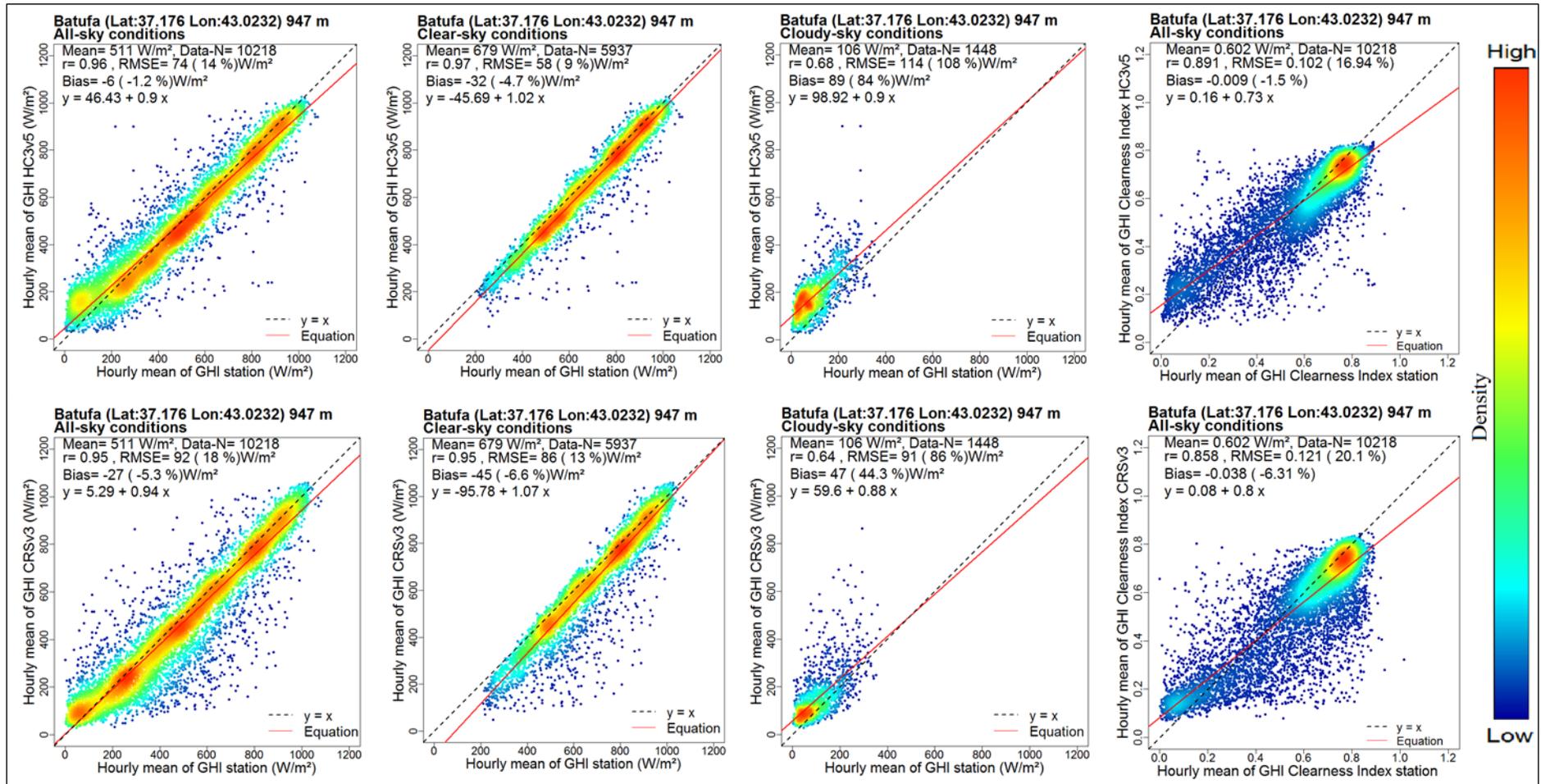
**Figure 5.8:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Batufa station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.
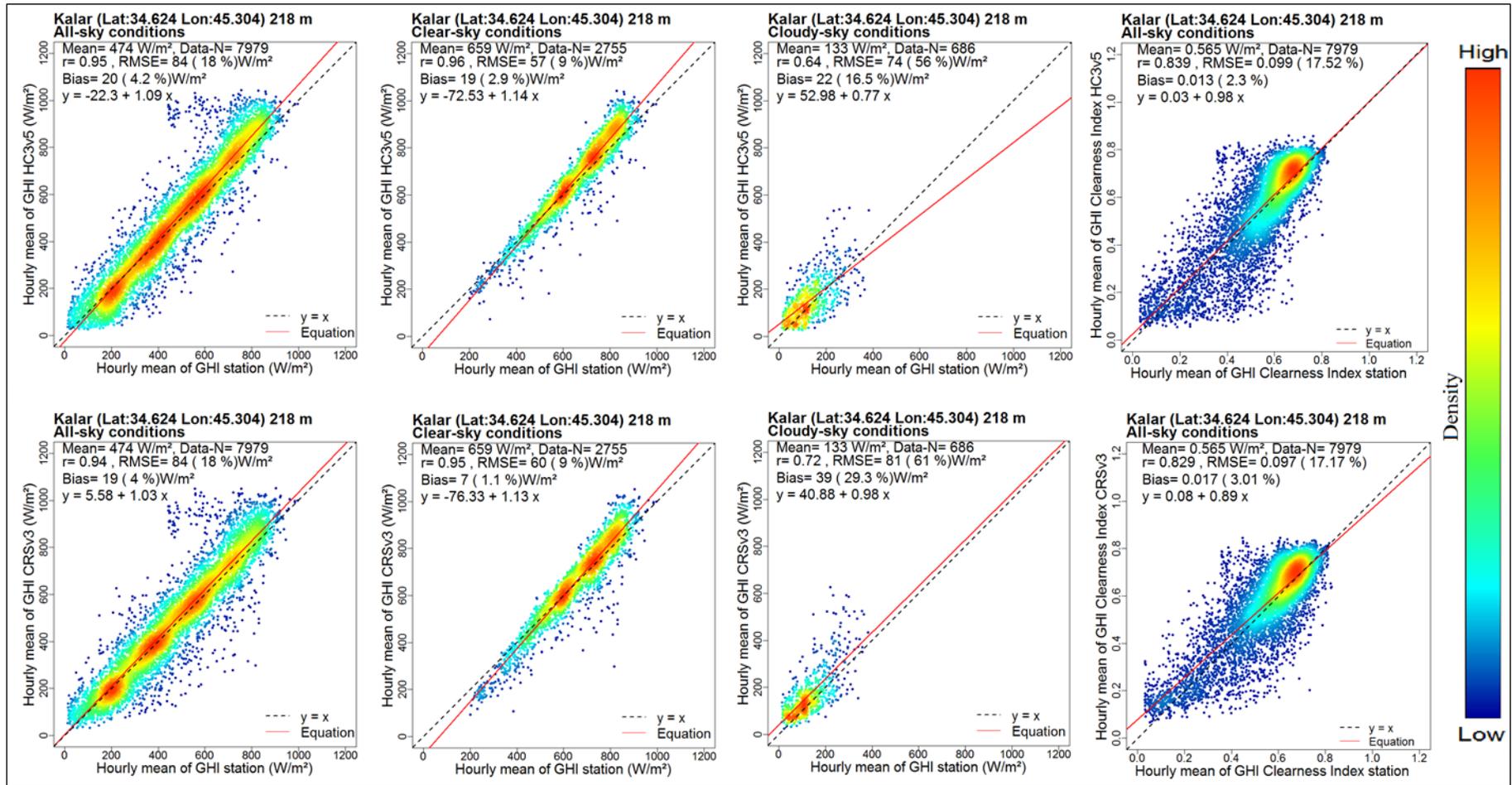
**Figure 5.9:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Kalar station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.
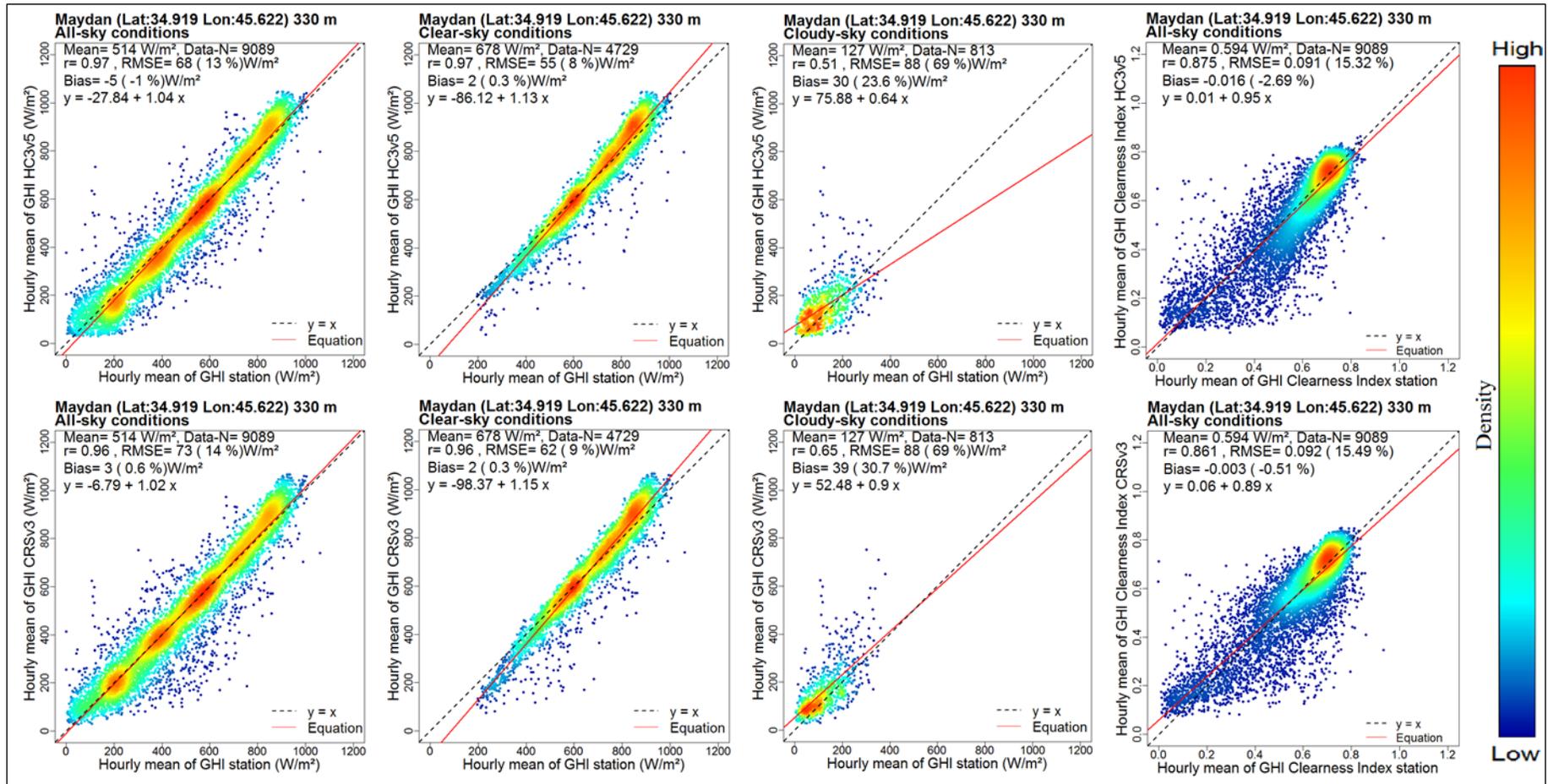
**Figure 5.10:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Maydan station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.

### 5.3.1 All-Sky Conditions

The results of the validation for the all-sky conditions are presented in Table 5.3. The range $r$ was 0.94–0.97 in all of the stations and 0.92–0.97 in the points around them. Interestingly, zero bias was recorded at Hojava station, and it was near zero in several other cases (Halsho, Maydan and Bazian stations) for both SDDs. Negative (underestimation) bias was recorded in several cases. It ranges from $-21$ W/m$^2$ ($-4\%$) to $-3$ W/m$^2$ ($-0.6\%$) for HC3v5, which is lower than CRSv3 in most cases, which ranges from $-27$ W/m$^2$ ($-5.3\%$) to $-2$ W/m$^2$ ($-0.4\%$). Moreover, bias (%) was lower than $-6\%$ for all of the stations. However, a positive (overestimation) bias was recorded in some of the cases. The highest case of bias in the chapter was recorded at Kalar station, which was 25 W/m$^2$ (5.3%) for CRSv3 and 21 W/m$^2$ (4.4%) for HC3v5. Some other positive rates were recorded at Halsho and Maydan stations, which were lower than 2% for both SDDs (Figure 5.6, Table 5.3).

The bias at each station was compared to the points around the station, and was nearly the same with no more than 4% difference for each station. Overall, the rate of bias in HC3v5 was less than that in CRSv3 (Figure 5.6, Table 5.3).

The RMSE was under 21% in all of the cases. Its lowest value, at Enjaksor, was 64 W/m$^2$ (12%), and increased to the highest value of 88 W/m$^2$ (19%) at Kalar station for HC3v5. It was generally high for CRSv3 ranging from 71 W/m$^2$ (14%) to 100 W/m$^2$ (20%). Most of the other rates of RMSE were between 14–18% for both SDDs.

The RMSE for the points around the stations compared to the station location are nearly the same (Figure 5.7). Overall, the RMSE for HC3v5 was less than that for CRSv3 (Figure 5.7).

The smooth scatter density plot illustrates the residual and correlation between ground data and SDDs in some cases. For example, Figure 5.8 for Batufa station shows that the density of observations were mostly under the 1:1 (y=x) line, which indicated a recorded negative bias, and the RMSE was acceptable for HC3v5 and high for CRSv3, while some of the other values above the 1:1 line are under 200W/m$^2$. However, Figure 5.9 (Kalar station) shows that the majority of observations are above the 1:1 line and some

points are far from the line. This corresponds to positive bias and high RMSE in the station.

Low rates of bias and RMSE were recorded at Maydan station, which is shown in Figure 5.10 for HC3v5 and CRSv3 respectively. The best-fit line in red is nearly the same as the 1:1 line, more so for HC3v5 than CRSv3 (at the same station).

The results of the clearness index in the all-sky conditions are represented in Table 5.4. The percentages of bias and RMSE were quite similar to the GHI. Hence, the $r$ values at all of the stations were lower than the GHI, which ranged from 0.82–0.89. The $r$ values were higher for HC3v3 than CRSv3 when comparing each station for both. The scatter density plot shows the highest density of observations at around 0.7 for all of the stations. Negative bias at Batufa, Maydan and positive bias at Kalar can be seen, although the values were low. Several values are far from the 1:1 line, resulting in RMSE to be from 15%–21% at all of the stations (Figures 5.8–5.10).

Results of the monthly mean and standard deviation of the GHI are represented in Figure 5.4 for HC3v5 and in Figure 5.5 for CRSv3. The two figures demonstrate the distribution of the two SDDs with ground data in each month, expressed by the standard deviation. However, some differences were recorded in the winter months at Batufa, Hojava, Halsho and Bazian stations, whereas for summer months differences were recorded at Maydan and Kalar stations for both SDDs.

### 5.3.2 Clear-Sky and Cloudy-Sky Conditions

The compared results of the two SDDs for clear-sky and cloudy-sky conditions are presented in Table 5.5. There are apparent differences in the $r$ values between the clear-sky and cloudy-sky conditions in all of the cases. The was 0.95–0.97 for clear-skies, whereas for the cloudy-skies, it was 0.51–0.72 .

Similarly, the ranges of bias were much higher in the cloudy-skies than in the clear-skies. The lowest bias was a 2 W/m$^2$ (0.3%) overestimation at Maydan under clear-sky conditions whereas in the same station it reached 30 W/m$^2$ (24%) under cloudy-sky conditions. The highest bias for the clear-sky conditions was recorded at Batufa station, which was −32 W/m$^2$ (−4.7%). The same station had the highest bias for the cloudy-sky conditions, which was 89 W/m$^2$ (84%). These were recorded for HC3v5. The bias for

CRSv3 was the same as HC3v3 for the low range in the clear-sky conditions. In others, the ranges start from −45 W/m$^2$ (−6.6%) underestimation at Batufa station to 7 W/m$^2$ (1%) overestimation at Kalar station, while for cloudy-skies, it ranged from 33 (29%) to 90 (85%) W/m$^2$ respectively. The variety of the two SDDs in term of bias is shown in Figure 5.6, which illustrates a moderate difference between cloudy-sky and clear-sky conditions from one station to the other. In addition, the range of bias was much lower in clear-skies than cloudy-skies and the bias in all-sky conditions was lower than that in clear-skies except for two stations for both SDDs where the bias was higher to some degree in the all-sky conditions than in the clear-skies (Figure 5.6).

Similarly, the RMSE was much higher under cloudy-skies than under clear-skies for the SDDs in each station. For example, at Halsho station for CRSv3 it was 61 W/m$^2$ (9%) in clear-skies and increased sharply to 127 W/m$^2$ (120%) under cloudy conditions. Nearly the same situation can be seen for Batufa station for HC3v5. The RMSE for both SDDs was lower than 14% at all of the stations for the clear-skies while it was above 58% for the cloudy-skies (Figure 5.7, Table 5.5). The RMSE in clear-skies was lower than in the case of all-sky conditions in all of the study areas for both SDDs (Figure 5.7).

The smooth scatter density plot for Batufa, Kalar and Maydan of both SDDs separately shows that the density of observations were above the 1:1 line, and the direction of distribution was towards the high value of SDD which recorded high overestimation and high RMSE in cloudy-sky conditions (Figures 5.8–5.10). In contrast, in the clear-sky conditions for nearly all of the stations, the distribution of observations was near the 1:1 line. This leads to low RMSE in clear-skies compared to cloudy-sky and all-sky conditions (Figures 5.8–5.10). The observations under the 1:1 line illustrated a negative bias at Batufa station, whereas the opposite—i.e. positive bias—occurred at Kalar station, and a minimal bias was seen at Maydan station relative to the normal distribution.

Overall, the results of the validation varied from one station to another. However, they are all acceptable according to bias (under 3%) and RMSE (under 14%) for hourly data. At some stations, the results were disappointing. The points around the stations had nearly the same ranges of bias and RMSE compared to the station location. In most of the cases, the bias and RMSE of HC3v5 were lower than CRSv3. The bias and RMSE were lower for the clear-sky and all-sky conditions than for the cloudy-skies.

### 5.3.3    Results of Bias Corrections for SDDs under Cloudy-Sky Conditions

Statistical results of bias correction under Cloudy-Sky Conditions for SDDs in the case study area are presented in Table 5.6, whereas the comparison of relative bias and relative RMSE before and after corrections are demonstrated in Figures 5.11 and 5.12 respectively. Slight improvement is noted for each of the individual cases of relative bias at all nine stations, whereas there is a narrow improvement for relative RMSE and with no improvement is noted at three stations for relative RMSE among eighteen cases. The bias removal rates were recorded between 10% to 80% from one station to another. For instance, bias reduced from 84% to 5% at Hojava, and it reduced from 16.5% to 5.3% at Kalar for HC3v5. It reduced to near zero percentage at Hojava and Bazian for HC3v5 and at Enjaksor for CRSv3. The overall improvements of relative RMSE after the bias correction ranged from 2% to 33% except three negative cases at all stations for HC3v5 and CRSv3.

**Table 5.6:** Statistical results of hourly GHI bias corrections for SDDs at nine stations. Bias and RMSE units are W/m$^2$.

| Stations | HC3v5-Corected | | | | CRSv3-Corected | | | |
|---|---|---|---|---|---|---|---|---|
| | Bias | Bias % | RMSE | RMSE % | Bias | Bias % | RMSE | RMSE % |
| Batufa | 5.19 | 4.96 | 78.8 | 75.42 | -4 | -3.9 | 80.4 | 80.7 |
| Enjaksor | 4.1 | 3.6 | 71.6 | 64.1 | -1 | -0.9 | 72.44 | 64.8 |
| Hojava | 0.8 | 0.7 | 74.6 | 66.1 | -10.7 | -9.5 | 84.2 | 75.09 |
| Jazhnikan | 14.5 | 12.6 | 83.8 | 72.9 | 4.8 | 4.2 | 71.9 | 62.64 |
| Tarjan | 12.3 | 10.6 | 68.3 | 59 | -16.1 | -13.9 | 72.9 | 63 |
| Halsho | -6.7 | -6.3 | 82.1 | 77.8 | -4.9 | -4.6 | 86.9 | 82.3 |
| Bazian | 0.05 | 0.04 | 77.7 | 67.2 | 4.4 | 3.8 | 79.48 | 68.6 |
| Maydan | 7.7 | 6.2 | 92.53 | 74.3 | 16.9 | 13.6 | 88.9 | 71.4 |
| Kalar | 7.3 | 5.3 | 83.03 | 60.39 | 4.7 | 3.4 | 73.3 | 53.34 |
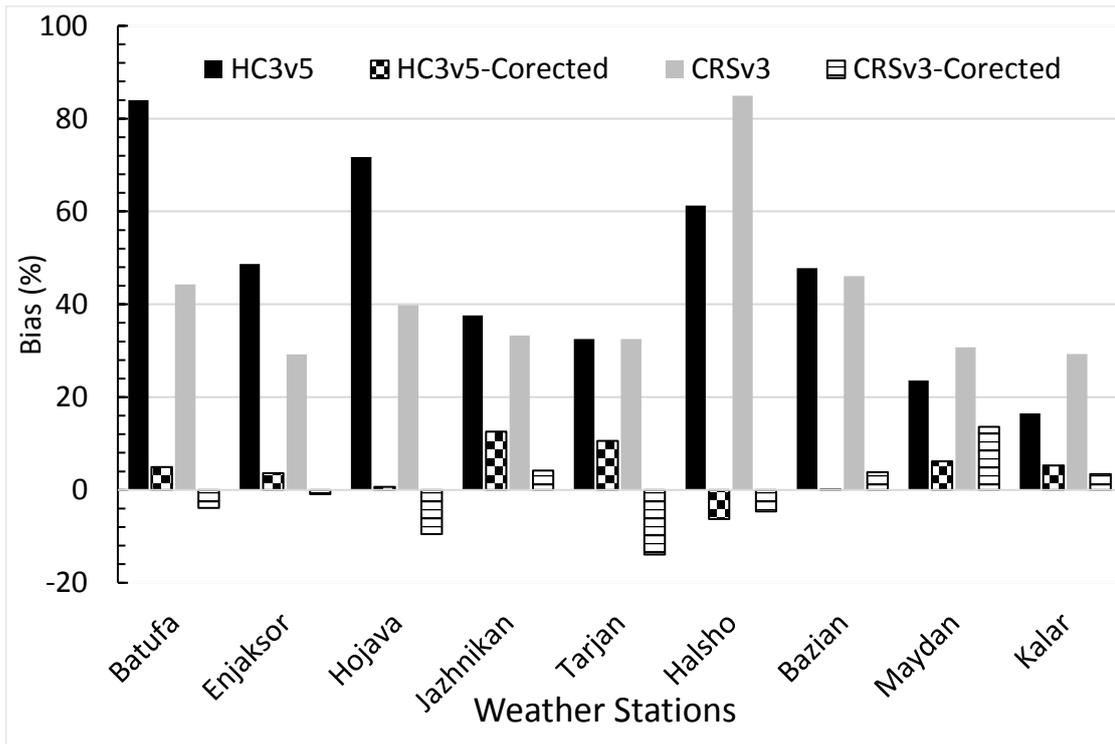
**Figure 5.11:** Comparison of bias (%) before and after corrections for SDDs.
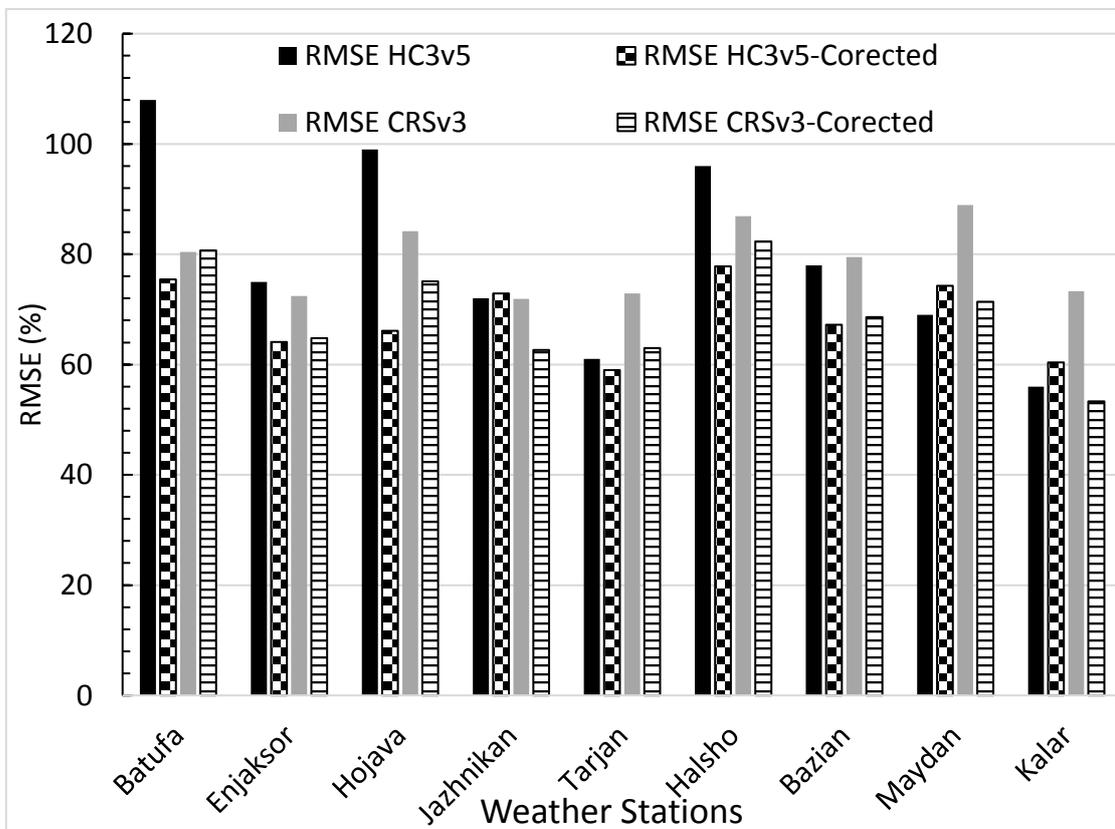


**Figure 5.12:** Comparison of RMSE (%) before and after corrections for SDDs.

## 5.4 Discussion

The validation results demonstrate good agreement between the ground data and SDDs in all-sky and clear-sky conditions (average $r = 0.95$, bias under 6% and RMSE under 21%), unlike the results for the cloudy-sky conditions (average $r = 0.61$, bias above 16% and RMSE above 61%). The results from the two neighbouring points at each station are close to the results at the station location with an average difference of 2%. Overall the performance of SDDs are in agreement with those from similar studies in other areas (Eissa et al., 2015a; Marchand et al., 2017; Thomas et al., 2016a; Thomas et al., 2016b), which also showed a better performance for HC3v5 over CRSv3 (Figure 5.6 and 5.7). This is mainly related to the models for creating each dataset, whether it is H2 or H4 (see this Chapter, Section 2.4). The results of bias correction demonstrate clear improvement of SDDs under cloudy-sky conditions.

### 5.4.1 All-Sky Conditions

The high rates of positive bias and RMSE at Kalar compared to all of the other stations (Figures 5.6–5.10) might be due to the quality of the recorded data (Chapter 4) and a partial shadow on the sensor at that station, because its mean is lower than that at the other stations by nearly 30 W/m$^2$ (Table 5.3), although those data did pass the quality control. A similar positive bias and RMSE are reported by other studies (Eissa et al., 2015a; Marchand et al., 2017). Low rates of bias (0–2%) were recorded for HC3v5 Hojava, Halsho, Bazian and Maydan, and for CRSv3 for the same stations except for Hojava, whose bias rate reaches −3.8% for all-sky conditions (Table 5.3, Figure 5.6). These show that the GHI ground data are explained well by satellite data (Figures 5.8–5.10 and B1–B6). Underestimations of bias (1–6%) and RMSE (12–20%) were recorded at most stations (Table 5.3, Figures 5.6 and 5.7). Comparable percentages were recorded in similar studies in other areas and climate regions, namely Egypt (Eissa et al., 2015a), Brazil (Thomas et al., 2016a), and some BSRN stations (Thomas et al., 2016b). The reasons are partly related to the local condition of the station, inputs to the Heliosat method—especially the atmospheric optical depth owing to its unavailability—, various cloud types, the resolution of satellite images and the aerosols effect (Kaskaoutis et al., 2016; Thomas et al., 2016a; Xia et al., 2017). This is because in some cases, the GHI ground data are well explained by SDDs, but in other cases only some error rates were recorded (Table 5.3). Those rates are quite reasonable for hourly GHI (Marchand et al.,

2017). Those rates of bias can be corrected or modified in some ways (Frank et al., 2018; Polo et al., 2015; Polo et al., 2016; Qu et al., 2014).

The low variabilities between the ground data and both SDDs are seen in Figures 5.4 and 5.5. This might be related to the geographical location and climatic condition, and another reason is that the data were aggregated, concealing some random errors between the two datasets. Hence, some error rates in winter months are related to the difficulty of the Heliosat methods to estimate GHI in cloudy conditions (Schillings et al., 2004; Blanc et al., 2011b). The performance of the clearness index (Table 5.4) is nearly the same for GHI in at all-sky conditions.

An interesting side of this study is that the results of the validation for both SDDs with the two neighbouring points at each station separately are slightly closer to those at the station location. The differences range between 0%–2% for bias and RMSE for each point at most of the locations (Table 5.3, Figures 5.6 and 5.7). The ±1%–4% difference between the station location and the neighbouring points with the station ground data GHI are mainly related to the elevation above sea level for each location. Other factors might be related to local land surface types such as land and water, agriculture and bare soil. This indicates that the GHI from SDDs can be used for regional planning for various purposes, and the ground data GHI can be used for neighbouring areas when there is a limitation of ground data. This validation is also considered to add further weight to the use of solar irradiance data for neighbouring area or interpolating ground data in a 25 km$^2$ area (Bojanowski et al., 2013; Bouchouicha et al., 2016; Janjai et al., 2011).

### 5.4.2 Clear-Sky and Cloudy-Sky Conditions

The validation results for both SDDs with the ground data for clear-sky conditions showed good agreement according to RMSE, which decreased at most stations (Figures 5.6–5.10). This is partly related to the inputs to the H2 method, especially in incorporating the visible images of MSG in cloud-free conditions into the model. Similarly, the increased performance of HC3v5 for clear-sky conditions has been reported (Eissa et al., 2015a) in Egypt. However, the remaining residuals of clear-sky conditions are caused by the factors that have been mentioned in the all-sky conditions above. However, the bias increased to some degree for both SDDs in most of the stations, which were recording underestimations for clear-sky conditions. This is partially related to the increase of the

mean GHI ground data in clear-sky conditions. It has also been recorded by several studies (Amillo et al., 2014; Zhang et al., 2016; Xia et al., 2017), which show that the bias is underestimated for clear-skies.

The study investigated the performance of SDDs on cloudy-sky conditions, reflected in the low performance of both HC3v5 and CRSv3 according to the high ratio of bias and RMSE (Figures 5.6 and 5.7). A close look at the samples of smooth scatter plots (Figures 5.8–5.10) shows how far the observations and their density are from the 1:1 line. This is related to difficulties in analysing cloudy pixels of MSG images (Schillings et al., 2004); the clouds prevented the ground being viewed from the sensor aboard the satellite (Marchand et al., 2017), and as such it is hard to differentiate between cloud albedo and ground albedo (Blanc et al., 2011b). These factors lead to a general overestimation of GHI as shown in all of the stations (Table 5.5) for bias, and much higher RMSE (Figures 5.5 and 5.6). Indeed, in some of the cases, it is above the mean of the observations. Similar high residuals for cloudy conditions have been reported in the literature (Amillo et al., 2014; Eissa et al., 2015a; Zhang et al., 2016; Xia et al., 2017). This indicates that the GHI ground data are well explained by the SDDs in clear-sky conditions, whereas they are not explained well in cloudy-sky conditions. The results of high bias and RMSE indicate that further research is required to correct the errors under cloudy-sky conditions, whereas several studies have done bias corrections for all-sky conditions (Frank et al., 2018; Polo et al., 2015; Polo et al., 2016; Qu et al., 2014).

### 5.4.3   Bias Corrections for SDDs under Cloud-Sky Conditions

The correction approach applied here substantially reduces the bias and enhances the SDDs under cloudy-sky conditions. This demonstrates the importance of applying bias correction approaches to improve the bankability of GHI data from SDDs. The various rates of improvement for relative bias among cases are related to the original SDDs, and local climate condition and landscape variability, also reported by the literature (Frank et al., 2018; Lange, 2018; Rincón et al., 2018). The low improvement of RMSE compared to bias after correction is in an agreement with similar studies (Polo et al., 2015; Rincón et al., 2018). These low improvements of RMSE could be associated with stochastic errors related to numerical random calculations.

The limitations of this study in this chapter are the different data timescales from one station to another and the limited information available for some parameters, such as the aerosols and local atmospheric properties. This might lead to a challenge to fully explain the reasons behind the results at each station. The simple bias correction method required short-term high quality ground measurement to capture the errors in SDDs over the long-term and this is a limitation of the method.

The validation results vary from one station to another, they are near the WMO standard, in which the bias should be less than 3 W/m$^2$ and 95% of errors should not exceed 20 W/m$^2$ (Lefèvre and Wald, 2016). However, the validation results in a minority of stations are above the WMO standard. It is therefore probable that the SDDs can be used for modelling and mapping solar irradiance with some modification and using results of bias correction under cloudy-sky conditions.

## 5.5    Conclusions

The study in this chapter validated hourly GHI from two SDDs ( HC3v5 and CRSv3) against ground data from nine stations in northeast Iraq for all-sky, clear-sky and cloudy-sky conditions values were compared for the station pixel and with two other adjacent pixels around the station in all-sky conditions. The simple method was applied for bias corrections in SDDs under cloudy-sky conditions. The temporal changes of ground data GHI were well represented by both SDDs; $r$ was above 0.94 for the all-sky and clear-sky conditions, and above 0.82 for the clearness index in most cases, while for cloudy-skies was between 0.51–0.72. The bias was negative (underestimation) for most cases except for two HC3v5 and three CRSv3 cases,; all of the absolute bias ranges were smaller than 8% of the mean GHI in all-sky and clear-sky conditions. For cloudy-sky conditions, bias was positive and varied from one station to another, by 17%–85% of the mean GHI. The same applies to RMSE. It ranged between 8–20% in all of the stations for all-sky and clear-sky conditions. In contrast, the RMSE range was much higher in cloudy-sky conditions: above 56%. The differences between neighbouring pixels and at-station pixels in the SDDs compared to the ground data of GHI for each site are very small, varying by 2% in most cases. The overall performance of HC3v5 is better than that of CRSv3.

Despite the high error rates at some stations, the GHI values from SDDs were closely related to the ground data at most of the stations. However, the resolution of MSG images is 5 km. The SDDs represent hourly GHI well, and this can be used to map solar resources and possibly for modelling GHI with ground data in areas with a low number of stations.

The results of bias correction revealed clear improvement of SDDs under cloudy-sky conditions by reducing the bias by 10%–80% in various cases and reducing bias to near zero at some stations.

Further research would be useful for validating the SDDs in other climates. Some studies are also required to address the inputs to the Heliosat method, according to regional and local factors, for a better estimation of GHI from satellite images.

Chapter 6 :   **Modelling Hourly Global Horizontal Irradiance from Satellite-Derived Datasets and Climate Variables as New Inputs with Artificial Neural Networks**

## 6.1   Introduction

This chapter addresses objective:

5. To assess the role of new input combinations (SDDs. Cs, TOA and climate variables) in estimating and modelling hourly GHI data.

Several studies have estimated GHI from various methods, but a high temporal resolution of GHI is likely necessary for several applications such as photovoltaic panel and concentrated solar power projects. Recently, demand for GHI that has increased for solar energy projects, due to problems related to non-renewable energies, a lack of other energy sources, increasing use of energy and potential availability of solar energy (Boussaada et al., 2018; Jadidi et al., 2018; Palmer et al., 2017; Palmer et al., 2018). However, stations with long historical measurements of GHI are limited because of the cost of installation and maintenance of pyranometers (Hassan et al., 2017b).

Therefore, this chapter aimed to address a model for obtaining high quality of GHI data on an hourly time step according to bias and RMSE by using new input combinations including SDDs, *Cs* model, TOA and climate variables (SD, AT, RH and WS) in a comparison of the results of ANN and regression models. This is a gap filling knowledge because the new input combinations have not been used by previous literature (Chapter 2, Section 4) to improve the quality of the results of GHI modelling.

This chapter used quality controlled GHI ground data (Chapter 4) and validated SDDs (Chapter 5) as the new input combinations in ten sets of inputs in ANN and regression models.

## 6.2 Materials and Methods

### 6.2.1 Study Site

The study area is in northeast Iraq, and was described in Chapter 3, Section 3.1. The same nine stations used in Chapter 5 were used here.

### 6.2.2 Ground Measurements

Hourly ground data for SD AT, RH, WS and GHI, were collected from two station types. First, the data are from tower stations are all above variables except SD. The pyranometer used for recording GHI in these stations is the Kipp and Zonen CMP6 Pyranometer. The data were collected for the period 2011–2014 from five stations, some of which lacked some years, from the Ministry of Electricity-KRG (Table 6.1). The stations are automatic weather stations, at which SD is also recorded as well as the above variables. The GHI equipped in these stations is the Vaisala QMS101 Pyranometer. The data were collected from 2013–2016 by the General Directorate of Meteorology and Seismology-KRG for four stations (Table 6.2). Both datasets are missing some months or years.

**Table 6.1:** Tower stations with hourly data of AT, RH, WS and GHI at ground measurements with GHI of two SDDs and with calculated Cs and TOA irradiance.

| Station | Coordinates (Degrees) | | Elevation a.s.l (m) | Period (dd/mm/yy) |
|---------|-----------|-----------|---------------------|-------------------|
| Batufa | 37.1764 N | 43.0236 E | 947 | 01/01/2011–31/12/2013 |
| Enjaksor | 37.0603 N | 42.4353 E | 509 | 01/01/2011–31/12/2014 |
| Hojava | 37.0075 N | 43.0369 E | 933 | 01/01/2011–31/12/2013 |
| Jazhnikan | 36.3564 N | 43.9556 E | 430 | 01/01/2011–31/10/2013 |
| Tarjan | 36.1258 N | 43.7353 E | 276 | 01/01/2011–31/12/2013 |

**Table 6.2:** Automatic stations as data in Table 6.1, plus SD.

| Station | Coordinates (Degrees) | | Elevation a.s.l (m) | Period (dd/mm/yy) |
|---------|-----------|-----------|---------------------|-------------------|
| Halsho | 36.2097 N | 45.2598 E | 1105 | 01/01/2013–31/12/2016 |
| Bazian | 35.6021 N | 45.1376 E | 892 | 01/04/2014–30/12/2016 |
| Maydan | 34.9194 N | 45.6224 E | 330 | 01/01/2014–31/12/2016 |
| Kalar | 34.6244 N | 45.3049 E | 218 | 01/01/2014–31/12/2016 |

### 6.2.3  Calculated data

These were described in (Chapter 4, Section 2.1.2).

### 6.2.4  Satellite-Derived Datasets (SDDs)

These were described in (Chapter 5, Section 5.2.4).

### 6.2.5  Quality Control of GHI Measurements and Evolution of SDDs

The output data from the nine stations, where ground measurements were tested (Chapter 4) and SDDs were evaluated (Chapter 5), were used in this chapter. In addition, the station data was harmonised with SDDs (HC3v5 and CRSv3), with TOA and Cs. All data were merged into one dataset. The dataset was configured based on the true solar time, when the solar elevation angle is above 15°. Systematic errors, NA values and a few questionable data points were then removed.

### 6.2.6  Data Pre-processing

The data were normalised to 0–1 as it is recommended for ANNs (Quaiyum et al., 2011; Jimenez et al., 2016; Fadare, 2009; Saberian et al., 2014). The input normalised data of SD, AT, RH, WS, Cs, TOA, HC3v5 and CRSv3 were set as ten different inputs. Each input contained some of the above variables in both tower and automatic stations (Table 6.3). Each set of input combinations were named model-1 (M1) to model-10 (M10) (Table 6.3). Hereafter, each combination of inputs at each station from M1 to M10 was used in ANN and regression models, and the results were presented using those names (M1–M10), as demonstrated in Table 6.3. The data in each station were randomly distributed for each model from M1–M10 as training (70%), validation (15%) and test (15%) data for ANN models. The training data were presented to the network during training, and the network was adjusted according to its error. The validation data were used to measure network generalization, and to halt training when generalization stops improving based on an increase in the mean square error. The test data have no effect on training and test data provide an independent measure of network performance during and after training (MATLAB, 2018).

The same data deviations for each station from (M1–M10) were used in multiple regression models, but the data were randomly distributed as training (70%) and test (30%).

**Table 6.3:** Inputs and output to the ANN models.

| Models | Inputs (automatic stations) | Inputs (tower stations) | Output |
|--------|------------------------------|--------------------------|--------|
| M1 | SD, AT, RH, WS | AT, RH, WS | GHI |
| M2 | SD, AT, RH, WS, Cs | AT, RH, WS, Cs | GHI |
| M3 | SD, AT, RH, WS, TOA | AT, RH, WS, TOA | GHI |
| M4 | SD, AT, RH, WS, HC3v5 | AT, RH, WS, HC3v5 | GHI |
| M5 | SD, AT, RH, WS, CRSv3 | AT, RH, WS CRSv3 | GHI |
| M6 | SD, AT, RH, WS, Cs, TOA | AT, RH, WS, Cs, TOA | GHI |
| M7 | SD, AT, RH, WS, Cs, HC3v5 | AT, RH, WS, Cs, HC3v5 | GHI |
| M8 | SD, AT, RH, WS, Cs, CRSv3 | AT, RH, WS, Cs CRSv3 | GHI |
| M9 | SD, AT, RH, WS, Cs, TOA, HC3v5 | AT, RH, WS, Cs, TOA, HC3v5 | GHI |
| M10 | SD, AT, RH, WS, Cs, TOA, CRSv3 | SD, AT, RH, WS, Cs, TOA, CRSv3 | GHI |

### 6.2.7 Artificial Neural Networks (ANNs)

ANNs are machine-learning techniques, which are based on how the human brain works and they are prepared to replicate the way that humans learn. ANNs are considered one of the most powerful algorithms for finding a relationship between inputs and outputs. They have been used broadly in the literature for modelling GHI and have been described in detail (Chapter 2, Section 2). ANNs are black box models because they do not provide any insights into the structure of the function being shaped, while they can approximate any function.

ANNs contain three primary layers, which are the input layer, hidden layer and output layer (Figure 6.1). Each layer consists of a number of neurons and every neuron in one layer is associated and interacts with other layers. Each neuron in the input layer only acts as a partition to distribute the input value to the neuron in the next layer. Each input to the neuron in that next layer is weighted by the strengths of the respective connections from the input layer and summed with the adjustable value (Bias) and then the output of the neuron is calculated with a function. The weight and bias in each layer's neurons are adjusted based on the activation function or transfer function (sigmoid, hyperbolic tangent

and linear), and algorithms (including Levenberg-Marquardt, Bayesian Regularization and Scaled Conjugate Gradient) for training the model, which depend on error minimising between the desired output and the target (Kheradmanda et al., 2016; López et al., 2018; Loutfi et al., 2017).
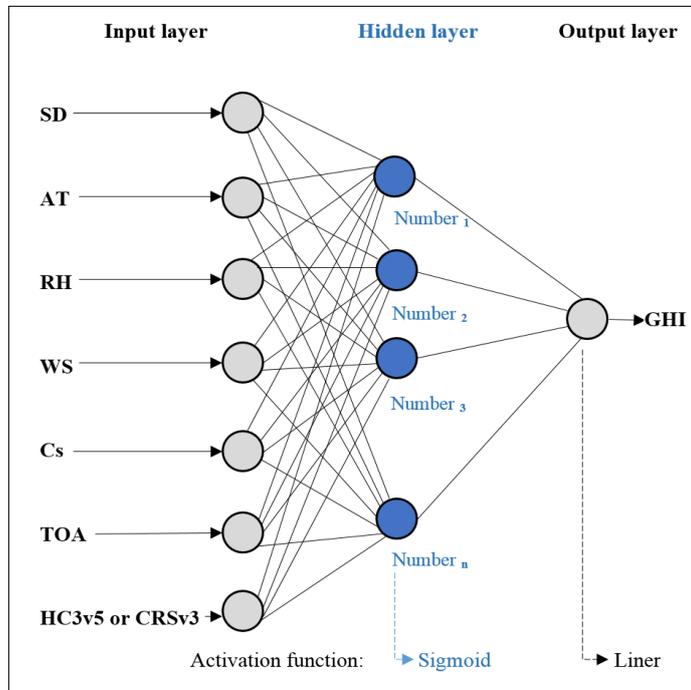


**Figure 6.1:** ANN architecture description.

The neural network fitting toolbox (nftool) of MATLAB R2016a academic use (MATLAB, 2018) was used in this chapter (see Appendix C, Figures C3–C10 and Table C10). After the data were normalised to 0–1 and inputs and outputs were designated as M1–M10 for each station, the data were divided into training, validation and test sets. The Levenberg-Marquardt back propagation algorithm (Moré, 1978) was used to train each model with activation functions, sigmoid in the hidden layer and linear in the output layer (Figure 6.1). The number of neurons in the hidden layer for each model was selected after several tests based on the performance and balance of under or overfitting of results among the training, validation and test datasets. All other processes such as initial weight and bias, and connections between layers were automatically completed.

The Levenberg-Marquardt algorithm was used due to its reduced time requirement for convergence (MATLAB, 2018), and its results are better than others were including Bayesian Regularization and Scaled Conjugate Gradient in the case of modelling GHI (Fadare, 2009; Khosravi et al., 2018a; López et al., 2018; Rao K et al., 2018). Another

reason is that the initial results of Levenberg-Marquardt better than those of the other two training algorithms.

The methodology steps in this chapter are illustrated in Figure 6.2. The results of each station and each model were aggregated for five towers and four automatic stations in one combination for each station type as overall results of the tower station and overall results of the automatic station; this is to avoid a complicated explanation of the results of each station separately. Hence, the results of each station are displayed in Appendix C.
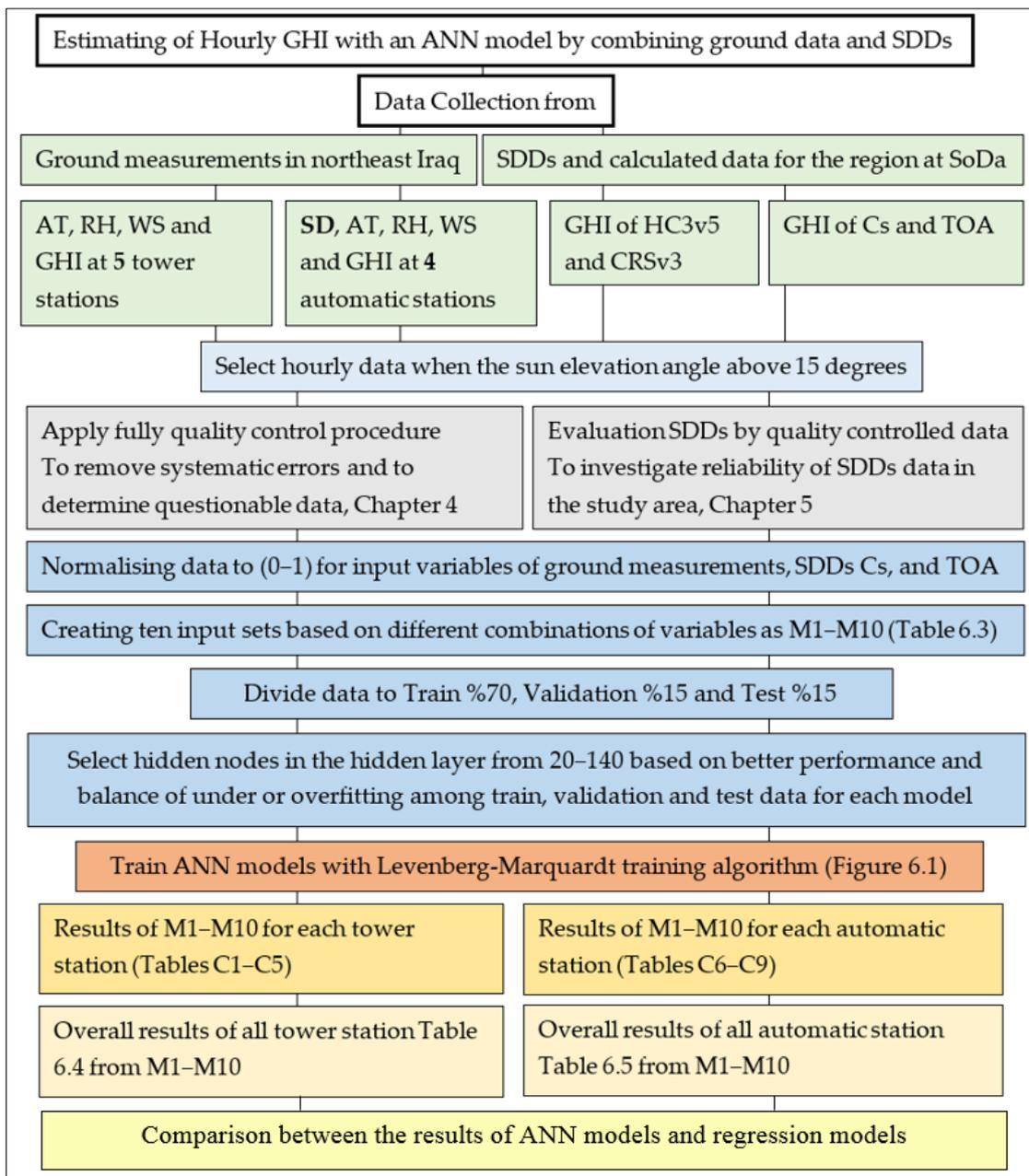


**Figure 6.2:** Flowchart of the methodology steps in this chapter.

## 6.3    Results

The results of hourly GHI with ANN and regression models from M1–M10 based on variable inputs for train, validation and test data were averaged for four automatic stations and five tower stations and are presented in Tables 6.4 and 6.5 respectively. The same results with the number of neurons in the hidden layer, number of datasets used and mean of GHI ground data for each induvial station types are presented in Appendix C, Tables (C1–C9) with two Figures (C1 and C2) of relative bias and RMSE of test data for further demonstration. The results of regression models are presented in Tables 6.6 and 6.7 for both station types. Figures 6.3 and 6.4 show further comparison of the relative bias and RMSE between the ANN and regression models and the station types in the test data. In addition, the results of M1–M10 in the test data are shown with scatter plots of ground vs ANN models and estimated vs residuals in Figures 6.5–6.8 for both station types respectively.

As can be seen in Tables 6.4 (tower stations) and 6.5 (automatic stations), there is no significant difference (the differences are lower than 3% in all individual cases) when comparing training and validation data with test data, which is in line with the stated methodology. Therefore, the results will be presented and discussed according to the models' test data, which is more important to demonstrate the reliability of each model.

The overall results of GHI estimation by ANN models compared to ground data shows the better performance for automatic stations than for tower stations in all models based on $r$ values, bias and RMSE (Tables 6.4 and 6.5 , Figures 6.3–6.8).

The lowest $r$ value ranges among the models between 0.601 and 0.755 in M1 for both station types respectively. The highest $r$ value is 0.983 for M9 applied to automatic stations and 0.976 in M10 tower stations. Other $r$ values range from 0.903–0.982 in both station types (Tables 6.4 and 6.5). Despite both high and low $r$ values, the values of M3 and M5 compared to other remaining values are low in automatic stations. This is also true for M2, M3 and M6 to others at tower stations.

The values of bias were significantly low in all cases in the study area, which is under 1% of mean ground data for M1–M10. In the tower stations, the highest bias was recorded in M1 (3.4 W/m$^2$) 0.7%. It was 0.4% (2.3 W/m$^2$) in M2, a negative bias of −0.4% (−2 W/m$^2$) in M7 and the other rates were below 0.3%. However, in the automatic stations,

the highest bias was recorded in M3 ($-2$ W/m$^2$) $-0.4\%$. It was 0.3% in M1 and M5 and the others were below that value. The lowest bias was recorded at M8, M9 and M10, which were close to zero in both station types (Tables 6.4 and 6.5, Figure 6.3). Figure 6.3 demonstrates the low rates of relative bias among M1–M10 for both station types.

The RMSE results were similar to those for bias. The highest RMSE for the tower stations was recorded in M1 (209.5 W/m$^2$ 41%). It decreased to 111.8 W/m$^2$ (21.5%) and to 104.4 W/m$^2$ (20.2%) in M3 and M2 respectively. The lowest recorded RMSE values were 57.8 W/m$^2$ (11.2%), 60 W/m$^2$ (11.6%) and 60.4 W/m$^2$ (11.6%) in M10, M9 and M7 respectively. Other rates are between 12%–19%.

On the other hand, the RMSE at automatic stations were low compared to the tower stations for each model. The highest was recorded in M1 (163.6 W/m$^2$ 33.7%). It decreased to 60 W/m$^2$ (12.4%) in M3. The lowest RMSE was recorded in M9 (46.3 W/m$^2$ 9.5%) and there were slightly higher values in M7 (47.6 W/m$^2$ 9.8%) and M10 (47.2 W/m$^2$ 9.8%). The other remaining values were between 10%–12% (Tables 6.4 and 6.5, Figure 6.4). Figure 6.4 shows kind of stability of relative RMSE in automatic stations after M1 among other models whereas it shows fluctuations for tower stations for the same situation.

The scatter plots of each individual model in Figures 6.5 (tower stations) and 6.7 (automatic stations) show the results of hourly GHI models for test data against ground measurements. The observations are concentrated around the 1:1 line in better performance models (M8, M9 and M10), where the regression lines are correspondingly close to the 1:1 line. The opposite is seen in M1 for both station types. However, in models M2, M3 and M6 (tower stations) the observations are far from the 1:1 line and the regression line in red is not close to the 1:1 line, corresponding to high recorded RMSE compared to other models (Figure 6.5).

Figures 6.6 (for tower stations) and 6.8 (for automatic stations) show the scatter plots of residuals against estimated hourly GHI of test data in each model. The clustered patterns of residuals are seen only in M1 in both station types whereas all other residuals are randomly distributed and the densities of observation are around zero. However, low performance can be noted at M2, M3 and M6 (Figure 6.6).

Regarding the results of regression models compared to ANN models, there is clearly a better performance using ANN than regression models according to *r* values and RMSE in each individual cases. However, according to bias in few cases regression is better than ANN models by minimal rates no more than 0.4% (Tables 6.4, 6.5., 6.6 and 6.7, Figures 6.3 and 6.4).

The results of regression are similar to ANN models by increasingly better performance after M1, M2 and M3 to other models (M4–M10). The lowest bias was recorded in M10 (0.01%), M6 (0.02%), M5 (0.07%) and M8 (0.07%) at tower stations and it was −0.06% in all M4, M5 and M7 at automatic stations. The lowest RMSE was 12% in both M7 and M9 at automatic stations and it was 13% in all M4, M7 and M9 at tower stations (Tables 6.6 and 6.7)
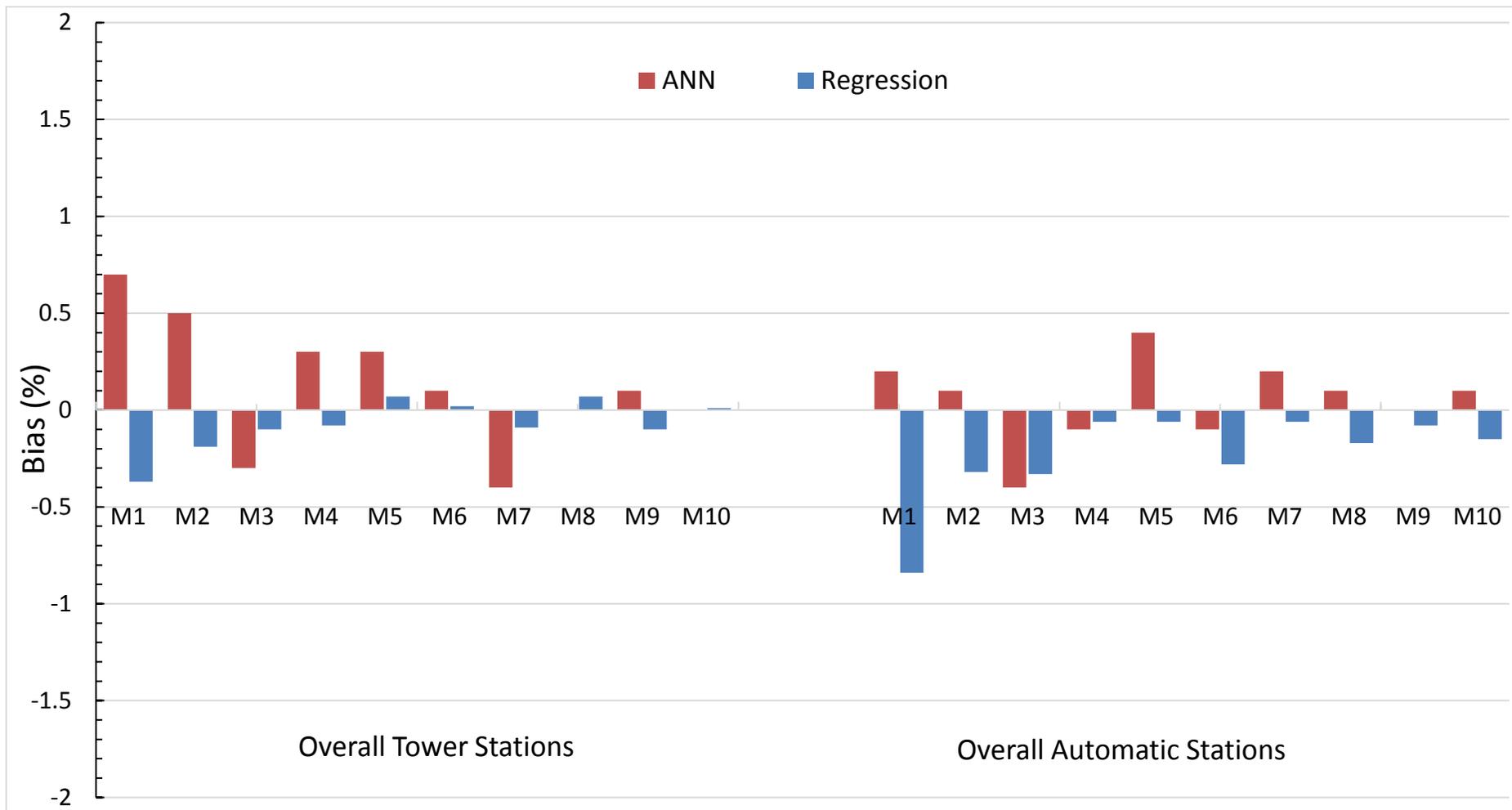
**Figure 6.3:** Comparison of bias (%) for the hourly GHI among ANN and regression models and for the overall results of station types.
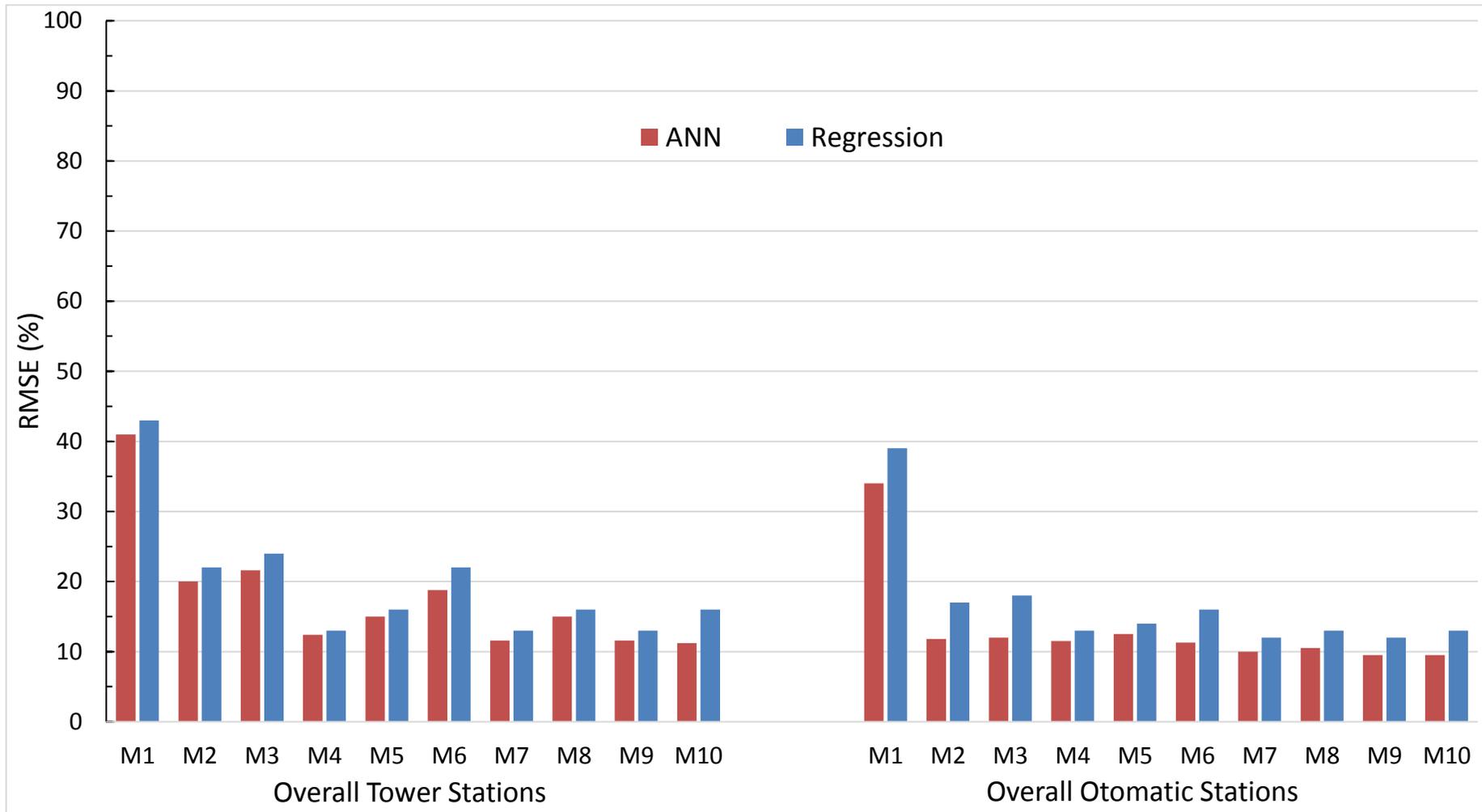
**Figure 6.4:** As in Figure 6.3, but for RMSE (%).

**Table 6.4:** Statistical results of hourly GHI models averaged for each ANN model M1–M10 for **tower stations**. Mean, Bias and RMSE units are W/m².

| Models | Train | | | | | | Validation | | | | | | Test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Bias | Bias % | RMSE | RMSE % | r | Mean | Bias | Bias % | RMSE | RMSE % | r | Mean | Bias | Bias % | RMSE | RMSE % | r |
| M1 | 517.2 | −0.4 | −0.1 | 206.6 | 40 | 0.617 | 516.2 | 0.2 | 0.0 | 208.4 | 40.2 | 0.600 | 511 | 3.4 | 0.7 | 209.5 | 41 | 0.601 |
| M2 | 516.2 | 3.7 | 0.7 | 103.4 | 20 | 0.919 | 513.6 | 4.0 | 0.8 | 105.4 | 20.4 | 0.916 | 516.7 | 2.3 | 0.4 | 104.4 | 20.2 | 0.917 |
| M3 | 514.8 | 1.7 | 0.3 | 111.6 | 21.6 | 0.905 | 518.8 | −0.4 | −0.1 | 112.8 | 21.6 | 0.903 | 519.7 | −1.7 | −0.3 | 111.8 | 21.5 | 0.903 |
| M4 | 515.8 | 1.8 | 0.3 | 62 | 12.2 | 0.971 | 518.4 | 1.0 | 0.2 | 62.6 | 12 | 0.971 | 516.7 | 1.5 | 0.3 | 62.9 | 12.2 | 0.971 |
| M5 | 515.6 | 0.2 | 0.0 | 76 | 15 | 0.957 | 516.4 | 1.0 | 0.2 | 76.6 | 14.8 | 0.956 | 517.2 | 1.7 | 0.3 | 75.6 | 14.6 | 0.958 |
| M6 | 516 | −0.6 | −0.1 | 95 | 18.4 | 0.932 | 514 | −0.8 | −0.2 | 98.4 | 19.4 | 0.927 | 518.7 | 0.9 | 0.2 | 98.1 | 18.9 | 0.929 |
| M7 | 515.6 | −1.5 | −0.3 | 59.8 | 11.4 | 0.974 | 513 | −1.1 | −0.2 | 60.4 | 11.6 | 0.973 | 521.1 | −2.0 | −0.4 | 60.4 | 11.6 | 0.973 |
| M8 | 515.8 | −0.7 | −0.1 | 73 | 14.2 | 0.961 | 519.2 | −1.4 | −0.3 | 74.2 | 14.2 | 0.960 | 514 | −0.1 | 0.0 | 75.3 | 14.6 | 0.958 |
| M9 | 515.6 | 0.4 | 0.1 | 59.2 | 11.4 | 0.974 | 518.6 | 0.4 | 0.1 | 59.6 | 11.4 | 0.974 | 516.3 | 0.3 | 0.1 | 60 | 11.6 | 0.974 |
| M10 | 517.2 | 1.3 | 0.2 | 55.8 | 10.8 | 0.977 | 512.6 | 1.8 | 0.4 | 58.2 | 11.4 | 0.975 | 513.8 | −0.2 | 0.0 | 57.8 | 11.2 | 0.976 |

**Table 6.5:** As in Table 6.4, but for **automatic stations**.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Bias | Bias % | RMSE | RMSE % | r | Mean | Bias | Bias % | RMSE | RMSE % | r | Mean | Bias | Bias % | RMSE | RMSE % | r |
| M1 | 485 | −0.2 | −0.1 | 160.5 | 33.0 | 0.743 | 481.2 | 1.2 | 0.2 | 162.0 | 33.8 | 0.736 | 482.2 | 1.4 | 0.3 | 163.6 | 33.7 | 0.755 |
| M2 | 484.2 | 0.4 | 0.1 | 54.3 | 11.5 | 0.974 | 486.5 | −0.2 | 0.0 | 55.8 | 11.5 | 0.973 | 479.9 | 0.7 | 0.1 | 56.8 | 11.8 | 0.974 |
| M3 | 484.5 | −0.6 | −0.1 | 55.5 | 11.8 | 0.973 | 479.2 | −1.2 | −0.3 | 59.0 | 12.3 | 0.970 | 485.4 | −2.0 | −0.4 | 58.8 | 12.1 | 0.971 |
| M4 | 484 | 0.0 | 0.0 | 53.8 | 11.3 | 0.975 | 488 | 0.8 | 0.2 | 54.5 | 11.3 | 0.974 | 480.4 | −0.5 | −0.1 | 54.6 | 11.4 | 0.976 |
| M5 | 484.7 | 1.5 | 0.3 | 58.8 | 12.5 | 0.970 | 480.5 | 2.0 | 0.4 | 61.8 | 13.0 | 0.967 | 484.4 | 1.6 | 0.3 | 60.0 | 12.4 | 0.970 |
| M6 | 482.5 | −0.7 | −0.1 | 52.8 | 10.8 | 0.976 | 491.2 | −1.3 | −0.3 | 54.0 | 11.0 | 0.974 | 482.6 | −0.5 | −0.1 | 54.1 | 11.2 | 0.976 |
| M7 | 483.5 | 1.2 | 0.3 | 46.0 | 9.3 | 0.982 | 486.5 | 0.4 | 0.1 | 47.3 | 9.5 | 0.981 | 483.3 | 1.1 | 0.2 | 47.6 | 9.8 | 0.981 |
| M8 | 484 | 0.1 | 0.0 | 48.5 | 10.3 | 0.979 | 481.7 | −0.1 | 0.0 | 50.0 | 10.3 | 0.978 | 486 | 0.3 | 0.1 | 49.5 | 10.2 | 0.980 |
| M9 | 484 | 0.7 | 0.1 | 43.5 | 9.0 | 0.984 | 480 | 0.0 | 0.0 | 44.3 | 9.3 | 0.983 | 487.9 | −0.2 | 0.0 | 46.3 | 9.5 | 0.983 |
| M10 | 484.5 | 0.3 | 0.1 | 44.5 | 9.5 | 0.983 | 483.2 | 0.6 | 0.1 | 45.3 | 9.3 | 0.982 | 482.1 | 0.3 | 0.1 | 47.2 | 9.8 | 0.982 |
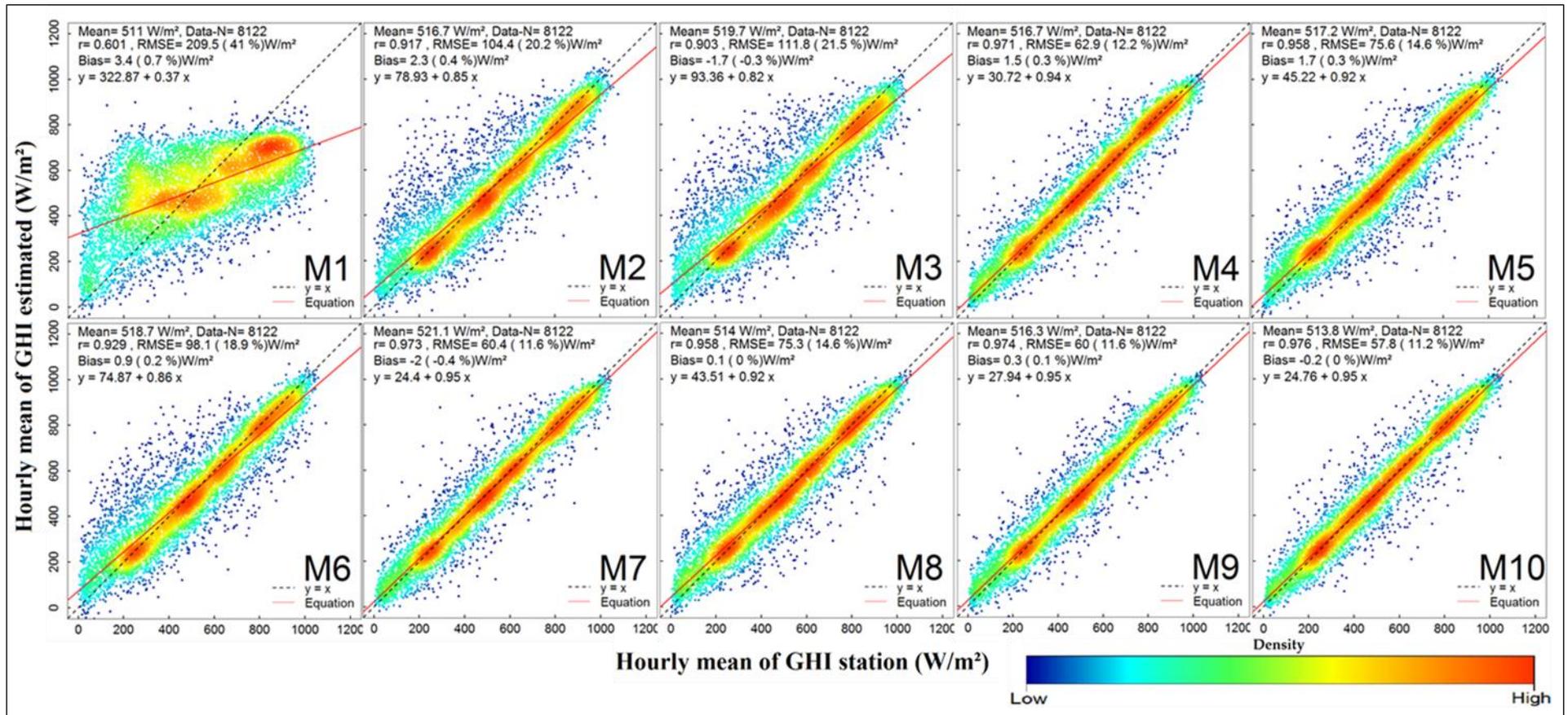
**Figure 6.5:** Scatter plots of hourly GHI ground measurements and ANN model estimated from M1–M10 at test data for overall results at tower stations.
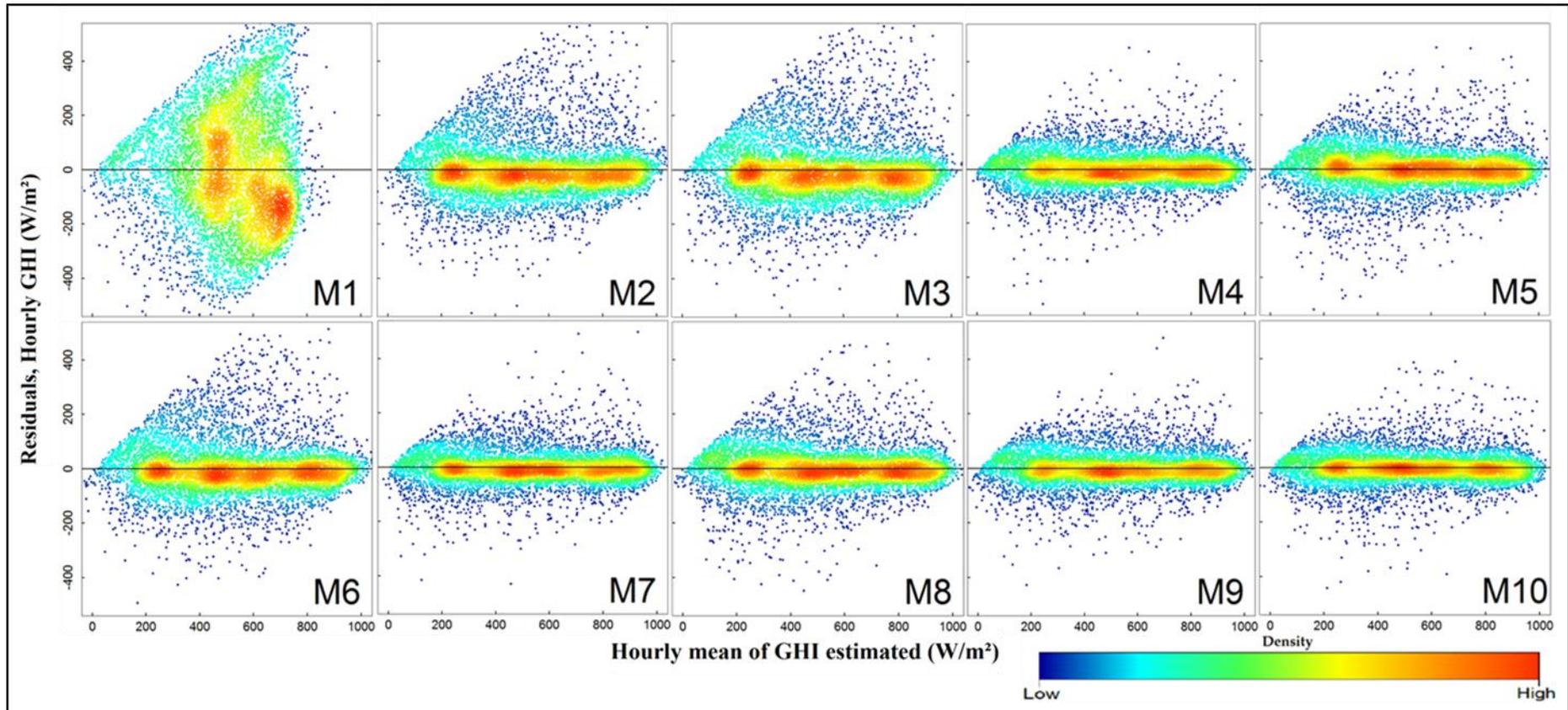
**Figure 6.6:** Scatter plots of hourly GHI residuals versus ANN model estimated from M1–M10 at test data for overall results at tower stations. The plots clearly reveal that the models were a good fit at M4, M5 and M7–M10 whereas it is not fit at M1 and low fit are seen at M2, M3 and M6.
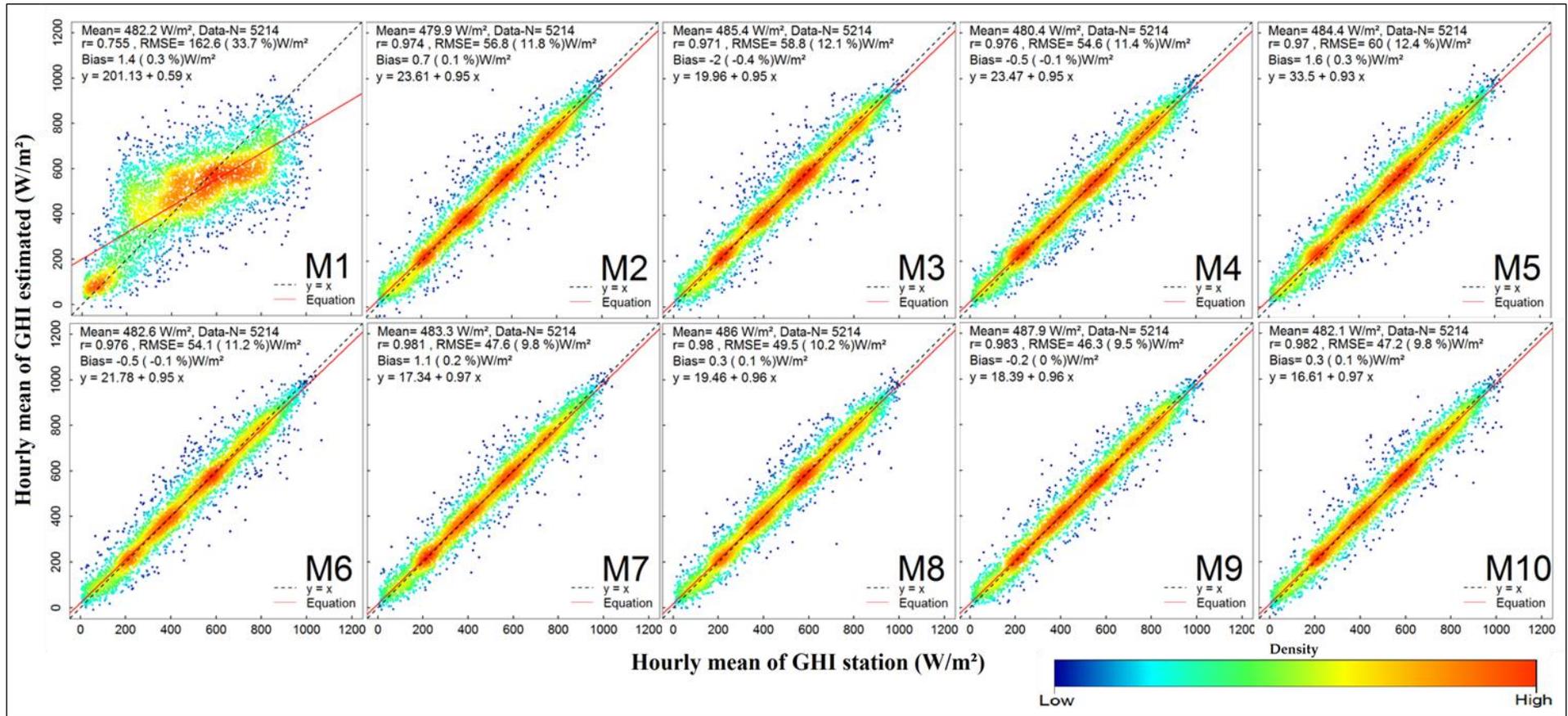
**Figure 6.7:** Scatter plots of hourly GHI ground measurements and ANN model estimated from M1–M10 at test data for overall results at automatic stations.
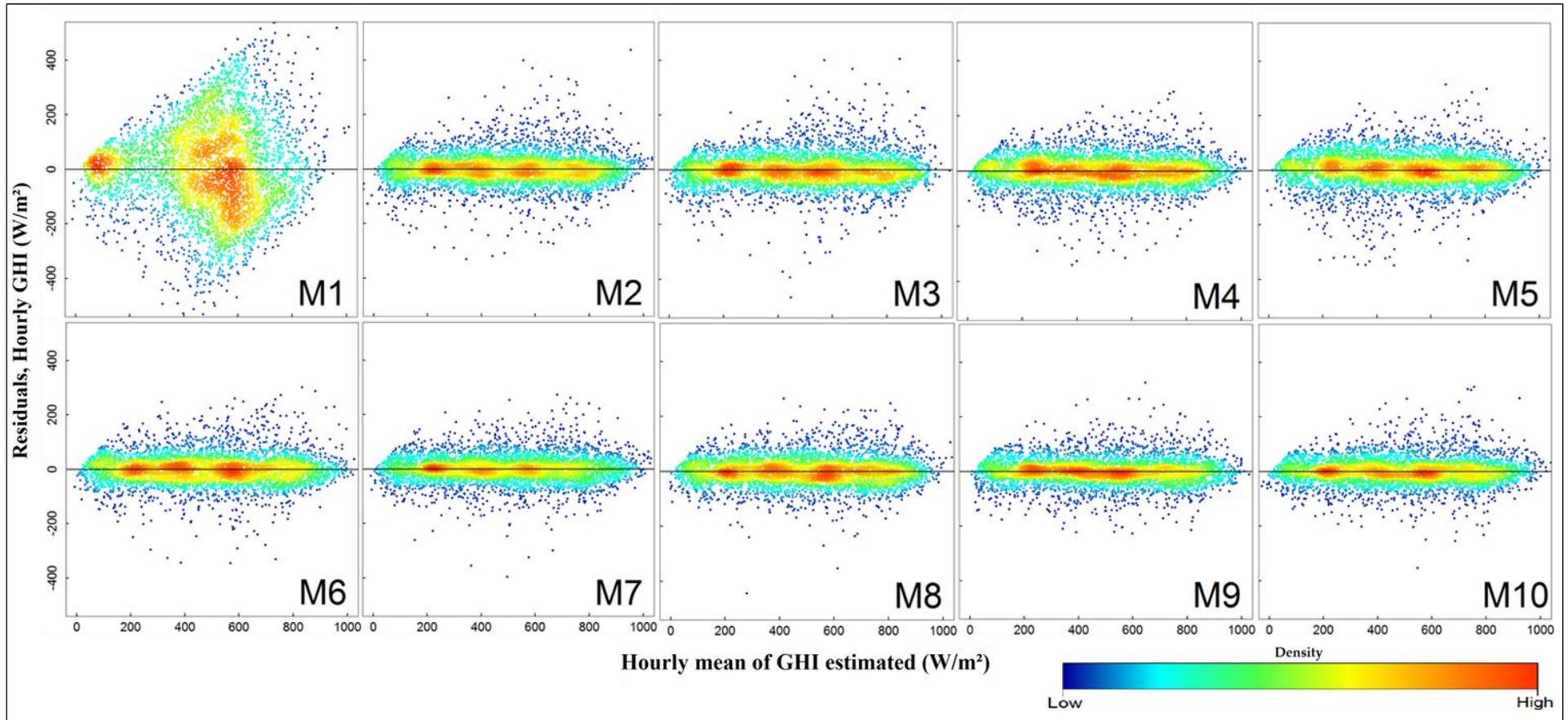
**Figure 6.8:** Scatter plots of hourly GHI residuals versus ANN model estimated from M1–M10 at test data for overall results at tower stations. The plots clearly reveal that the models were a good fit for all except M1.

**Table 6.6:** Statistical results of hourly GHI models averaged for each regression model M1–M10 for **tower stations.** Mean, Bias and RMSE units are W/m$^2$.

| Models | Test data | | | | | |
|---|---|---|---|---|---|---|
| | **Mean** | **Bias** | **Bias %** | **RMSE** | **RMSE %** | *r* |
| M1 | 516 | -1.9 | -0.37 | 220 | 43 | 0.55 |
| M2 | 516 | -1 | -0.19 | 115 | 22 | 0.9 |
| M3 | 516 | -0.5 | -0.1 | 125 | 24 | 0.88 |
| M4 | 516 | -0.42 | -0.08 | 67 | 13 | 0.967 |
| M5 | 516 | 0.36 | 0.07 | 81 | 16 | 0.952 |
| M6 | 516 | 0.12 | 0.02 | 111 | 22 | 0.907 |
| M7 | 516 | -0.45 | -0.09 | 67 | 13 | 0.967 |
| M8 | 516 | 0.35 | 0.07 | 80 | 16 | 0.953 |
| M9 | 516 | -0.53 | -0.1 | 67 | 13 | 0.967 |
| M10 | 516 | 0.05 | 0.01 | 80 | 16 | 0.953 |

**Table 6.7:** As in Table 6.6, but for **automatic stations.**

| Models | Test data | | | | | |
|---|---|---|---|---|---|---|
| | **Mean** | **Bias** | **Bias %** | **RMSE** | **RMSE %** | *r* |
| M1 | 490 | -4.14 | -0.84 | 189 | 39 | 0.644 |
| M2 | 490 | -1.56 | -0.32 | 81 | 17 | 0.945 |
| M3 | 490 | -1.61 | -0.33 | 87 | 18 | 0.936 |
| M4 | 490 | -0.31 | -0.06 | 62 | 13 | 0.968 |
| M5 | 490 | -0.29 | -0.06 | 71 | 14 | 0.958 |
| M6 | 490 | -1.38 | -0.28 | 80 | 16 | 0.946 |
| M7 | 490 | -0.3 | -0.06 | 59 | 12 | 0.971 |
| M8 | 490 | -0.84 | -0.17 | 66 | 13 | 0.964 |
| M9 | 490 | -0.41 | -0.08 | 58 | 12 | 0.972 |
| M10 | 490 | -0.72 | -0.15 | 65 | 13 | 0.965 |

## 6.4    Discussion

Hourly GHI was estimated for nine stations in northeast Iraq using observed meteorological inputs (SD, AT, RH and WS), calculated inputs (TOA and Cs) and new input from SDDs (HC3v5 or CRSv3) to the ten M1–M10 ANN models based on the number and combination of inputs. The results of the overall performance have $r$ values from 0.601 to 0.976, bias from −0.4% to 0.7% and RMSE from 11.2% to 41% at tower stations and $r$ values from 0.755 to 0.983, bias from −0.4% to 0.3% and RMSE from 9.5% to 33.7% at automatic stations. Excellent performance was recorded in M9 (9.5%) and M10 (11.2%) and low performance was recorded in M1 at automatic and tower stations respectively. The better results of those models at hourly time scales compared to the previous studies for similar estimation (Chapter 2, Table 2.4) are related to the new inputs such as Cs, TOA and SDDs together in this study and in this chapter.  The results of ANN models recorded better performance than regression models in most cases. This demonstrates the capability of the ANN algorithm for capturing the relationships between inputs and outputs.

The overall better performance for automatic stations than for tower stations in all models, was obtained by the using of SD as an input in automatic stations—SD is unrecorded by tower stations. It is also reported by literature (Ampratwum and Dorvlo, 1999; Dahmani et al., 2016; Loutfi et al., 2017; Rao K et al., 2018) that the role of SD increases the performance of models. This is unsurprising; it is recording the same phenomenon.

The low performance of M1 for both station types is related to the small number of inputs, which do not include any of the calculated inputs. The calculated inputs such as *Cs* have a unique role to increase the model performance as seen in M2 compared to M1 in both stations. This is in agreement to improved results in some limited studies, which used Cs as inputs either for modelling or forecasting GHI (Lima et al., 2016; Aguiar et al., 2015; Ramedani et al., 2014; Voyant et al., 2011).

The low recorded bias in most of the models is related to the good estimation of GHI by ANN models as mentioned in several studies (Dahmani et al., 2016; Hassan et al., 2017a; Jimenez et al., 2016; Loutfi et al., 2017). The overall bias among stations are presented, which led to a decrease in the bias because of positive bias in some stations

and negative bias in others in the same model; whereas the bias in all individual stations was, lower than 2% except one case of 2.2% (Appendix C, Tables C1–C9, Figure C1).

The fluctuation of RMSE among models at tower stations and its stability among models at automatic stations (Figures 6.4 and C2, Tables 6.4 and 6.5) is mainly related to the role of SD, which was used as an input in the latter. The highest record of RMSE in M1 in both station types are related to inputs which contain only four climate variables. This is reported by literature where GHI was estimated at a daily time scale (Ozgoren et al., 2012; Rao K et al., 2018). The improved performance in M2 and M3 compared to M1 is related to the use of additional variables of Cs and TOA in those models respectively (Table 6.3). Hence, the low performance of M2, M3 and M6 compared to better performance in M4 and M5 are related to the use of SDDs as new inputs with climate variables. This has been reported by studies, which have used SDDs in foresting GHI (Aguiar et al., 2016; Aguiar et al., 2015). The better performance of M7 and M8 compared to the previous M1–M6 are related to the use of Cs with SDDs in those models. The role of Cs is mentioned in the literature (Lima et al., 2016; Voyant et al., 2011), but in those cases, it was not combined with SDDs. The overall better performances of M9 and M10 from the other models (M1–M8) are related to the fact all variables were used in those models. These demonstrate the better performance of this study compared to similar studies (Dahmani et al., 2016; Hassan et al., 2017b; Ibrahim and Khatib, 2017; Loutfi et al., 2017). This is because the best RMSE recorded by this study is 9.5%, whereas the other studies are 13.3%, 17%, 18.9% and 13.1 respectively.

The performance of HC3v5 as an input with only four climate variables is better than CRSv3 as demonstrated in the comparison between M4 and M5 in both station types, whereas in other models (M7–M10) the difference between them are minimal. The former results (M4–M5) are related to the accurate reproduction of the GHI ground data by HC3v5 as described in the literature (Ameen et al., 2018b; Marchand et al., 2017; Thomas et al., 2016a). The latter (M7–M10) is related to the use of Cs and TOA as separate inputs.

This study revealed that using SDDs, Cs, and TOA with climate variables in ANN models has improved the results of estimation for hourly GHI with all data combined $r$ value of 0.980, bias lower than 2% and RMSE lower than 10% compared to similar studies with no combination of those inputs (Dahmani et al., 2016; Hassan et al., 2017b; Ibrahim and Khatib, 2017; Loutfi et al., 2017).

The results of this study demonstrate that this way of modelling allows the retrieval or management of a dataset of GHI for decades where the inputs are available, but where GHI is not recorded as in most areas in the case study and similar regions with a scarcity of ground data, it can be achieved by using the trained models.

The overall better GHI estimation results of ANN models than regression models (Figures 6.3 and 6.4), demonstrated by this study, have also been reported in literature (Jiang, 2009; Jimenez et al., 2016; Kaba et al., 2018; Moreno et al., 2011; Sharifi et al., 2016; Zou et al., 2016; Kumar et al., 2015). In addition, recently ANN and other machine learning approaches have been used broadly for GHI modelling (Chapter 2, Section 2.4).

The improvement in the results of regression models for GHI estimation from M1, M2 and M3 to M4–M10 like ANN models but with low rates at both station types (Figures 6.3 and 6.4) are related to the role of new inputs (including Cs and SDD). The best results of regression models in this study, which recorded bias of lower than 1% and RMSE of 12%–13% indicates that the regression models can also be used with the new input combinations for modelling GHI due to their ease of use.

The new inputs of SDDs and Cs, improved the results and are easily and openly available for most regions (SoDa, 2018) unlike other variables such as cloud cover and SD (Ameen et al., 2018a; Hassan et al., 2016; Urraca et al., 2017c). Therefore, the mentioned new variables can be used for modelling and forecasting the solar components for better results.

The limitations are principally as follows: This chapter estimated GHI but no other solar components, which are required directly in fields such as DNI in concentrated solar power. Hence, some studies have estimated DNI and DHI from GHI (Loutfi et al., 2017; Pérez-Burgos et al., 2018; Lee et al., 2017). However, further research is required for that in such areas with a scarcity of ground data. Another limitation is the scarcity of long-term GHI ground data at timescales beyond five years or more, which are better for training this kind of model.

## 6.5 Conclusions

This chapter aimed to use a new input of SDDs together with Cs, TOA and observed climate variables SD, AT, RH and WS as new input combinations in ten ANN models to

estimate GHI at the hourly time scale with a Levenberg-Marquardt training algorithm. The inputs were arranged into ten different sets, as models M1–M10 to demonstrate the role of new inputs. The data at four automatic stations of all the above variables and five tower stations without SD in northeast Iraq were used. The ANN results were compared to regression models for the same inputs.

The test results demonstrated a good improvement from M1 to M10 based on adding the new inputs such as TOA with observed variables (M3), Cs with observed variables (M2), SDDs with all observed climate variables (M4–M5), other combinations (M6–M8) and all together (M9–M10) with low percent fluctuation between both station types. The best results are $r = 0.983$, RMSE = 9.5% and bias = 0.0% in M9 and $r = 0.976$, RMSE = 11.2% and bias = 0.0% in M10 and the worst results are $r = 0.755$, RMSE = 33.7% and bias = 0.3% in M1 and $r = 0.601$, RMSE = 41% and bias = 0.7% in M1 at automatic and tower stations respectively. The ANN results were better than the regression models in most cases. The regression models were also improved by new input combinations.

This study in this chapter demonstrated the role of new input combinations for estimating hourly GHI with high accuracy. While the models have been trained with a few years of data, it would be better to train them with more years of data with such algorithms.

Further research is required for using new inputs with other machine learning approaches and other empirical models.

Chapter 7 : **General Discussion, Conclusions and Research Contributions**

## 7.1 Introduction

This chapter is organised as follows. The knowledge contribution to the literature from this study is identified in Section (7.2). Meeting the aim and the objectives of the study and the main conclusions are highlighted in Sections 7.3 and 7.4. Finally, recommendations for future research based on the results and the limitations of the study are provided in Section 7.5.

## 7.2 Contribution to knowledge

This study presented new results of a novel QC approach for testing GHI; it demonstrated the difference between two QC tests and developing the local QC test; it validated two SDD values for GHI in a new area (northeast Iraq) with applying a method for bias correction of those SDDs under cloud-sky conditions, and it revealed the role of new input combinations to improve modelling hourly GHI. The key contribution of the thesis for each chapter can be listed as follows:

In Chapter 4, the study applied a new QC method to detect systematic errors and questionable data on GHI hourly time scales by utilising SD and AT climate variables. This is to cover the limitation of applied QC approaches where some of their subtests, such as consistency or comparison tests, cannot be applied when only GHI is available and other solar components are unavailable. The study showed that the new QC approach could be used as a consistency test partially. Moreover, the new QC procedure has detected a rate of errors in hourly GHI data, which has not been identified by other approaches.

Another contribution of this chapter is a comparison between two existing QC tests (BSRN and TOACs) that has not previously been documented in the literature. The study illustrated the difference between them and their subtests for flagging erroneous GHI data or revealing questionable data. Based on the differences of data flagging and error detection between those two QC tests, a new QC test as a local test was developed to check GHI. The ratio of data flagging by the local test is quite reasonable. This is because 1) the local test detected systaltic errors in which had not been detected by BSRN test, and 2) the local test was not flagged high rate of observation as errors as they were flagged by TOACs test.

In Chapter 5, the study contributed to previous studies by validating two SDDs for GHI in a new region (northeast Iraq), where those and other SDDs have not been validated in that area in the past. The study investigated the reliability of each SDD to reproduce GHI hourly ground data in all-sky, clear-sky and cloud-sky conditions. The findings showed that SDDs can be a good source of GHI data in the study area. However, they did not reproduce ground data well under cloudy-sky conditions. The study modified a method for bias correction and applied it to correct SDDs under cloudy-sky conditions; the corrected SDDs data can be used for forming a GHI dataset. The chapter also investigated a new validation method by validating two other point pixel data of SDDs around a station with station data. The finding showed minor differences in the two-point pixel data with station pixel data. This offers further insight for using GHI data of the neighbouring area and interpolating ground data for the same target in the case of limited recording observations in an area.

The contribution of Chapter 6 demonstrates the role of several sets of new input combinations of SDDs, TOA, Cs and climate variables in the ANN models to obtain high accuracy of hourly GHI data modelling. A significant finding was that HC3v5 and CRSv3 decreased RMSE in ANN models compared to other input sets. The results also demonstrated that all inputs together have a high accuracy of results than when some of the inputs are lacking. This study also proved the better performance of ANNs than regression models as seen by previous studies (Jiang, 2009; Jimenez et al., 2016; Kaba et al., 2018; Moreno et al., 2011; Sharifi et al., 2016; Zou et al., 2016; Kumar et al., 2015).

## 7.3 Meeting the Aim and Objectives

Solar irradiance data is considered as an essential parameter in the fields of solar energy, agriculture, architecture, hydrology and others. Therefore, it is highly required, and it is obtained from direct ground measurements, SDD, and modelling. Unfortunately, the recorded ground data are limited, especially in developed countries (e.g. Middle East) and the ground measurements need to be checked with QC approaches due to operational and equipment errors. An issue with QC methods is their error rates detection are different, and some of the QC methods cannot be applied when only GHI is available.

SDD is an option for obtaining solar irradiance data, whereas these data need to be evaluated with high-quality ground measurements. Obtaining GHI from a variety of

methods has been addressed in the literature, but achieving highly accurate results from hourly GHI data requires a new approach.

Given this background, this study aimed to develop a model for obtaining high accuracy of GHI hourly data in three steps from QC approaches, evaluating and correcting SDDs, and combining SDDs with ground measurements for that goal. The method could be used to create a GHI hourly dataset in the study area, and it can be applied in other areas for the same purpose. The following objectives were formed to achieve the aims of the study:

**1. To determine the differences between (TOACs and BSRN) the QC set of tests for error detection in hourly GHI data, to develop a local test for error detection based on that difference and to evaluate the role of the local test.**

The two QC (BSRN and TOACs) tests to detect questionable data of hourly GHI were compared for the upper and lower limits of physically possible and extremely rare observations at 20 stations in northeast Iraq (Chapter 4, Section 3.2). The results demonstrate the high percentage difference between BSRN and TOACs (up to 30% among stations) for each subtest due to the different limits. This indicates that BSRN tests cannot be used when only GHI is available because most errors will not be detected. Hence, the extremely rare limit of the TOACs test flagged high rates of GHI data as errors. The local test was developed based on the difference limit between BSRN and TOACs QC tests (Chapter 4, Section 2.2.3). The results of the local test covered the limitations in the BSRN and TOACs QC tests by detecting systematic errors and passing high rates values of the extremely rare limit test. This confirms the role of the local test, and it indicates that those QC tests needed to be modified based on a local climate condition. This is limited in the literature (Geiger et al., 2002; Khaliliaqdam and Soltani, 2012; Long and Shi, 2008; Moradi, 2009; Moreno-Tejera et al., 2015; Ntsangwane et al., 2018; Pashiardis and Kalogirou, 2016; Perez-Astudillo et al., 2018; Roesch et al., 2011; Tang et al., 2010; Zo et al., 2017).

**2. To develop a new QC method for GHI by using the SD and AT variables and to evaluate it, if it works as a consistency test and to detect systematic errors and questionable data by graphics**.

In Chapter 4, the new QC tests were developed by using SD and AT variables to check the consistency of GHI hourly data and to use graphics for a general check of those variables. The new QC tests were applied at 20 stations in northeast Iraq, and the QC tests were validated at six stations worldwide. The results revealed that SD can be used as a partial consistency test, which has been supported by the validation results. On the contrary, AT has not been supported as a suitable test. The new approach based on SD and AT flagged several questionable data, which were not detected by BSRN tests nor by TOACs tests. This indicates that the new methods can improve the two other tests. However, the results also demonstrated that graphics such as fingerprint plot and plotting SD, AT and GHI together can be used to detect systematic errors in the GHI data. This suggests that using other climate variables for checking GHI data is a good option. The QC tests were written in R codes, which is easy for using, modifying and comparing the results of flags for each QC test.

**3. To validate the SDDs in new areas and to investigate the spatiotemporal features of SDDs to ground data, and to investigate the difference of two pixel point data from SDDs around a station compared with ground data of a station.**

Hourly GHI from two SDDs ( HC3v5 and CRSv3) were validated against ground data from nine stations in northeast Iraq for all-sky, clear-sky and cloudy-sky conditions, and two other adjacent pixels data of the two SDDs around a station in all-sky conditions were also compared to the ground data of a station at nine meteorological stations in Chapter 5. The results of validation revealed high accuracy of SDDs in the new region (northeast Iraq) in all-sky and clear-sky condition (*r* values above 0.94) and low performance in the cloudy conditions nearly the same as other studies in other areas for the same SDDs (Amillo et al., 2014; Eissa et al., 2015a; Marchand et al., 2017; Thomas et al., 2016a; Thomas et al., 2016b). This indicates that those SDDs are considered as a good source for solar irradiance data in all-sky and clear conditions despite a rate of bias. The results of the other two points show a minor difference with ground data. This adds further weight for using neighbouring data for a location in the case of a GHI data record limitation.

**4. To evaluate the ability of a simple method for bias correction in SDDs under cloudy sky conditions.**

The simple method of bias correction was applied in Chapter 5 to correct the two SDDs (HC3v5 and CRsv3) in cloudy-sky conditions at nine stations in northeast Iraq. The results illustrated a definite improvement of the two SDDs to represent ground data in cloudy-sky conditions after bias correction. The method required one season or one year of GHI ground measurement for training and calculating coefficients of the regression equation. With one year of trained data, this study has provided the equations and coefficients that can be used to correct SDDs in similar climate regions, and they can be used to create a long-term dataset of GHI from SDDs in the study area.

**5. To assess the role of new input combinations (SDDs. Cs, TOA and climate variables) in estimating and modelling hourly GHI data.**

In Chapter 6, the hourly GHI data has been estimated with ANN and regression models at nine stations in northeast Iraq using several sets of inputs from climate variables, new input (SDD), Cs and TOA. The results of GHI estimation were slightly improved due to new inputs in different scenarios in both ANN and regression models. When some combination inputs were used without SDDs, the results are nearly the same as reported by other literature. However where all inputs from SDDs, TOA, Cs and climate variables were used the RMSE decreased slightly in this study compared to other studies (Dahmani et al., 2016; Hassan et al., 2017b; Ibrahim and Khatib, 2017; Loutfi et al., 2017). This is because those studies did not use SDDs as additional inputs for the hourly GHI modelling. This indicates that the ANN trained models can be used for other stations to estimate GHI with high accuracy (bias near zero and RMSE under 10%) when the inputs are available, but GHI is not available. This is because the ANN models have already been tested and validated with good quality of GHI ground data. In the same way, a long-term dataset of GHI can be formed. The results confirm the role of SDD as inevitable input in GHI modelling. Therefore, it is suggested that the SDD be used in all modelling and prediction of GHI and for other solar components for better results.

## 7.4    Conclusions

This thesis has presented an approach to obtain good quality hourly GHI data in an area of scarce ground measurements, using northeast Iraq, as an example. It contributes to developing a new method for QC GHI, modifying the QC tests according to the local data, validating two SDDs in the new region, applying a method to adjust GHI from SDDs under cloudy-sky conditions and using new inputs in modelling hourly GHI to improve the accuracy of results. The method of QC can be used in the case study to test GHI. This is because, to date, no QC approaches have been applied in the Iraqi Kurdistan region to check GHI data. In addition, it can also be implemented by engineering and researchers in other areas because the study demonstrated that modifying QC tests is working precisely and SD can be used as a consistency test. The study also revealed that SDDs are a good source of solar irradiance data and it applied a simple method with high performance (reducing bias by 10–80%) for correcting bias in SDDs under cloudy-sky conditions. The thesis proved that hourly GHI data with nearly 0% bias and RMSE under 10% can be obtained with new input combinations from SDDs, climate variables, Cs and TOA. Researchers can also apply the new input combinations in models to obtain high accuracy of hourly GHI data. In addition, M9 and M10 of ANN models are highly recommended to be used in a data bank of solar irradiance in Iraq and in Middle East. In general, the results of the study illustrated the idea for obtaining GHI and forming GHI datasets, which can be used by policymakers for solar energy projects or by scientists for the purpose of agriculture, architecture and hydrological modelling.

## 7.5    Limitations and Future Research Recommendations

There are some limitations in the data and in the methods that have been utilised in this thesis, which require further research. They will be provided for each chapter.

**In Chapter 4**, first, the chosen limits for the SD and AT tests might not be perfect. For example, the upper limit of SD was set to 50 and from 30–50 minutes in one hour, which is based on an assumption about why SD is high and GHI is low; there might be some questionable data under this limit. Second, the mean and the half of mean AT in each month were used to test GHI with 10% and 35% of TOA, which tends not to be perfect. This is mainly because there are some times when this situation may happen naturally. Next, owing to the limited record of minimum and maximum AT in many

stations in the study area, the chapter used mean AT for tests. The SD and AT do not check every observation because they have a limit border, unlike the consistency test. Finally, the methods for the quality check cannot detect several random errors such as minor calibration error and low performance of the equipment and others. This requires developing new methods by using SDDs or using reference pyranometers or any other possible approach to detect random errors. In addition, further research is also required to compare BSRN and TOACs tests in other areas. Using several arguments with mean AT or with minimum and maximum AT, and also using other climate variables to check the quality of GHI data will be useful.

**In Chapter 5**, the challenges are the different data timescales from one station to another, lack of several years of ground data and the limited information available for some variables including aerosols and local atmospheric properties in which they are useful to analyse the rate of bias and RMSE in each sky condition.

Further research would also be useful for validating the SDDs in other climates with several years of data, and to compare most of the available SDDs in a local area or large regions based on their availability. This is to show the reliability of each of them based on the data and the method used for forming them. Some further studies are also required to address the inputs to the Heliosat method, according to regional and local factors, for a better estimation of GHI from satellite images.

**In Chapter 6**, the study focused on only estimating GHI but no other solar components, which are required for some fields of solar energy, such as concentrated solar power. Further research might estimate DNI and DHI from GHI in areas with a scarcity of ground data, like other areas where other studies have been conducted. Again, another limitation is the scarcity of long-term GHI ground data at time scales beyond five years or more. It would be better to train ANN models with more years of data to improve estimation. It might be possible to apply the same method to an area of long-term data for further investigation.

Further research is also required to utilise the new input combinations provided in this thesis with other machine learning approaches, namely support vector machines and random forests, or with bagging and boosting algorithms.

**Appendix A – Chapter 4 Additional Figures and Tables**
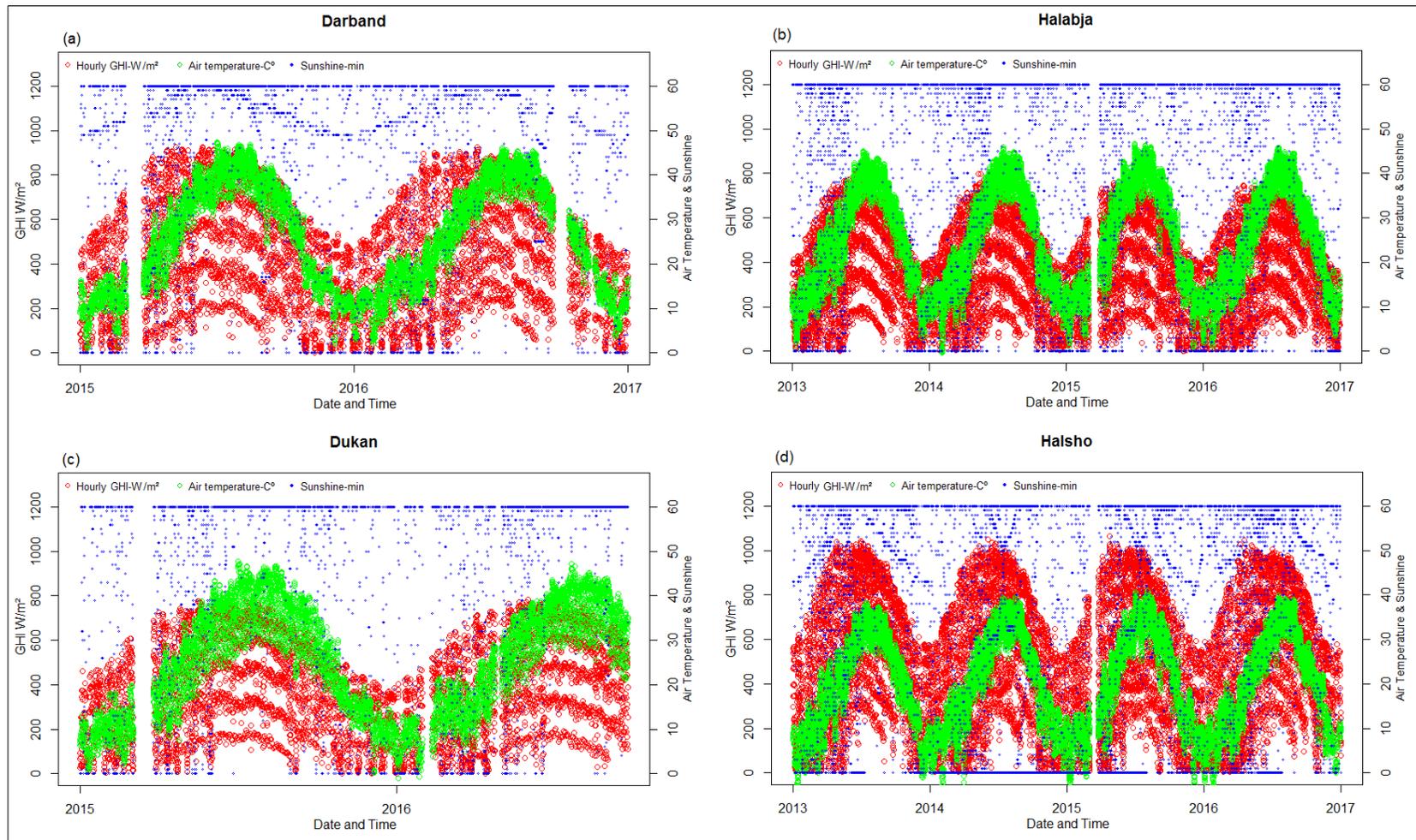
**Figure A1:** Scatterplots of GHI W/m² (first y-axis) and SD in minutes per hour with AT in °C (second y-axis) for all hourly time series data in each station.
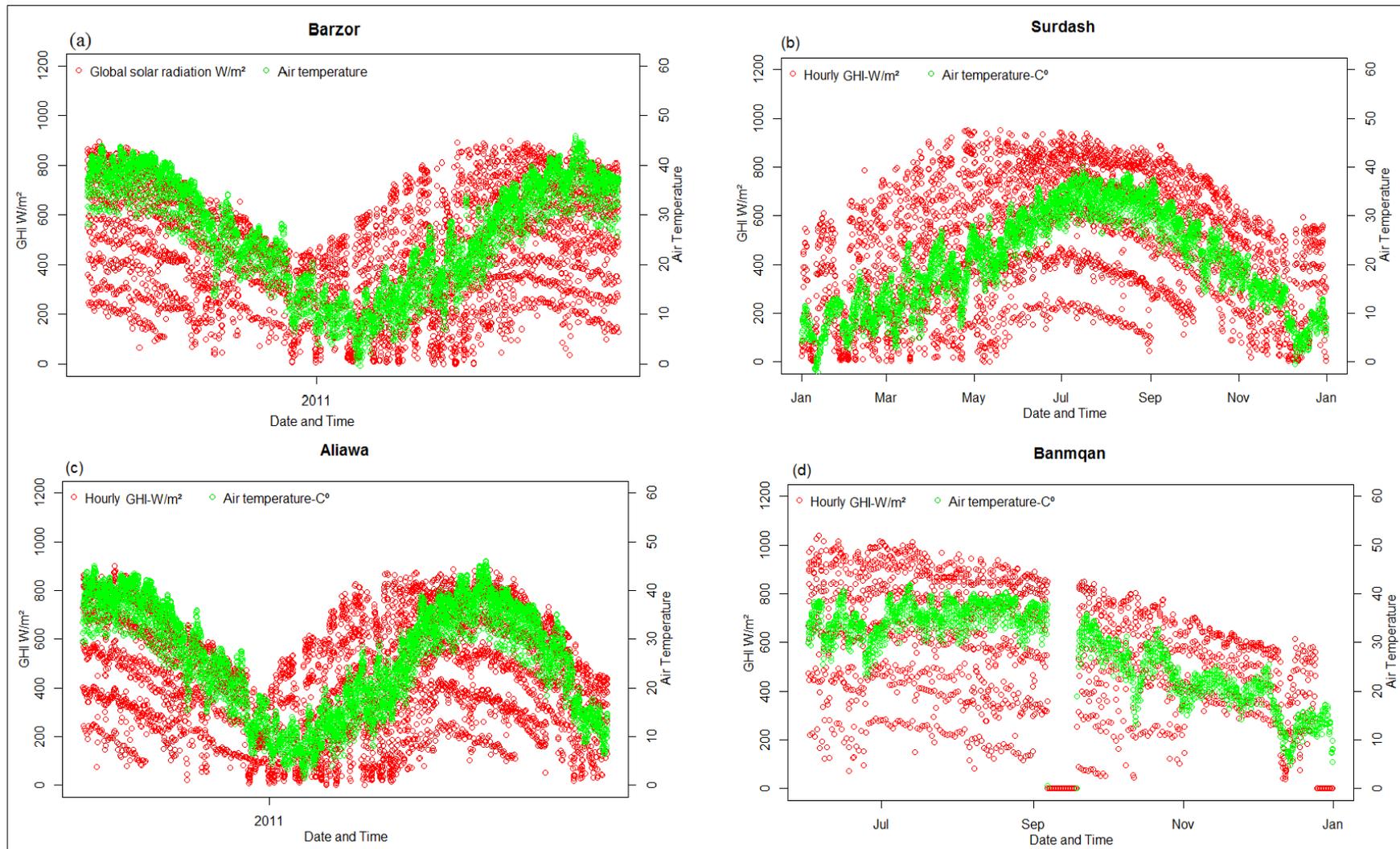
**Figure A2:** Scatterplots of GHI W/m² (first y-axis) and AT in °C (second y-axis) for all hourly time series data in each station.

141
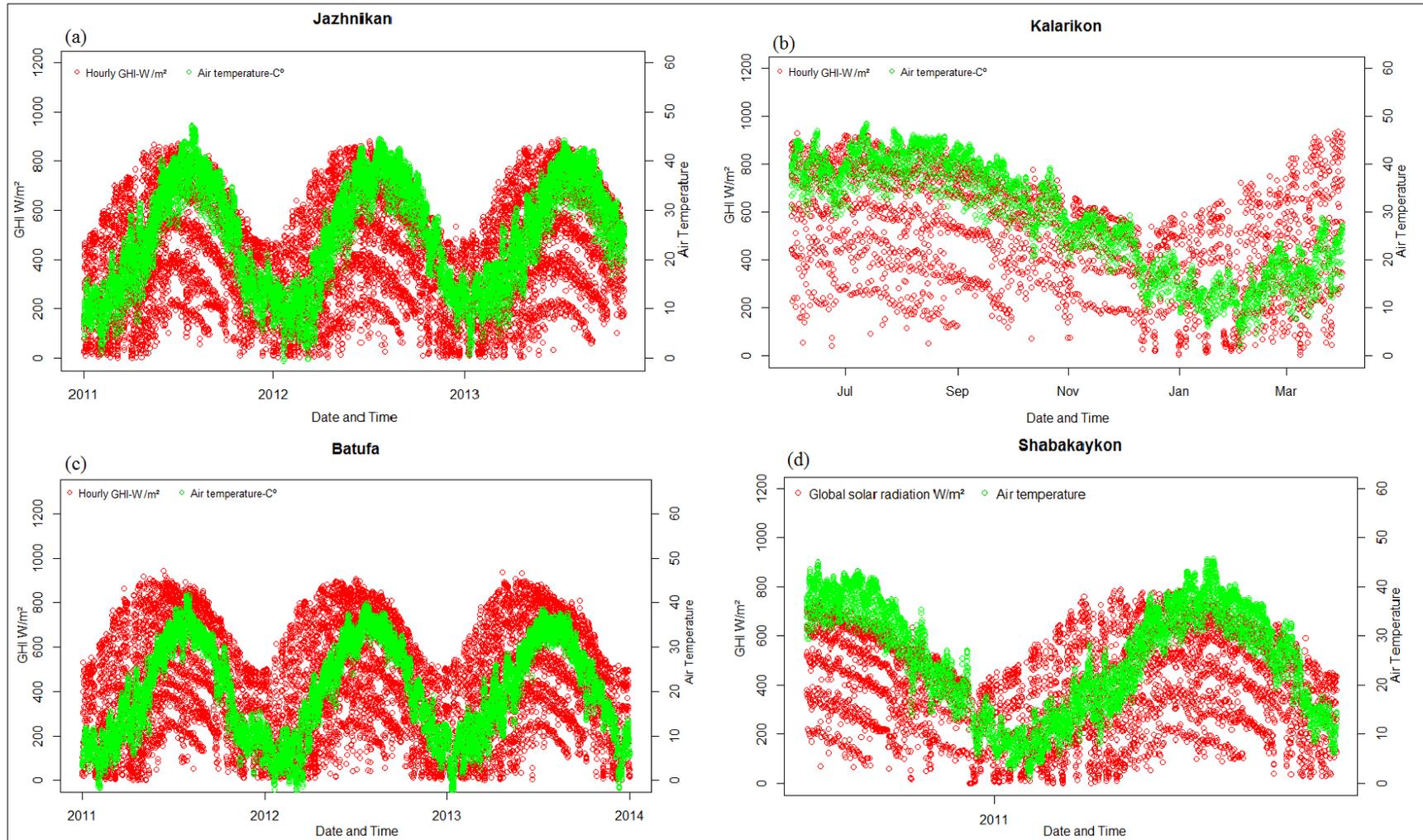
**Figure A3:** Scatterplots of GHI W/m² (first y-axis) and AT in °C (second y-axis) for all hourly time series data in each station.
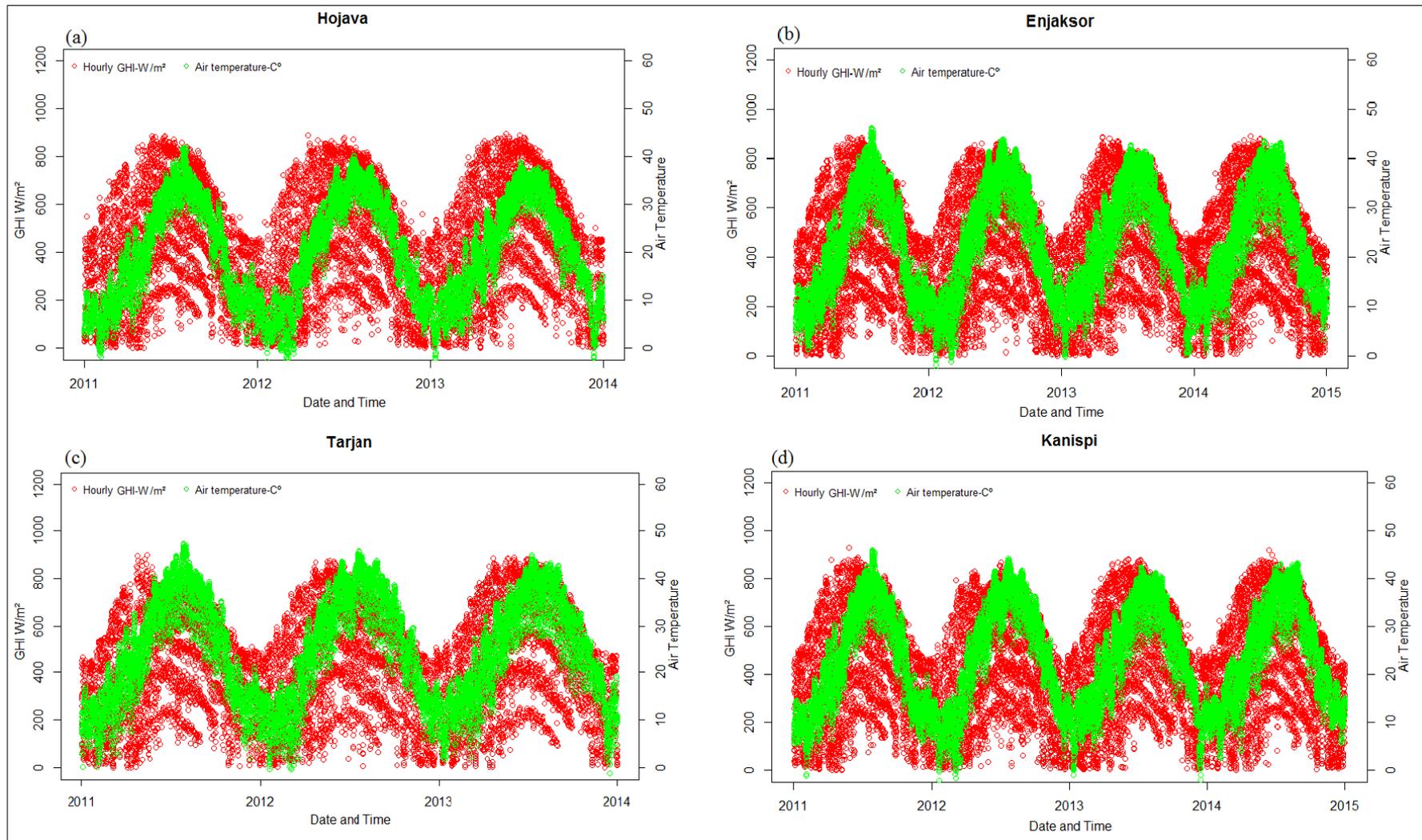
**Figure A4:** Scatterplots of GHI W/m² (first y-axis) and AT in °C (second y-axis) for all hourly time series data in each station.
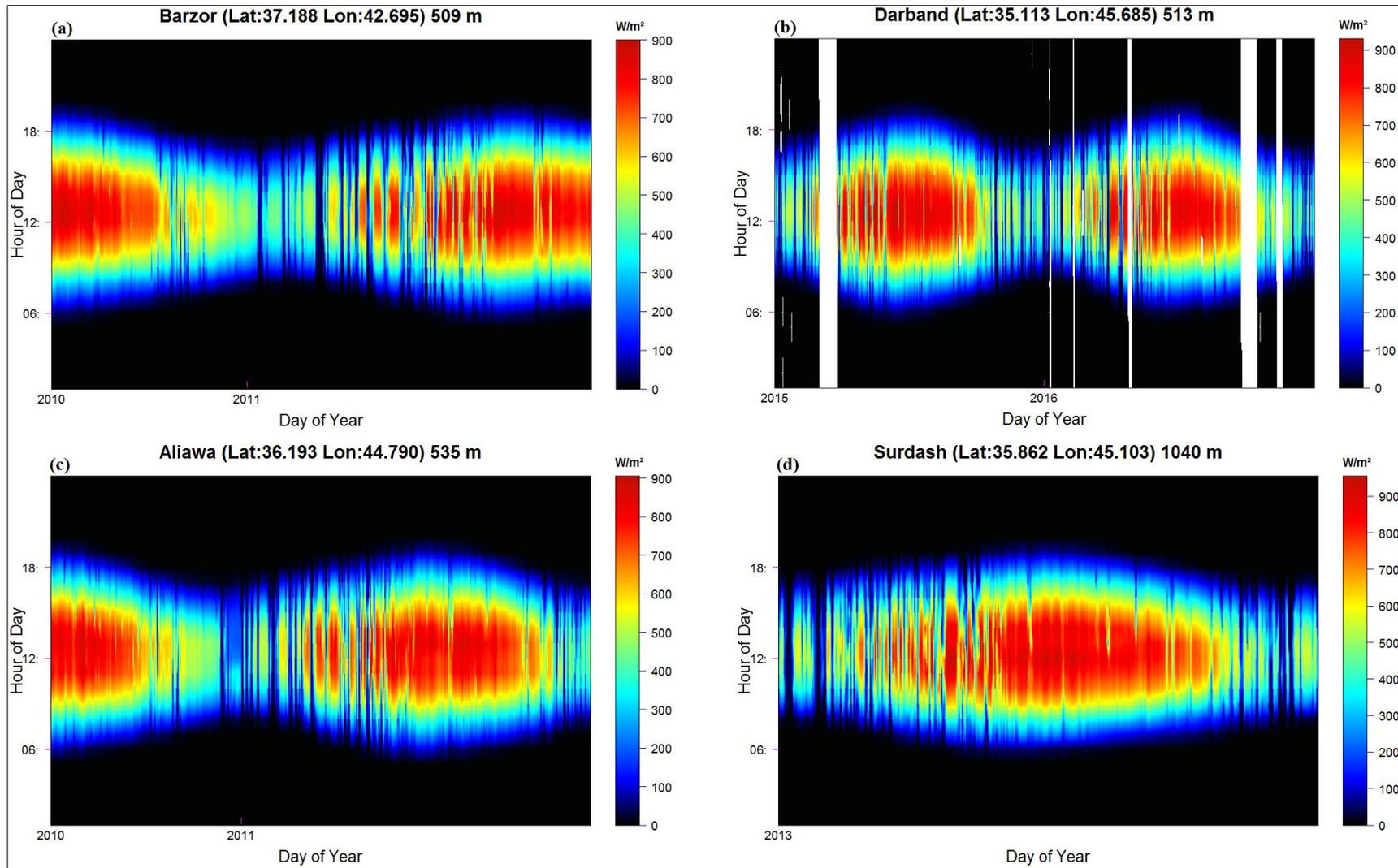
**Figure A5:** Fingerprint plot GHI time series of hourly data for each station, white colour shows NA values.
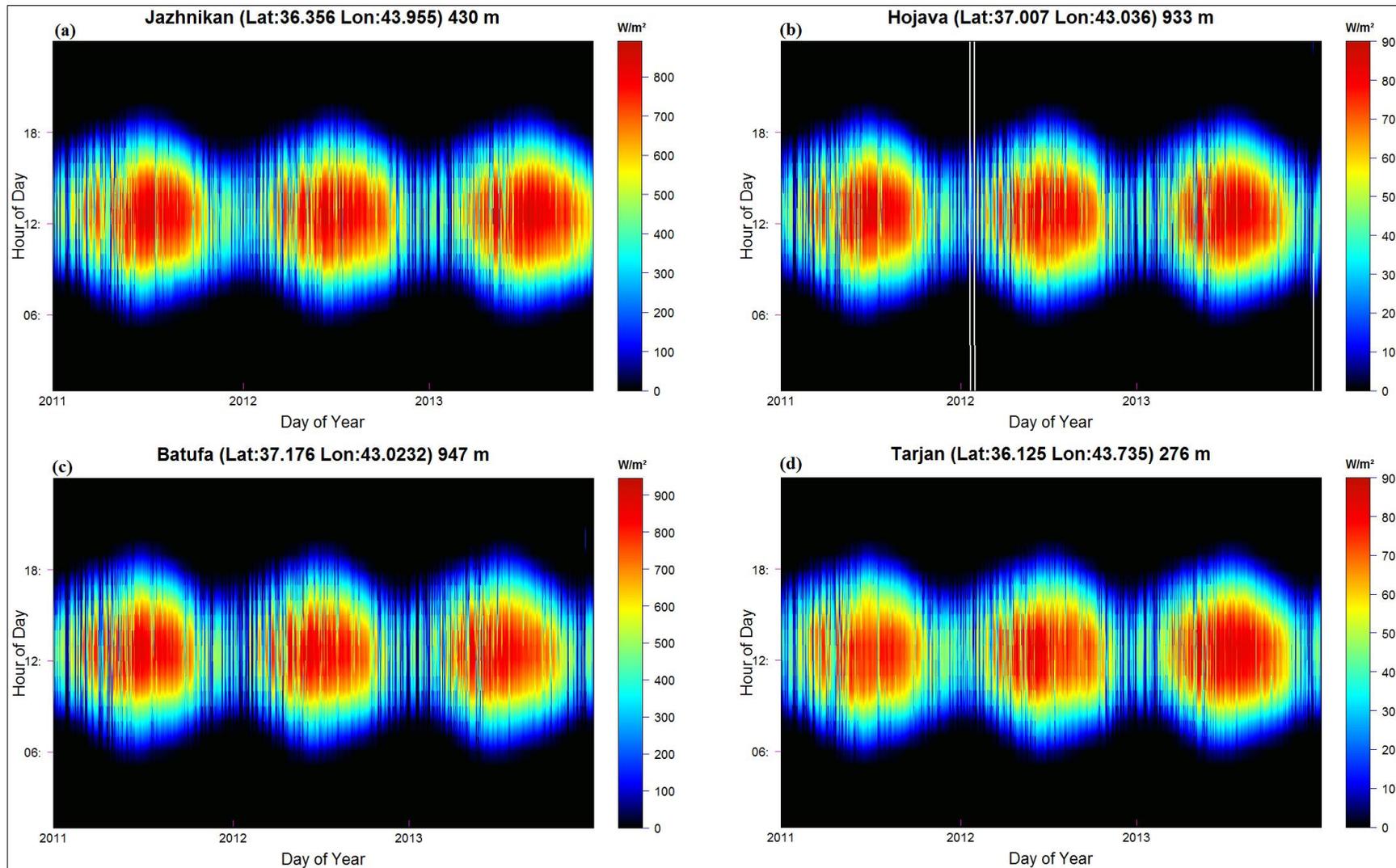
**Figure A6:** Fingerprint plot GHI time series of hourly data for each station, white colour shows NA values.
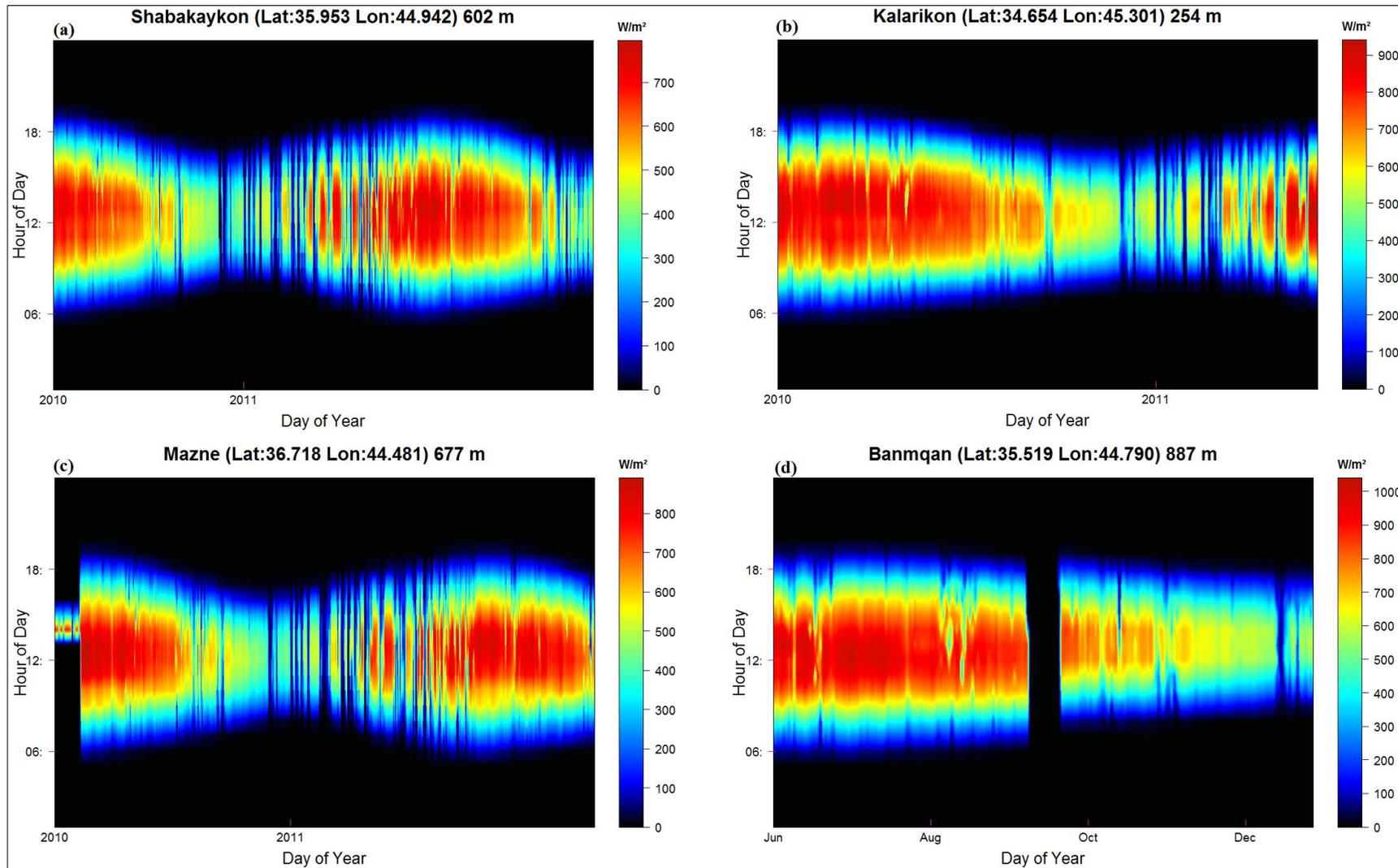
**Figure A7:** Fingerprint plot GHI time series of hourly data for each station, white colour shows NA values.
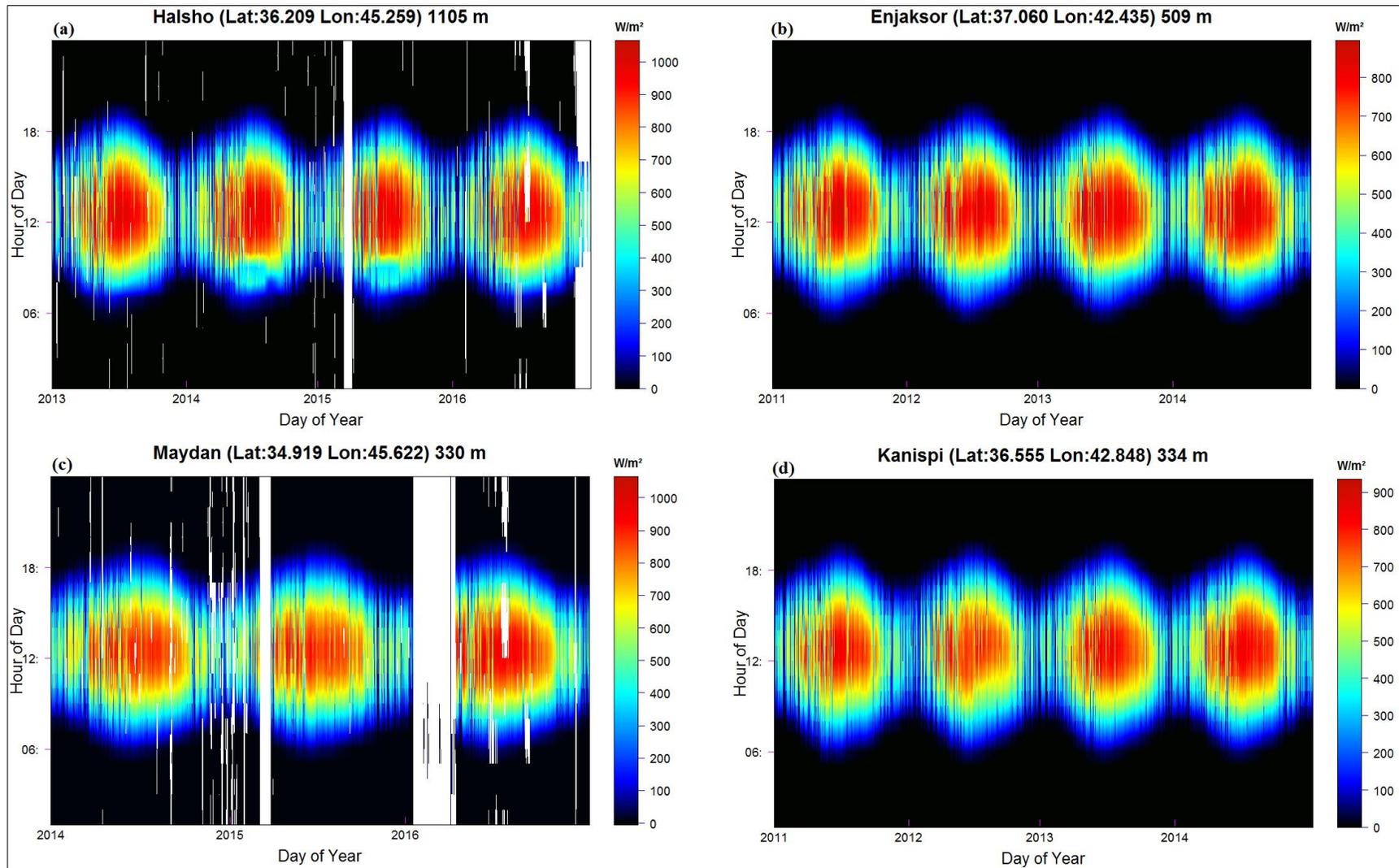
**Figure A8:** Fingerprint plot GHI time series of hourly data for each station, white colour shows NA values.

**Table A1:** R codes for fingerprint plot to test GHI data.

```
# Fingerprint plot to check GHI data
# Note: The code was originally written by Dr Ross Morrison, Biometeorologist, CEH Wallingford for
DEFRA SP1210 Lowland Peatlands purpose (2014). Then it was adopted by Dr Alex Cumming,
Department of Geography, University of Leicester for air temperature variable (2016). Finally, it was
modified by Bikhtiyar Ameen, Department of Geography, University of Leicester for GHI hourly data
(September 2017).
# The example here is applied at Halsho station in the study area.
# Need a variable (VAR.x), Year & Month
VAR.GHI<- Halsho$GHI # put GHI data as variable here
# Create a Year variable as a column variable.
Halsho[,"Year"]<- strftime((as.POSIXct(Halsho$date, format= "%Y-%m-%d %HH:%MM")), format=
"%Y")
Year<- Halsho$Year
# Create a Month variable as a column variable.
Halsho[,"Month"]<- strftime((as.POSIXct(Halsho$date, format= "%Y-%m-%d %HH:%MM")),
format=  "%m")
Month<- Halsho$Month
##
Year <- as.numeric(Year)
Month <- as.numeric(Month)
# 24 returns to the number of data records of each hour at daytime and nighttime.  # If data is recorded
at each half of hour, it will change the 24 to 48.
mat.var <- matrix(data = GHI.x, nrow =24, ncol = length(VAR.x)/24)
y <- 1:nrow(mat.var)
x <- 1:ncol(mat.var)
par(mar=c(5,4.5,4,3)+0.1,mgp=c(2.2,1,0)) # To set margins
par(fg = NA,col="black") # to remove line colour in the legend
### Set up colour scheme
fJetColors <- colorRampPalette(c('#00007F', 'blue', '#007FFF', 'cyan', '#7FFF7F', 'yellow', '#FF7F00',
'red', '#7F0000'))
Jet.n <- 200
# plot with countor
filled.contour(x=x,
        y=y,
        z=t(mat.var),
        color.palette=colorRampPalette(c(fJetColors(Jet.n))),
        plot.title=title(main="Halsho (Lat:36.209 Lon:45.259) 1105 m",cex.main=1.8, sub=" ",
                xlab="Day of Year", ylab="Hour of Day",cex.lab=1.6),
        nlevels=256,
        plot.axes = { axis(side = 1, at = c(1,365,721,1086), labels
=c("2013","2014","2015","2016"),pos = 1.5, outer = F,tick = T,col.ticks ="#fe03ff",cex.axis=1.2)
         axis(side = 2, at = c(6,12,18), labels = c("06:","12:","18:"),cex.axis=1.2,pos = 1.5, outer =
F,tick =T,col.ticks ="#fe03ff")},
        key.title=title(main="W/m²",adj = 0,line = 0.8),
        key.axes = axis(4, line = F,col.ticks =4, seq(0, 1200, by = 100),cex.axis=1.2))
#End
```

**Table A2:** R codes for full applied QC method.

```
# QC of GHI data based on BSRN, TOACs, and SD and AT tests. #September 2017#
#Managing the data for each station from its hourly time series with GHI, SD and AT of ground data,
Cs, TOA, SE, SZϴ and Cos SZϴ Calculated data as data frame named StationDF = data frame of a
station with variables from the relevant sources.
# Import station csv data file to R which contains all relevant variables.
StationDF <- read.csv(file.choose(), header = T)
# Creating a full-time series base on the first available hour data and the end available hour data of a
station data named as date.time.
```

```r
xStationDF <- seq(as.POSIXct("2013-01-01 0:00:00", tz = "UTC"), as.POSIXct("2016-12-31
23:00:00", tz="UTC"), by="hour")
names(xStationDF)[1] <- "date.time"
# Merging the created time series with a time series of the station to detect any gaps and set them as
NAs automatically.
# Wth new name StationDF1
StationDF1 <- merge(xStationDF, StationDF, by = "date.time", all.x = TRUE)
#Calculating the variables that required for QC test based on the solar constant (1367) and other
variables in the station data frame.
# To test Upper physically possible limit BSRN. See table (4.1)
StationDF1$BSRN.P<- round((((1367/StationDF1$SE^2*1.5*(StationDF1$CosSZΘ^1.2))+100),1)
# To test Extremely rare limit BSRN.
StationDF1$BSRN.R <- round((((1367/StationDF1$SE^2*1.2*(StationDF1$CosSZΘ^1.2))+50),1)
# To test Lower physical possibly limit TOACs.
StationDF1$M0.03 <- round(StationDF1$TOA*0.03,1)
# To count 10% of TOA and used it with SD and AT tests.
StationDF1$TO0.1 <- round(StationDF1$TOA*0.1,1)
# To count 35% of TOA and used it with SD and AT tests.
StationDF1$Mdiff <- round(StationDF1$TOA*0.35,1)
# To test Extremely rare limit TOACs.
StationDF1$C1.1 <- round(StationDF1$Cs*1.1,1)
# To count 120% of TOA and used it to test Upper physically possible limit local test.
StationDF1$TO1.2 <- round(StationDF1$TOA*1.2,1)
# To test Lower physical possibly limit local test.
StationDF1$M0.005 <- round(StationDF1$TOA*0.005,1)
# To test Extremely rare limit local test.
StationDF1$C1.25 <- round(StationDF1$Cs*1.25,1)
#create a new variable as Month to count the number of each month with its number in the time series
for calculating the mean of AT in each month to later uses of AT tests.
StationDF1[,"Month"]<- strftime((as.POSIXct(StationDF$date.time, format= "%Y-%m-%d
%HH:%MM")), format= "%m")
# Calculate mean AT for each month # Number 18 at the end of each code is a variable sequence of
AT in the station data frame.
s1 <- summary(StationDF1[StationDF1$Month=="01",18])
s2 <- summary(StationDF1[StationDF1$Month=="02",18])
s3 <- summary(StationDF1[StationDF1$Month=="03",18])
s4 <- summary(StationDF1[StationDF1$Month=="04",18])
s5 <- summary(StationDF1[StationDF1$Month=="05",18])
s6 <- summary(StationDF1[StationDF1$Month=="06",18])
s7 <- summary(StationDF1[StationDF1$Month=="07",18])
s8 <- summary(StationDF1[StationDF1$Month=="08",18])
s9 <- summary(StationDF1[StationDF1$Month=="09",18])
s10 <- summary(StationDF1[StationDF1$Month=="10",18])
s11 <- summary(StationDF1[StationDF1$Month=="11",18])
s12 <- summary(StationDF1[StationDF1$Month=="12",18])
# make data frame of mean AT
meanATStationDF <- rbind(s1,s2,s3,s4,s5,s6,s7,s8,s9,s10,s11,s12)
meanATStationDF <- as.data.frame (meanATStationDF)
# set new variable as half of mean AT.
meanATStationDF [,7] <- meanATStationDF[,4]/2
############################################################
# Apply QC method with flags
# Setting flags according to tests (BSRN and TOACs) see table (4.1)
StationDF1$f2 <- ifelse(StationDF1$GHI<StationDF1$BSRN.E, c(1), c(2))
StationDF1$f3 <- ifelse(StationDF1$GHI<StationDF1$TOA, c(1), c(3))
enjaksorD$f4 <- ifelse(enjaksorD$mV> -2, c(1), c(4))
StationDF1$f5 <- ifelse(StationDF1$GHI>StationDF1$M0.03, c(1), c(5))
StationDF1$f6 <- ifelse(StationDF1$GHI<StationDF1$BSRN.R, c(1), c(6))
StationDF1$f7 <- ifelse(StationDF1$GHI<StationDF1$C1.1, c(1), c(7))
# Setting flags according to local tests see table (4.1)
```

```
StationDF1$f8 <- ifelse(StationDF1$GHI<StationDF1$TOA1.2, c(1), c(8))
StationDF1$f9 <- ifelse(StationDF1$GHI>StationDF1$M0.005, c(1), c(9))
StationDF1$f10 <- ifelse(StationDF1$GHI<StationDF1$C1.25, c(1), c(10))
## SD test for its upper bound, Why SD is high whereas GHI is low.
StationDF1$f11 <- ifelse(StationDF1$SD > 50 & StationDF1$GHI < StationDF1$Mdiff |
StationDF1$SD < 50 & StationDF1$SD > 30 & StationDF1$GHI < StationDF1$TO0.1, c(11), c(1))
# SD test for its lower bound, Why SD is low whereas GHI is high.
StationDF1$f12 <- ifelse(StationDF1$SD==0 & StationDF1$GHI > StationDF1$Mdiff, c(12), c(1))
# AT test for its upper bound, Why AT is higher than its mean in the month, whereas GHI is lower
than 10% of TOA. # Number[x,4] returns to the column number of mean AT in a (meanATStationDF)
data frame.
StationDF1$f13 <-
ifelse(StationDF1$Month=="01"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATSt
ationDF[1,4] |
StationDF1$Month=="02"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[2,4] |
StationDF1$Month=="03"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[3,4] |
StationDF1$Month=="04"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[4,4] |
StationDF1$Month=="5"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStationD
F[5,4] |
StationDF1$Month=="06"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[6,4] |
StationDF1$Month=="07"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[7,4] |
StationDF1$Month=="08"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[8,4] |
StationDF1$Month=="09"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[9,4] |
StationDF1$Month=="10"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[10,4] |
StationDF1$Month=="11"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[11,4] |
StationDF1$Month=="12"&StationDF1$GHI<StationDF1$TO0.1&StationDF1$AT>meanATStation
DF[12,4], c(13), c(1))
# AT test for its lower bound, Why AT is lower than its half of mean in the month, whereas GHI is
higher than 35% of TOA. # Number[x,7] returns to the column number of half of mean AT in a
(meanATStationDF) data frame.
StationDF1$f14 <-
ifelse(StationDF1$Month=="01"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATSt
ationDF[1,7] |
StationDF1$Month=="02"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[2,7] |
StationDF1$Month=="03"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[3,7] |
StationDF1$Month=="04"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[4,7] |
StationDF1$Month=="5"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[5,7] |
StationDF1$Month=="06"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[6,7] |
StationDF1$Month=="07"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[7,7] |
StationDF1$Month=="08"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[8,7] |
StationDF1$Month=="09"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[9,7] |
StationDF1$Month=="10"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[10,7] |
```

```r
StationDF1$Month=="11"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[11,7] |
StationDF1$Month=="12"&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationD
F[12,7], c(14), c(1))
# All arguments of SD and AT for their upper bound with low GHI in one argument for each month.
To check both SD, AT and GHI together.
StationDF1$f15 <-
ifelse(StationDF1$Month=="01"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&Statio
nDF1$AT>meanATStationDF[1,4] |
StationDF1$Month=="02"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[2,4]|
StationDF1$Month=="03"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[3,4] |
StationDF1$Month=="04"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[4,4] |
StationDF1$Month=="5"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1$
AT>meanATStationDF[5,4] |
StationDF1$Month=="06"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[6,4] |
StationDF1$Month=="07"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[7,4] |
StationDF1$Month=="08"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[8,4] |
StationDF1$Month=="09"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[9,4] |
StationDF1$Month=="10"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[10,4] |
StationDF1$Month=="11"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[11,4] |
StationDF1$Month=="12"&StationDF1$SD>50&StationDF1$GHI<StationDF1$TO0.1&StationDF1
$AT>meanATStationDF[12,4], c(15), c(1))
# All arguments of SD and AT for their lower bound with high GHI in one argument for each month.
To check both SD, AT and GHI together.
StationDF1$f16 <-
ifelse(StationDF1$Month=="01"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&Statio
nDF1$AT<meanATStationDF[1,7] |
StationDF1$Month=="02"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[2,7] | StationDF1$Month=="03"&StationDF1$SD
==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$AT<meanATStationDF[3,7] |
StationDF1$Month=="04"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[4,7] |
StationDF1$Month=="5"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[5,7] |
StationDF1$Month=="06"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[6,7] |
StationDF1$Month=="07"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[7,7] |
StationDF1$Month=="08"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[8,7] |
StationDF1$Month=="09"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[9,7] |
StationDF1$Month=="10"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[10,7] |
StationDF1$Month=="11"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[11,7] |
StationDF1$Month=="12"&StationDF1$SD==0&StationDF1$GHI>StationDF1$Mdiff&StationDF1$
AT<meanATStationDF[12,7], c(16), c(1))
####################################################
# Before final count to the rate of flags, the flags can be selected based on the solar zenith angle (SZƟ)
for selecting specific time or day time.
```

```r
StationDF1 <- StationDF1 [ which(StationDF1$SZϴ < 75),]
# for daytime
StationDF1 <- StationDF1 [ which(StationDF1$SZϴ < 90),]
# To set night-time to zero(0) for all flags as the two examples is shown, before final calculating of the
rates.
StationDF1$f2 <- ifelse(StationDF1$SZϴ>90,  c(0), c(StationDF1$f2))
StationDF1$f3 <- ifelse(StationDF1$SZϴ>90,  c(0), c(StationDF1$f3))
# To delete NAs before final count if it is required.
StationDF1 <- na.omit(StationDF1)
#######################################################
# Calculating data pass according to BSRN tests.
StationDF1$BSRN <- ifelse(StationDF1$f2==1 & StationDF1$f6==1 , c(1), (15))
# Calculating data pass according to TOACs tests.
StationDF1$HeliClim <- ifelse(StationDF1$f3==1 & StationDF1$f5==1 & StationDF1$f7==1 , c(1),
(16))
# Calculating data pass according to local test.
StationDF1$HeliClim <- ifelse(StationDF1$f8==1 & StationDF1$f9==1 & StationDF1$f10==1 , c(1),
(20))
# Calculating data pass according to AT and all other previous tests.
StationDF1$ATtest <- ifelse(StationDF1$f2==1 & StationDF1$f3==1 & StationDF1$f5==1 &
StationDF1$f6==1 & StationDF1$f7==1 & StationDF1$f13==1 & StationDF1$f14==1, c(1), (17))
# Calculating data pass according to AT and all other previous tests.
StationDF1$SDtest <- ifelse(StationDF1$f2==1 & StationDF1$f3==1 &  StationDF1$f5==1 &
StationDF1$f6==1 & StationDF1$f7==1 & StationDF1$f11==1 & StationDF1$f12==1, c(1), (18))
# Calculating data pass according to mix arguments of SD and AT tests and all BSRN and TOACs
tests.
StationDF1$mixATSD <- ifelse(StationDF1$f2==1 & StationDF1$f3==1 &  StationDF1$f5==1 &
StationDF1$f6==1 & StationDF1$f7==1 & StationDF1$f15==1 & StationDF1$f16==1, c(1), (19))
#########################
# For counting percentage of each flag
tblFun <- function(x){
  tbl <- table(x)
  res <- cbind(tbl,round(prop.table(tbl)*100,2))
  colnames(res) <- c('Count','Percentage')
  res
}
# This [27:43] returns to the column number of flags.
perStationDF <- do.call(rbind,lapply(StationDF1[27:43],tblFun))
###############################################################
# To save data as tables.
write.csv(perStationDF, "StationDFper.csv")
write.csv(StationDF1, "StationDF1.csv")
#End
```

**Appendix B – Chapter 5 Additional Figures and Tables**
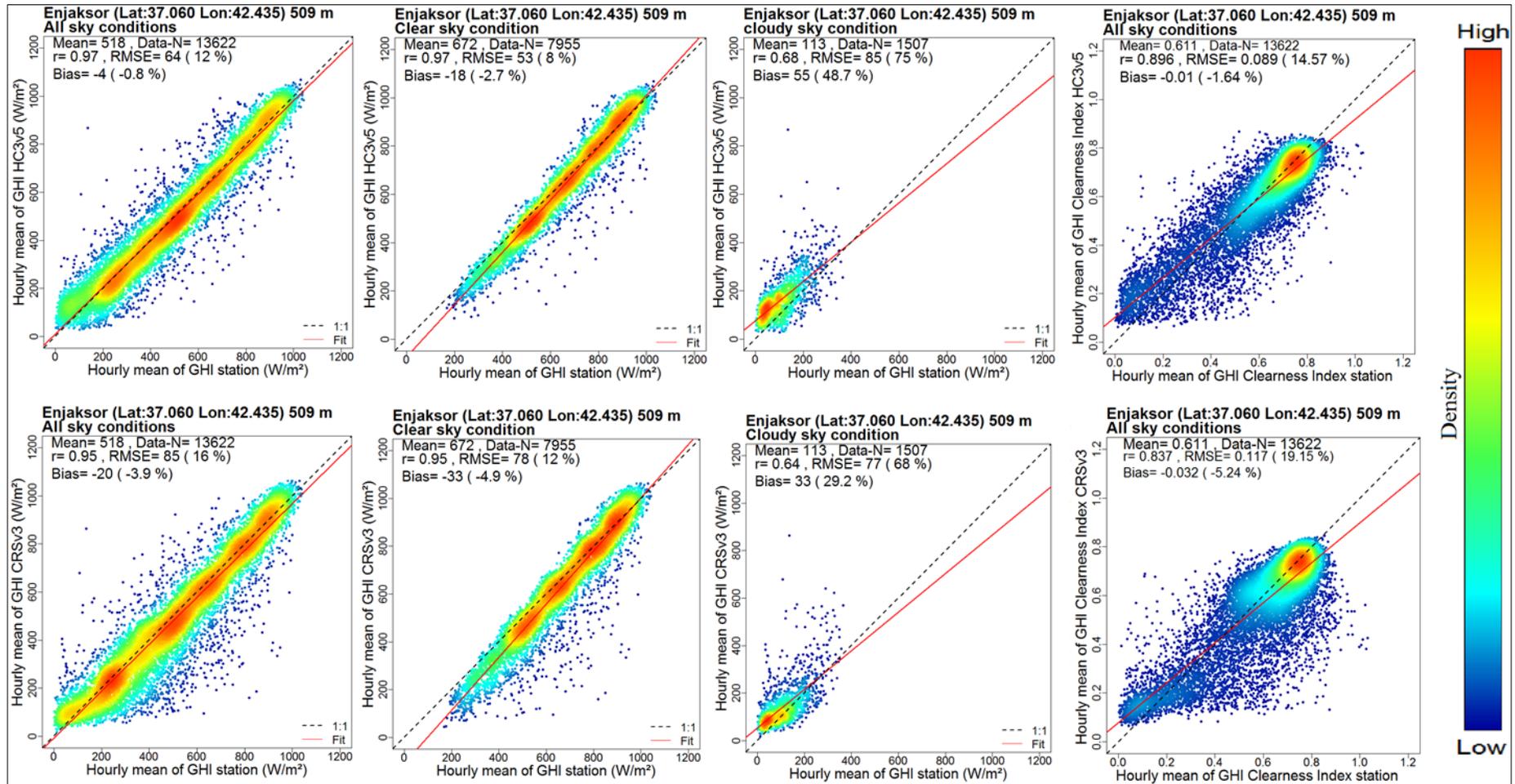
**Figure B1:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Enjaksor station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.
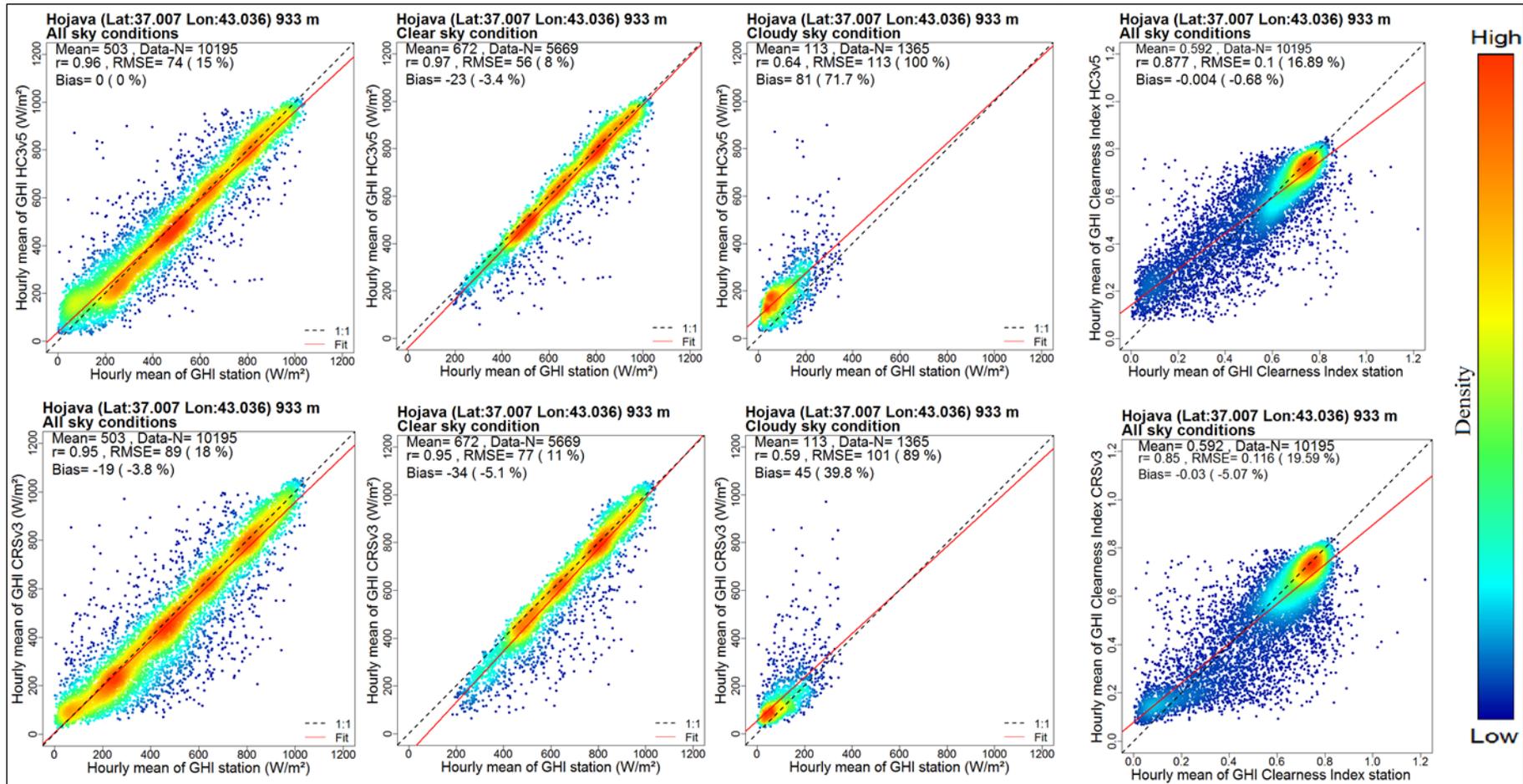
**Figure B2:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Hojava station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.
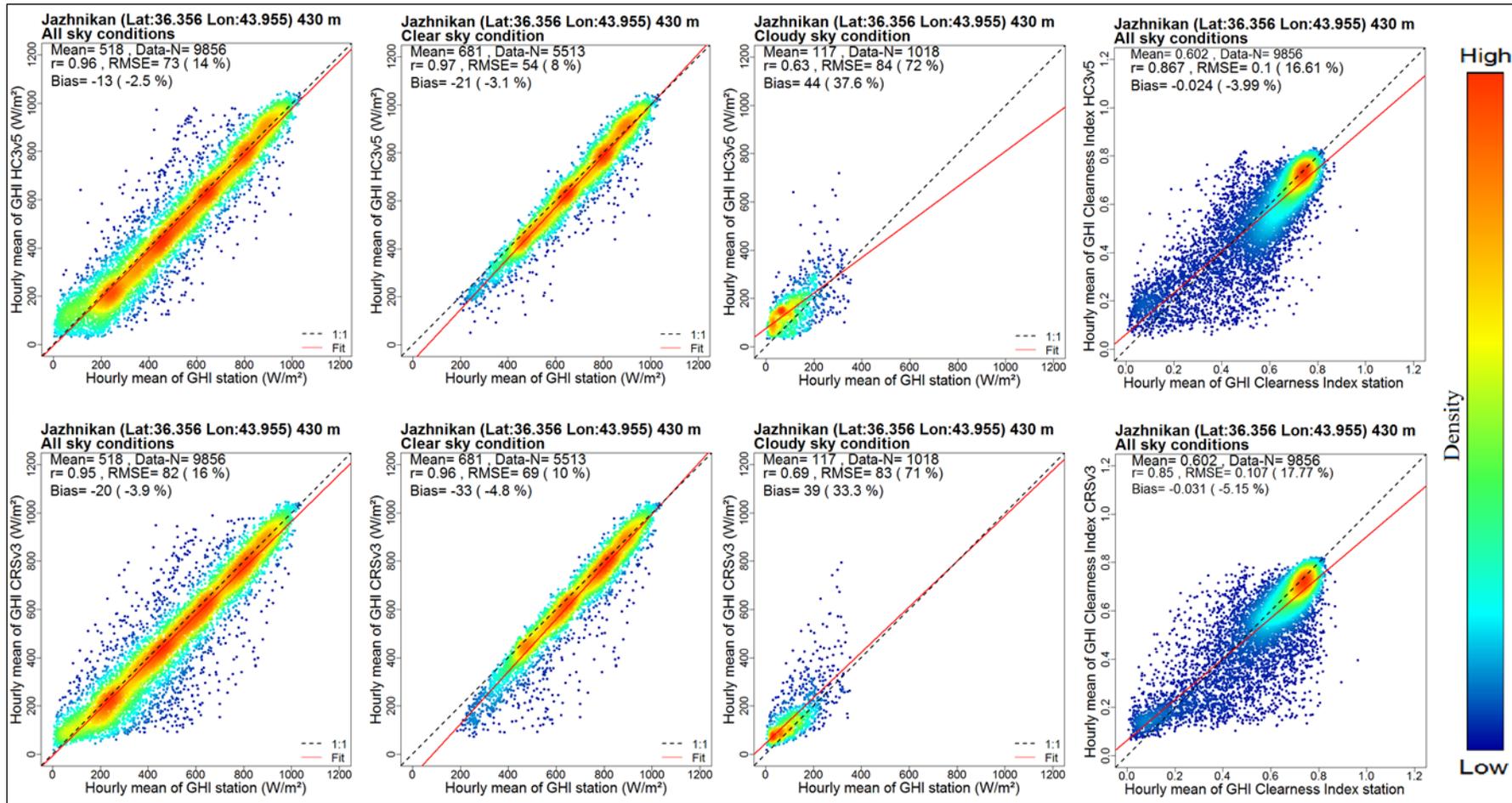
**Figure B3:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Jazhnikan station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.
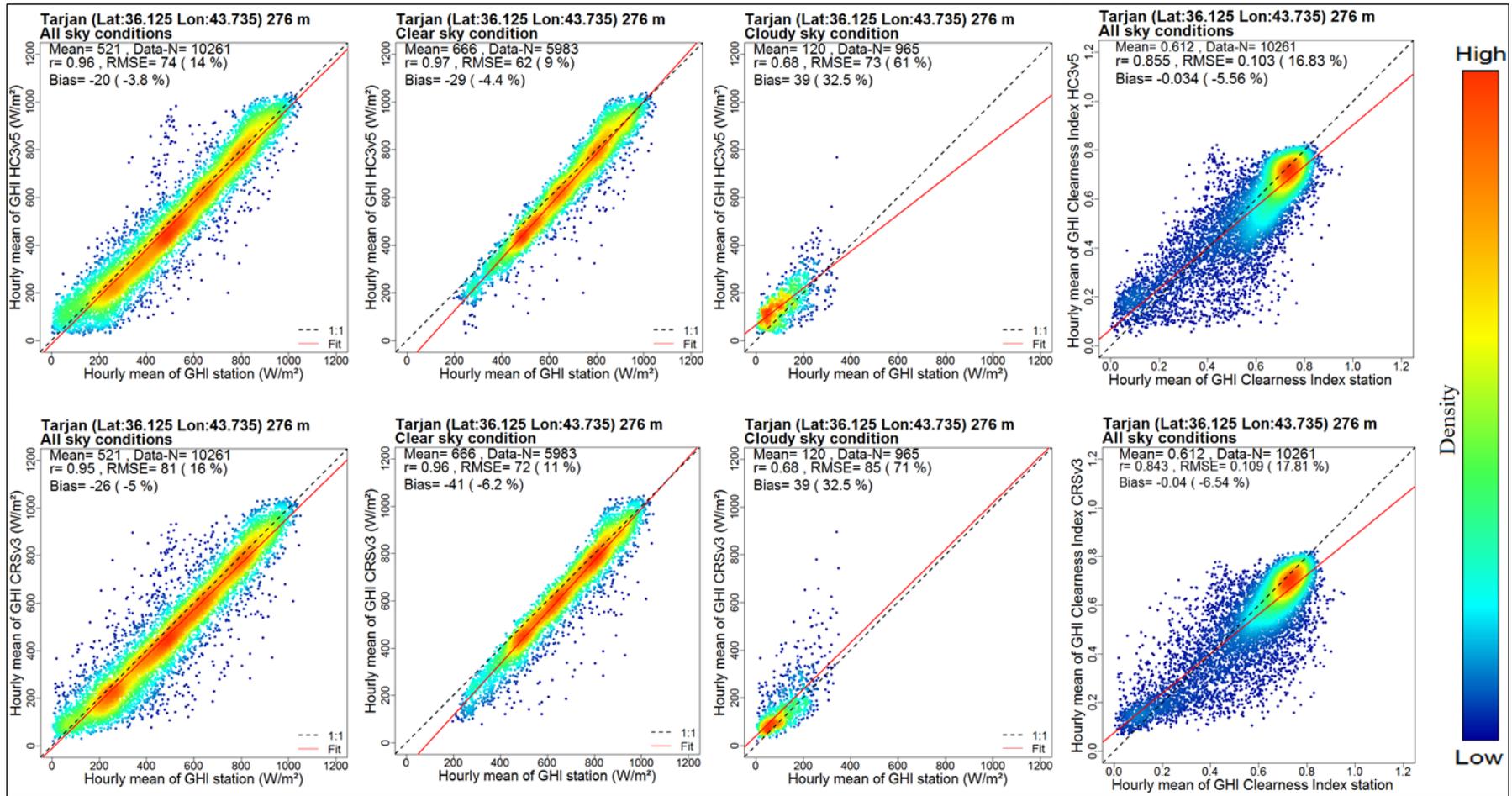
**Figure B4:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Tarjan station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.
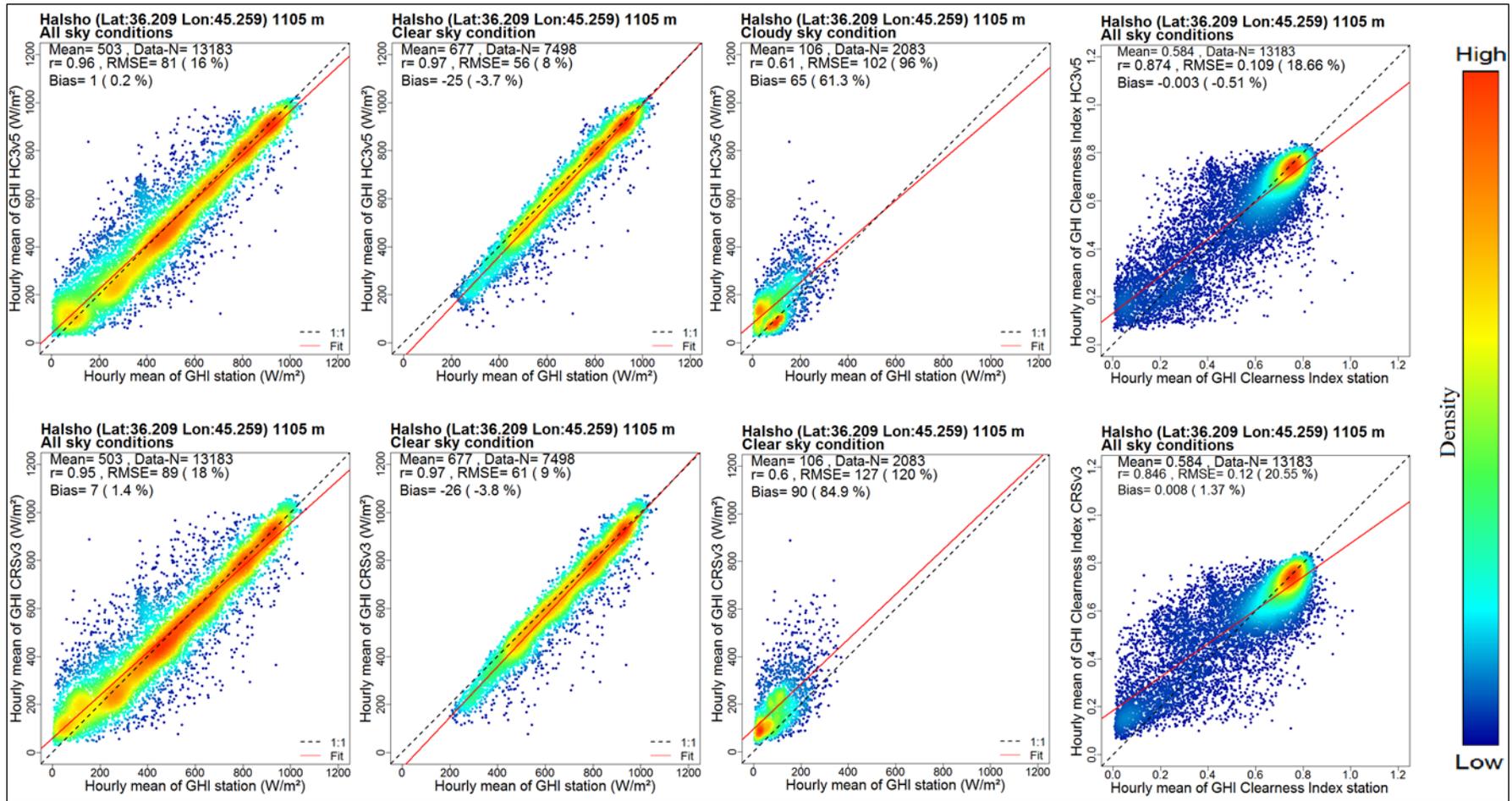
**Figure B5:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Halsho station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.
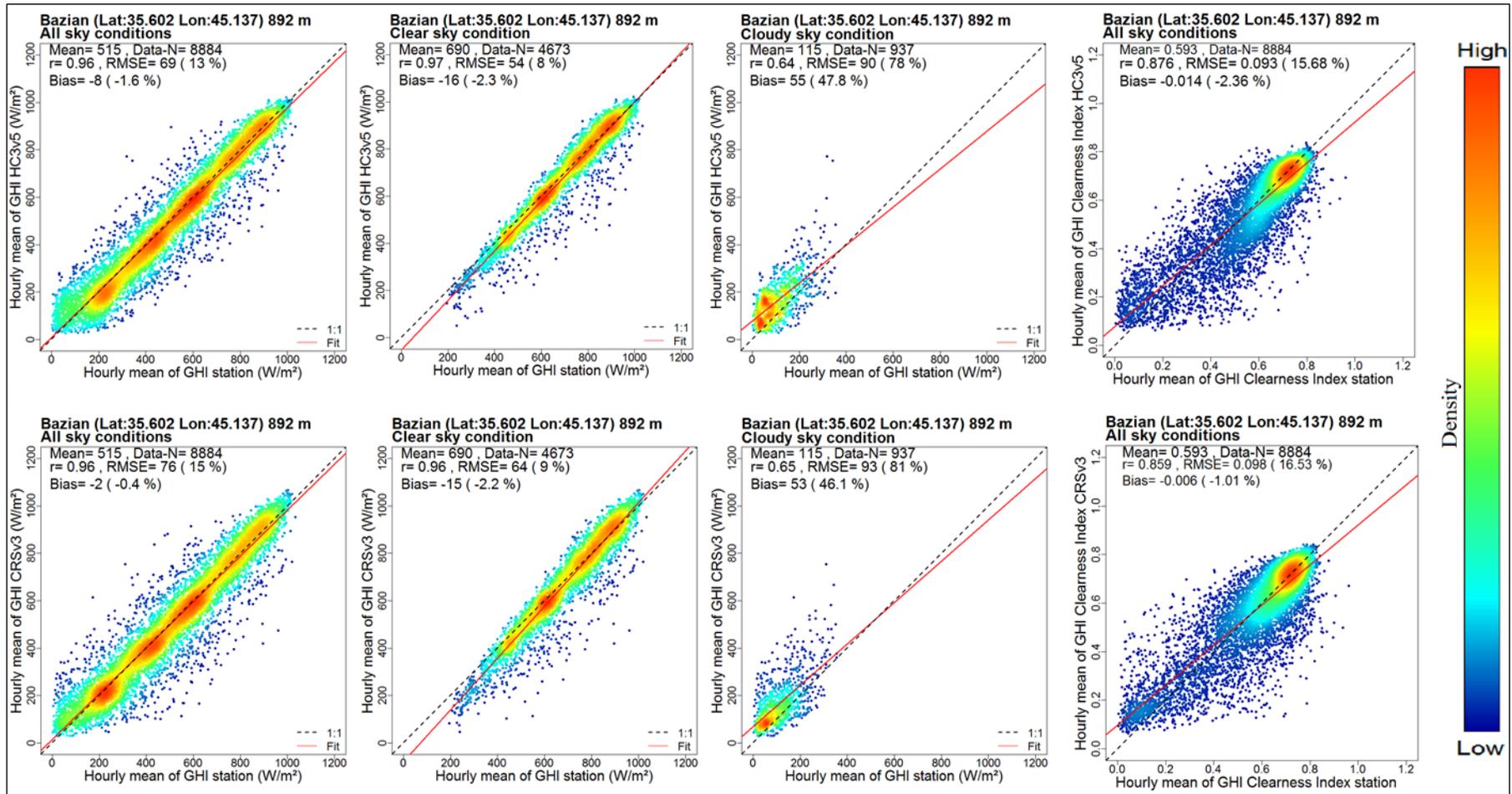
**Figure B6:** Example scatter plots between hourly GHI ground measurements and SDDs (HC3v5 top and CRSv3 bottom) for Bazian station for all-sky left, clear-sky mid and cloudy-sky right conditions. Also shows in the clearness index (Kt) for right.

**Table B1:** R codes for Appling the simple method of bias correction for SDD under cloudy sky conditions.

```r
# Simple method from (polo, 2015) applied for bias correction under cloudy-sky conditions.
# Divide dataset to use one year data for bias correction
Station2011 <- Station[which(Station$Year==2011),]
Station2012.2013 <- Station[which(Station$Year > 2011),]
# estimate HC3v5 from GHI ground data in one year to calculate coefficients
Station2011m <- lm(Station2011$HC3v5 ~ Station2011$GHIGD)
coefStation2011m <- coef(Station2011m)
interceptStation2011m <- round(coefStation2011m[1],3)
slopStation2011m <- round(coefStation2011m[2],3)
# Estimate new SDD of HC3v5 as ynew from previous steps and coefficients
Station2011$ynew <- Station2011$HC3v5-((slopStation2011m-1)*Station2011$GHIGD+interceptStation2011m)
# calculate the coefficients from ynew and HC3v5
Station2011m1 <- lm(Station2011$ynew ~ Station2011$HC3v5)
coefStation2011m1 <- coef(Station2011m1)
interceptStation2011m1 <- round(coefStation2011m1[1],3)
slopStation2011m1 <- round(coefStation2011m1[2],3)
# Use the coefficients from pervious steps to calculate Corected-HC3v5 in the new dataset.
Station2012.2013$Corected-HC3v5 <- Station2012.2013$HC3v5*slopStation2011m1+interceptStation2011m1
```

**Table B2:** Coefficients of regression models for bias correction under cloudy-sky conditions of both SDDs and at each station.

| Stations | HC3v5 | CRSv3 |
|---|---|---|
| Batufa | Corrected-HC3v5 = HC3v5* 1.062−97.406 | Corrected-CRSv3 = CRSv3*1.057−58.668 |
| Enjaksor | Corrected-HC3v5 = HC3v5*1.126−72.964 | Corrected-CRSv3 = CRSv3*1.076−44.475 |
| Hojava | Corrected-HC3v5 = HC3v5*1.021−84.253 | Corrected-CRSv3 = CRSv3*1.047−59.811 |
| Jazhnikan | Corrected-HC3v5 = HC3v5*1.163−61.212 | Corrected-CRSv3 = CRSv3*0.988 −33.993 |
| Tarjan | Corrected-HC3v5 = HC3v5*1.145−54.289 | Corrected-CRSv3 = CRSv3*0.963 −43.837 |
| Halsho | Corrected-HC3v5 = HC3v5*1.036−75.792 | Corrected-CRSv3 = CRSv3*1.006−94.783 |
| Bazian | Corrected-HC3v5 = HC3v5*1.064−69.250 | Corrected-CRSv3 = CRSv3*1.0877−64.905 |
| Maydan | Corrected-HC3v5 = HC3v5*1.123−44.535 | Corrected-CRSv3 = CRSv3*1.080 −42.400 |
| Kalar | Corrected-HC3v5 = HC3v5*1.156−42.402 | Corrected-CRSv3 = CRSv3*0.995 −35.299 |

# Appendix C – Chapter 6 Additional Figures and Tables

**Table C1:** Statistical results of hourly GHI ANN models and neuron numbers in the hidden layer for **Batufa** tower station. Mean, Bias and RMSE units are W/m$^2$.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|--------|-------|------|-------|------|----|-------|------------|-------|-------|------|----|-------|------|-------|-------|------|----|-------|---------|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 523 | 0.03 | 0.01 | 211 | 40 | 0.654 | 514 | 7.11 | 1.38 | 212 | 41 | 0.639 | 512 | 9.87 | 1.93 | 215 | 42 | 0.639 | 70 |
| M2 | 523 | 3.34 | 0.64 | 110 | 21 | 0.919 | 511 | 7.61 | 1.49 | 112 | 22 | 0.914 | 517 | 4.54 | 0.88 | 110 | 21 | 0.92 | 80 |
| M3 | 519 | 2.68 | 0.52 | 115 | 22 | 0.912 | 528 | -0.42 | −0.08 | 117 | 22 | 0.906 | 518 | -0.18 | -0.03 | 120 | 23 | 0.904 | 140 |
| M4 | 517 | 0.93 | 0.18 | 66 | 13 | 0.971 | 534 | 0.86 | 0.16 | 65 | 12 | 0.973 | 522 | 0.3 | 0.06 | 66 | 13 | 0.972 | 70 |
| M5 | 518 | −0.6 | −0.12 | 83 | 16 | 0.955 | 523 | 1.11 | 0.21 | 79 | 15 | 0.959 | 528 | 0.18 | 0.03 | 82 | 16 | 0.956 | 70 |
| M6 | 521 | 1.32 | 0.25 | 100 | 19 | 0.932 | 506 | 1.41 | 0.28 | 108 | 21 | 0.923 | 533 | 0.83 | 0.16 | 101 | 19 | 0.935 | 90 |
| M7 | 522 | −5.23 | −1 | 64 | 12 | 0.974 | 514 | −2.29 | −0.45 | 66 | 13 | 0.971 | 519 | −4.86 | −0.94 | 64 | 12 | 0.973 | 70 |
| M8 | 521 | 0.53 | 0.1 | 79 | 15 | 0.959 | 525 | −1.32 | −0.25 | 80 | 15 | 0.957 | 513 | 0.63 | 0.12 | 81 | 16 | 0.957 | 70 |
| M9 | 519 | 2.51 | 0.48 | 64 | 12 | 0.974 | 526 | 1.79 | 0.34 | 64 | 12 | 0.974 | 522 | 1.5 | 0.29 | 65 | 12 | 0.973 | 50 |
| M10 | 522 | −0.55 | −0.11 | 61 | 12 | 0.976 | 513 | 0.98 | 0.19 | 63 | 12 | 0.975 | 521 | 0.07 | 0.01 | 61 | 12 | 0.976 | 90 |

**Table C2:** As in Table C1, but for **Enjaksor** tower station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|--------|------|------|------|------|-----|------|------|------|------|------|-----|------|------|------|------|------|-----|------|---------|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 518 | −1.36 | −0.26 | 207 | 40 | 0.608 | 521 | −3.85 | −0.74 | 208 | 40 | 0.6 | 515 | −2.67 | −0.52 | 209 | 41 | 0.596 | 90 |
| M2 | 518 | −2.02 | −0.39 | 103 | 20 | 0.919 | 520 | −2.97 | −0.57 | 105 | 20 | 0.916 | 512 | −0.57 | −0.11 | 104 | 20 | 0.917 | 120 |
| M3 | 514 | 2.27 | 0.44 | 112 | 22 | 0.903 | 530 | 0.42 | 0.08 | 113 | 21 | 0.901 | 524 | −1.84 | −0.35 | 113 | 22 | 0.898 | 100 |
| M4 | 514 | 0.21 | 0.04 | 57 | 11 | 0.976 | 529 | −2.32 | −0.44 | 57 | 11 | 0.975 | 524 | 0.04 | 0.01 | 57 | 11 | 0.976 | 60 |
| M5 | 517 | 3.93 | 0.76 | 75 | 15 | 0.958 | 521 | 2.45 | 0.47 | 75 | 14 | 0.958 | 517 | 4.09 | 0.79 | 75 | 15 | 0.959 | 90 |
| M6 | 518 | −0.83 | −0.16 | 98 | 19 | 0.926 | 519 | −2.64 | −0.51 | 102 | 20 | 0.92 | 517 | 4.12 | 0.8 | 100 | 19 | 0.925 | 70 |
| M7 | 519 | 0.24 | 0.05 | 54 | 10 | 0.979 | 512 | −1.03 | −0.2 | 56 | 11 | 0.977 | 519 | −2.45 | −0.47 | 56 | 11 | 0.977 | 60 |
| M8 | 517 | 5.47 | 1.06 | 73 | 14 | 0.96 | 521 | 4.34 | 0.83 | 73 | 14 | 0.961 | 519 | 4.49 | 0.87 | 72 | 14 | 0.961 | 50 |
| M9 | 516 | −0.45 | −0.09 | 54 | 10 | 0.978 | 525 | 0.56 | 0.11 | 55 | 10 | 0.977 | 520 | 0.49 | 0.09 | 54 | 10 | 0.979 | 60 |
| M10 | 520 | 0.73 | 0.14 | 52 | 10 | 0.98 | 513 | 0.63 | 0.12 | 51 | 10 | 0.98 | 511 | −0.14 | -0.03 | 52 | 10 | 0.98 | 100 |

**Table C3:** As in Table C1, but for **Hojava** tower station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|--------|------|------|------|------|-----|------|------|------|------|------|-----|------|------|------|------|------|-----|------|---------|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 504 | −1.63 | -0.32 | 205 | 41 | 0.634 | 514 | −6.44 | -1.25 | 207 | 40 | 0.618 | 491 | 6.66 | 1.36 | 208 | 42 | 0.621 | 90 |
| M2 | 504 | 5.03 | 1 | 106 | 21 | 0.918 | 495 | 0.47 | 0.09 | 106 | 21 | 0.916 | 509 | −0.44 | −0.09 | 103 | 20 | 0.92 | 90 |
| M3 | 504 | 1.82 | 0.36 | 112 | 22 | 0.908 | 503 | −0.78 | -0.16 | 111 | 22 | 0.906 | 501 | 1.25 | 0.25 | 110 | 22 | 0.908 | 90 |
| M4 | 503 | 0.45 | 0.09 | 64 | 13 | 0.97 | 501 | 0.11 | 0.02 | 64 | 13 | 0.97 | 509 | 0.3 | 0.06 | 65 | 13 | 0.969 | 80 |
| M5 | 504 | −2 | -0.4 | 79 | 16 | 0.955 | 498 | 2.35 | 0.47 | 83 | 17 | 0.947 | 504 | −1.46 | −0.29 | 79 | 16 | 0.955 | 80 |
| M6 | 501 | −4.56 | -0.91 | 98 | 20 | 0.93 | 512 | −0.81 | -0.16 | 100 | 20 | 0.925 | 504 | −3.23 | −0.64 | 99 | 20 | 0.927 | 100 |
| M7 | 501 | −3.24 | -0.65 | 64 | 13 | 0.97 | 507 | −2.2 | -0.43 | 62 | 12 | 0.972 | 511 | −2.67 | −0.52 | 64 | 13 | 0.97 | 80 |
| M8 | 504 | −4.46 | -0.88 | 75 | 15 | 0.959 | 504 | −3.53 | -0.7 | 78 | 15 | 0.957 | 500 | -6.12 | -1.22 | 78 | 16 | 0.957 | 80 |
| M9 | 504 | 0.32 | 0.06 | 64 | 13 | 0.971 | 503 | −0.83 | -0.17 | 64 | 13 | 0.97 | 499 | 2.7 | 0.54 | 63 | 13 | 0.971 | 80 |
| M10 | 502 | 4.01 | 0.8 | 57 | 11 | 0.976 | 514 | 4.16 | 0.81 | 63 | 12 | 0.972 | 499 | 2.56 | 0.51 | 61 | 12 | 0.974 | 80 |

**Table C4:** As in Table C1, but for **Jazhnikan** tower station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 518 | –1.85 | –0.36 | 208 | 40 | 0.59 | 518 | –0.13 | –0.03 | 208 | 40 | 0.584 | 519 | –6.91 | –1.33 | 209 | 40 | 0.586 | 100 |
| M2 | 516 | 8.12 | 1.57 | 99 | 19 | 0.924 | 516 | 10.19 | 1.97 | 102 | 20 | 0.918 | 530 | 3.36 | 0.63 | 102 | 19 | 0.917 | 110 |
| M3 | 518 | 0.99 | 0.19 | 108 | 21 | 0.907 | 512 | 0.03 | 0.01 | 111 | 22 | 0.906 | 526 | –1.78 | –0.34 | 109 | 21 | 0.903 | 110 |
| M4 | 521 | 3.41 | 0.65 | 62 | 12 | 0.97 | 509 | 3.59 | 0.71 | 63 | 12 | 0.97 | 516 | 3.67 | 0.71 | 66 | 13 | 0.968 | 120 |
| M5 | 518 | –1.5 | –0.29 | 72 | 14 | 0.96 | 521 | –3.16 | –0.61 | 75 | 14 | 0.958 | 514 | 2.21 | 0.43 | 71 | 14 | 0.96 | 100 |
| M6 | 518 | 0.63 | 0.12 | 89 | 17 | 0.938 | 514 | –1.4 | –0.27 | 91 | 18 | 0.934 | 522 | 0.39 | 0.07 | 96 | 18 | 0.927 | 130 |
| M7 | 514 | –0.4 | –0.08 | 58 | 11 | 0.974 | 524 | –1.07 | –0.2 | 60 | 11 | 0.973 | 530 | 0.01 | 0 | 60 | 11 | 0.972 | 120 |
| M8 | 517 | –2.8 | –0.54 | 70 | 14 | 0.963 | 524 | –3.44 | –0.66 | 69 | 13 | 0.964 | 516 | –0.39 | –0.08 | 75 | 15 | 0.955 | 140 |
| M9 | 516 | –0.58 | –0.11 | 56 | 11 | 0.976 | 525 | –1.27 | –0.24 | 57 | 11 | 0.976 | 523 | –3.2 | –0.61 | 59 | 11 | 0.974 | 140 |
| M10 | 520 | 3.27 | 0.63 | 53 | 10 | 0.979 | 503 | 3.45 | 0.69 | 59 | 12 | 0.972 | 523 | 0.65 | 0.12 | 59 | 11 | 0.973 | 130 |

**Table C5:** As in Table C1, but for **Tarjan** tower station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 523 | 3.01 | 0.58 | 202 | 39 | 0.598 | 514 | 4.39 | 0.85 | 207 | 40 | 0.557 | 517 | 11.55 | 2.23 | 207 | 40 | 0.552 | 100 |
| M2 | 520 | 3.83 | 0.74 | 99 | 19 | 0.917 | 526 | 4.76 | 0.9 | 102 | 19 | 0.917 | 518 | 5.33 | 1.03 | 102 | 20 | 0.912 | 80 |
| M3 | 519 | 0.89 | 0.17 | 111 | 21 | 0.895 | 521 | –1.26 | –0.24 | 112 | 21 | 0.894 | 528 | –6.07 | –1.15 | 106 | 20 | 0.907 | 70 |
| M4 | 524 | 4.08 | 0.78 | 61 | 12 | 0.97 | 519 | 2.51 | 0.48 | 64 | 12 | 0.969 | 510 | 3.79 | 0.74 | 62 | 12 | 0.967 | 100 |
| M5 | 521 | 1.2 | 0.23 | 71 | 14 | 0.959 | 519 | 2.33 | 0.45 | 71 | 14 | 0.959 | 523 | 2.82 | 0.54 | 71 | 14 | 0.959 | 20 |
| M6 | 522 | 0.49 | 0.09 | 90 | 17 | 0.932 | 519 | –0.66 | –0.13 | 91 | 18 | 0.935 | 518 | 1.22 | 0.24 | 93 | 18 | 0.929 | 80 |
| M7 | 522 | 1.01 | 0.19 | 59 | 11 | 0.972 | 508 | 0.95 | 0.19 | 58 | 11 | 0.972 | 527 | 0.18 | 0.03 | 59 | 11 | 0.972 | 60 |
| M8 | 520 | –2.17 | –0.42 | 68 | 13 | 0.962 | 522 | –2.81 | –0.54 | 71 | 14 | 0.96 | 521 | 0.41 | 0.08 | 71 | 14 | 0.96 | 60 |
| M9 | 523 | 0.1 | 0.02 | 58 | 11 | 0.973 | 514 | 1.69 | 0.33 | 58 | 11 | 0.974 | 517 | –0.04 | –0.01 | 60 | 12 | 0.971 | 40 |
| M10 | 522 | –1.16 | –0.22 | 56 | 11 | 0.975 | 520 | –0.32 | –0.06 | 55 | 11 | 0.975 | 516 | –4.35 | –0.84 | 57 | 11 | 0.973 | 60 |

**Table C6:** As in Table C1, but for **Halsho** automatic station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|--------|-------|------|-------|------|-----|-------|-------|------|------|------|-----|-------|-------|------|-------|------|-----|-------|---------|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 482 | 1.13 | 0.23 | 153 | 32 | 0.835 | 468 | 2.39 | 0.51 | 153 | 33 | 0.834 | 474 | 8.11 | 1.71 | 155 | 33 | 0.827 | 80 |
| M2 | 479 | –0.76 | –0.16 | 57 | 12 | 0.979 | 487 | –1.08 | –0.22 | 59 | 12 | 0.977 | 472 | 1.11 | 0.24 | 61 | 13 | 0.976 | 120 |
| M3 | 478 | –2.87 | –0.6 | 62 | 13 | 0.974 | 475 | 0.32 | 0.07 | 67 | 14 | 0.97 | 485 | –3.28 | –0.68 | 66 | 14 | 0.972 | 120 |
| M4 | 481 | –0.11 | –0.02 | 55 | 11 | 0.98 | 476 | 2.36 | 0.5 | 55 | 12 | 0.98 | 474 | –0.49 | –0.1 | 56 | 12 | 0.98 | 60 |
| M5 | 477 | 0.64 | 0.13 | 60 | 13 | 0.976 | 481 | 1.44 | 0.3 | 62 | 13 | 0.974 | 484 | 0.31 | 0.06 | 62 | 13 | 0.975 | 100 |
| M6 | 479 | –0.52 | –0.11 | 58 | 12 | 0.978 | 482 | –0.14 | –0.03 | 57 | 12 | 0.978 | 477 | 1.6 | 0.34 | 60 | 13 | 0.976 | 40 |
| M7 | 478 | 1.45 | 0.3 | 49 | 10 | 0.984 | 479 | 1.37 | 0.29 | 50 | 10 | 0.984 | 481 | 1.81 | 0.38 | 49 | 10 | 0.984 | 20 |
| M8 | 478 | –0.37 | –0.08 | 51 | 11 | 0.983 | 481 | 0.95 | 0.2 | 54 | 11 | 0.981 | 482 | 0.65 | 0.13 | 51 | 11 | 0.983 | 30 |
| M9 | 477 | –0.14 | –0.03 | 48 | 10 | 0.985 | 483 | 1.54 | 0.32 | 48 | 10 | 0.985 | 484 | –0.1 | –0.02 | 49 | 10 | 0.984 | 50 |
| M10 | 476 | 0.4 | 0.08 | 47 | 10 | 0.985 | 489 | 0 | 0 | 46 | 9 | 0.986 | 481 | 0.73 | 0.15 | 49 | 10 | 0.985 | 40 |

**Table C7:** As in Table C1, but for **Bazian** automatic station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|--------|-------|------|-------|------|-----|-------|-------|------|------|------|-----|-------|-------|------|-------|------|-----|-------|---------|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 482 | –0.46 | –0.1 | 166 | 34 | 0.724 | 477 | –3.28 | –0.69 | 168 | 35 | 0.712 | 468 | 5.49 | 1.17 | 168 | 36 | 0.716 | 70 |
| M2 | 480 | 3.41 | 0.71 | 62 | 13 | 0.966 | 482 | 0.25 | 0.05 | 63 | 13 | 0.965 | 473 | 0.78 | 0.16 | 62 | 13 | 0.967 | 60 |
| M3 | 479 | –0.11 | –0.02 | 60 | 13 | 0.968 | 475 | –0.63 | –0.13 | 63 | 13 | 0.965 | 484 | –1.32 | –0.27 | 61 | 13 | 0.967 | 70 |
| M4 | 477 | 0.94 | 0.2 | 55 | 12 | 0.973 | 490 | 4.31 | 0.88 | 55 | 11 | 0.973 | 477 | 1.11 | 0.23 | 56 | 12 | 0.973 | 50 |
| M5 | 480 | 2.78 | 0.58 | 61 | 13 | 0.967 | 471 | 4.94 | 1.05 | 64 | 14 | 0.963 | 483 | 2.97 | 0.61 | 61 | 13 | 0.969 | 60 |
| M6 | 478 | –0.31 | –0.06 | 57 | 12 | 0.971 | 484 | –0.26 | –0.05 | 61 | 13 | 0.967 | 476 | 0.88 | 0.18 | 58 | 12 | 0.971 | 70 |
| M7 | 478 | 2.47 | 0.52 | 50 | 10 | 0.979 | 486 | –0.62 | –0.13 | 50 | 10 | 0.978 | 476 | –2.14 | –0.45 | 50 | 11 | 0.978 | 20 |
| M8 | 480 | –0.04 | –0.01 | 52 | 11 | 0.976 | 476 | 2.2 | 0.46 | 53 | 11 | 0.976 | 476 | 1.18 | 0.25 | 53 | 11 | 0.975 | 20 |
| M9 | 478 | 2.6 | 0.54 | 46 | 10 | 0.982 | 484 | 0.74 | 0.15 | 46 | 10 | 0.982 | 477 | 2.93 | 0.61 | 48 | 10 | 0.979 | 40 |
| M10 | 480 | 0.37 | 0.08 | 46 | 10 | 0.981 | 480 | 1.82 | 0.38 | 47 | 10 | 0.98 | 472 | –0.31 | –0.07 | 49 | 10 | 0.978 | 50 |

**Table C 8**: As in Table C1, but for **Maydan** automatic station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|--------|-------|------|-------|------|-----|-------|-------|------|-------|------|-----|-------|-------|-------|-------|------|-----|-------|---------|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 503 | 0.06 | 0.01 | 165 | 33 | 0.723 | 501 | 13.04 | 2.6 | 170 | 34 | 0.701 | 507 | -0.72 | -0.14 | 169 | 33 | 0.7 | 60 |
| M2 | 503 | –0.1 | –0.02 | 43 | 9 | 0.983 | 506 | –1.09 | –0.22 | 45 | 9 | 0.982 | 499 | -0.16 | -0.03 | 46 | 9 | 0.982 | 60 |
| M3 | 506 | 0.42 | 0.08 | 45 | 9 | 0.982 | 494 | 1.64 | 0.33 | 48 | 10 | 0.98 | 498 | 0.02 | 0 | 47 | 9 | 0.981 | 50 |
| M4 | 503 | 0.17 | 0.03 | 48 | 10 | 0.98 | 509 | –1.94 | –0.38 | 50 | 10 | 0.978 | 499 | 0.19 | 0.04 | 49 | 10 | 0.978 | 60 |
| M5 | 504 | –2.29 | –0.45 | 54 | 11 | 0.974 | 499 | –2.64 | –0.53 | 59 | 12 | 0.969 | 505 | -2.93 | -0.58 | 57 | 11 | 0.971 | 70 |
| M6 | 500 | –1.41 | –0.28 | 42 | 8 | 0.984 | 514 | –1.11 | –0.22 | 43 | 8 | 0.983 | 506 | -3.86 | -0.76 | 44 | 9 | 0.982 | 80 |
| M7 | 504 | 0.96 | 0.19 | 37 | 7 | 0.988 | 499 | 1.26 | 0.25 | 39 | 8 | 0.986 | 503 | 2.02 | 0.4 | 40 | 8 | 0.986 | 60 |
| M8 | 503 | 0.16 | 0.03 | 40 | 8 | 0.986 | 503 | –0.32 | –0.06 | 42 | 8 | 0.984 | 505 | 0.28 | 0.06 | 43 | 9 | 0.983 | 60 |
| M9 | 505 | 0.6 | 0.12 | 35 | 7 | 0.989 | 488 | –1.37 | –0.28 | 35 | 7 | 0.989 | 511 | -2.43 | -0.48 | 39 | 8 | 0.987 | 50 |
| M10 | 504 | 0.33 | 0.07 | 38 | 8 | 0.987 | 502 | 0.18 | 0.04 | 40 | 8 | 0.986 | 500 | -0.46 | -0.09 | 41 | 8 | 0.985 | 70 |

**Table C9:** As in Table C1, but for **Kalar** automatic station.

| Models | Train | | | | | | Validation | | | | | | Test | | | | | | Neurons |
|--------|-------|------|-------|------|-----|-------|-------|------|-------|------|-----|-------|-------|-------|-------|------|-----|-------|---------|
| | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | Mean | Bias | % | RMSE | % | r | |
| M1 | 473 | –1.63 | –0.34 | 158 | 33 | 0.689 | 479 | –7.26 | -1.52 | 157 | 33 | 0.695 | 479 | –9.24 | –1.93 | 161 | 34 | 0.705 | 70 |
| M2 | 475 | –0.91 | –0.19 | 55 | 12 | 0.968 | 471 | 1.17 | 0.25 | 56 | 12 | 0.966 | 475 | 0.84 | 0.18 | 56 | 12 | 0.966 | 60 |
| M3 | 475 | 0.03 | 0.01 | 55 | 12 | 0.968 | 473 | –6.31 | –1.33 | 58 | 12 | 0.965 | 474 | –3.22 | –0.68 | 58 | 12 | 0.962 | 50 |
| M4 | 475 | –1.14 | –0.24 | 57 | 12 | 0.966 | 477 | –1.44 | –0.3 | 58 | 12 | 0.964 | 473 | –2.85 | –0.6 | 57 | 12 | 0.964 | 60 |
| M5 | 478 | 4.84 | 1.01 | 60 | 13 | 0.962 | 471 | 4.31 | 0.92 | 62 | 13 | 0.96 | 464 | 6.89 | 1.48 | 60 | 13 | 0.961 | 90 |
| M6 | 473 | –0.48 | –0.1 | 54 | 11 | 0.97 | 485 | –3.77 | –0.78 | 55 | 11 | 0.969 | 471 | –0.73 | –0.15 | 52 | 11 | 0.971 | 80 |
| M7 | 474 | –0.02 | 0 | 48 | 10 | 0.976 | 482 | –0.47 | –0.1 | 50 | 10 | 0.975 | 472 | 2.12 | 0.45 | 51 | 11 | 0.972 | 60 |
| M8 | 475 | 0.77 | 0.16 | 51 | 11 | 0.972 | 467 | –3.03 | –0.65 | 51 | 11 | 0.972 | 479 | –1.02 | –0.21 | 51 | 11 | 0.972 | 40 |
| M9 | 476 | –0.21 | –0.04 | 45 | 9 | 0.979 | 465 | –0.77 | –0.17 | 48 | 10 | 0.975 | 477 | –0.66 | –0.14 | 48 | 10 | 0.977 | 40 |
| M10 | 478 | 0.03 | 0.01 | 47 | 10 | 0.977 | 462 | 0.34 | 0.07 | 48 | 10 | 0.976 | 474 | 1.05 | 0.22 | 49 | 10 | 0.973 | 60 |
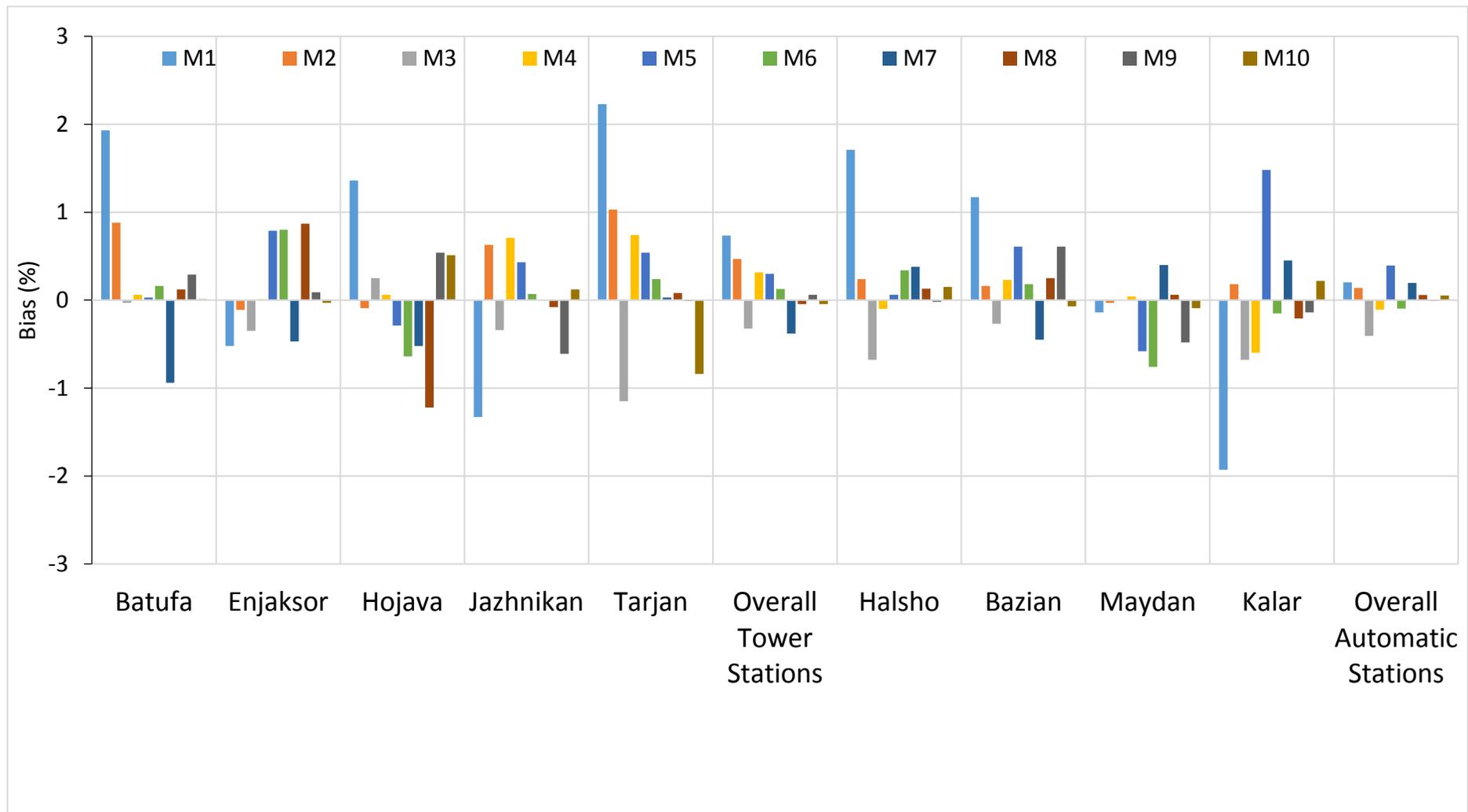
**Figure C1:** Comparison of bias (%) for the hourly GHI among ANN models, stations, and overall results of station types.
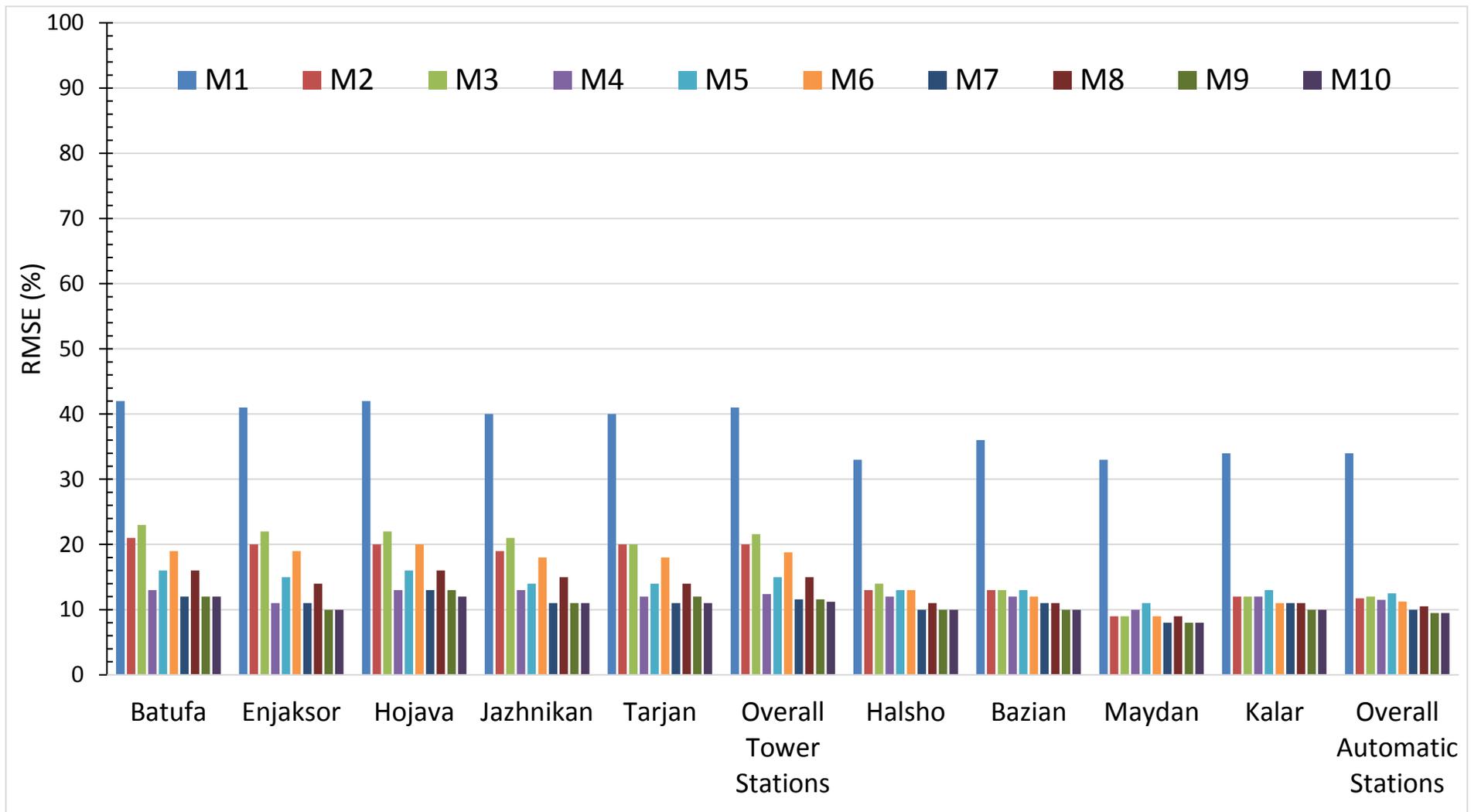
**Figure C2:** As in Figure C1, but for RMSE (%).
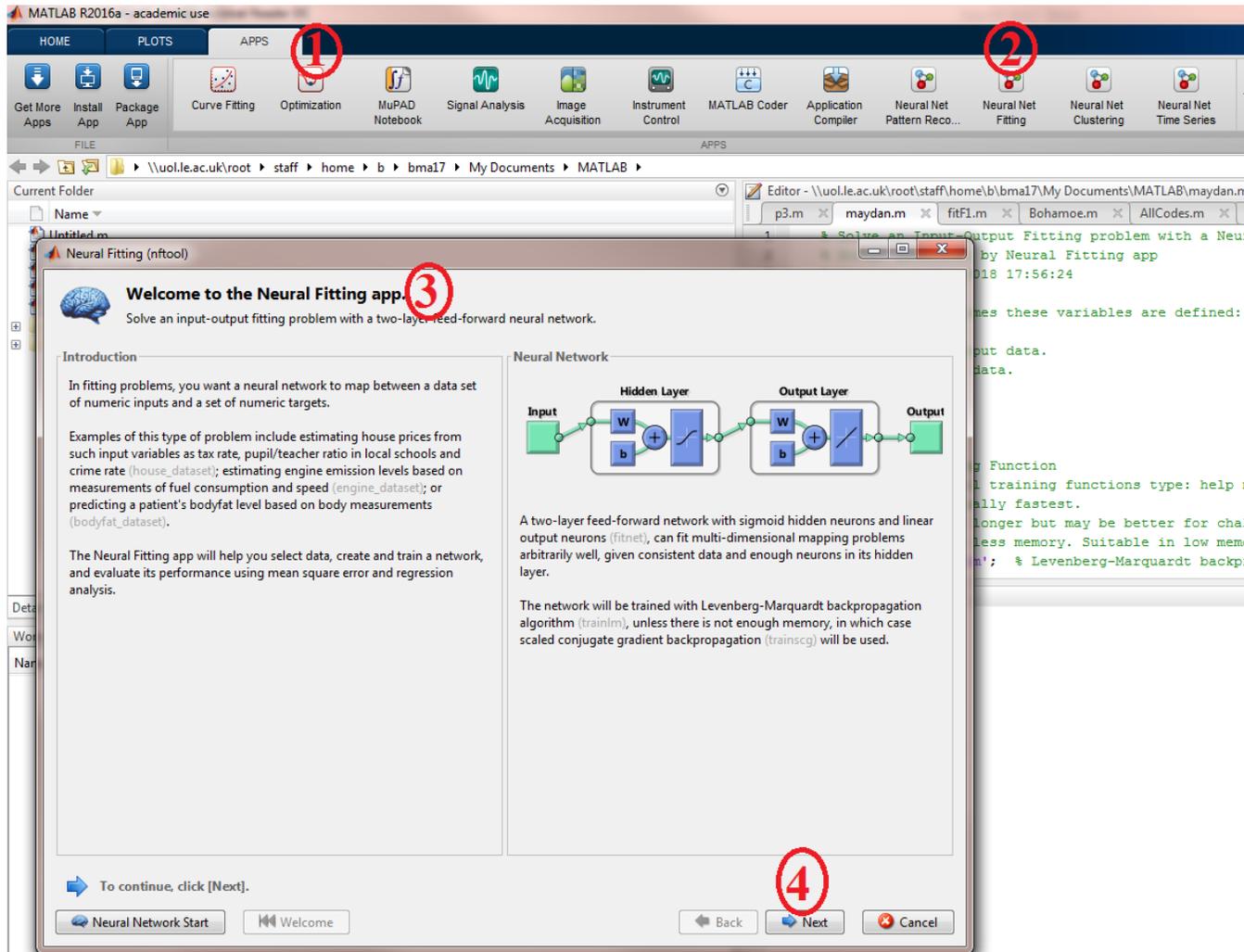
**Figure C3:** Using ANN APP in MATLAB software to run ANN models as number 1 in red, Selecting Neural Network fitting tool as number 2–4 in red to next step.
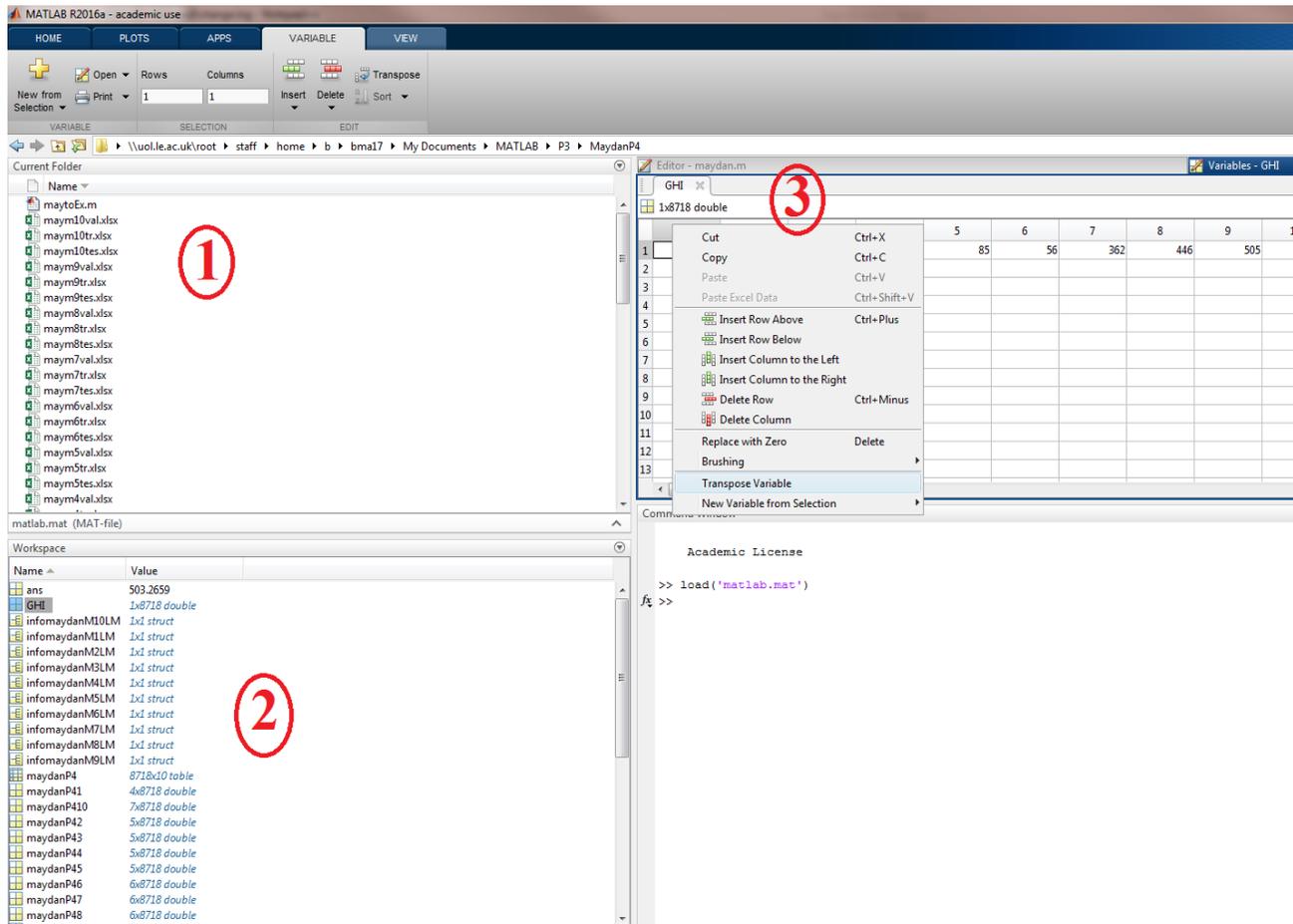
**Figure C4:** The first step in the Neural Network fitting tool is import data to MATLAB number 1 in red, second mange data frames and variables in MATLAB-Workspace as target variable and several sets of input variables, number 2 in red. Then, open data frames and table of variables to change (Transpose Variable) the variables from vertical to horizontal as it is required in MATLAB number 3 in red.
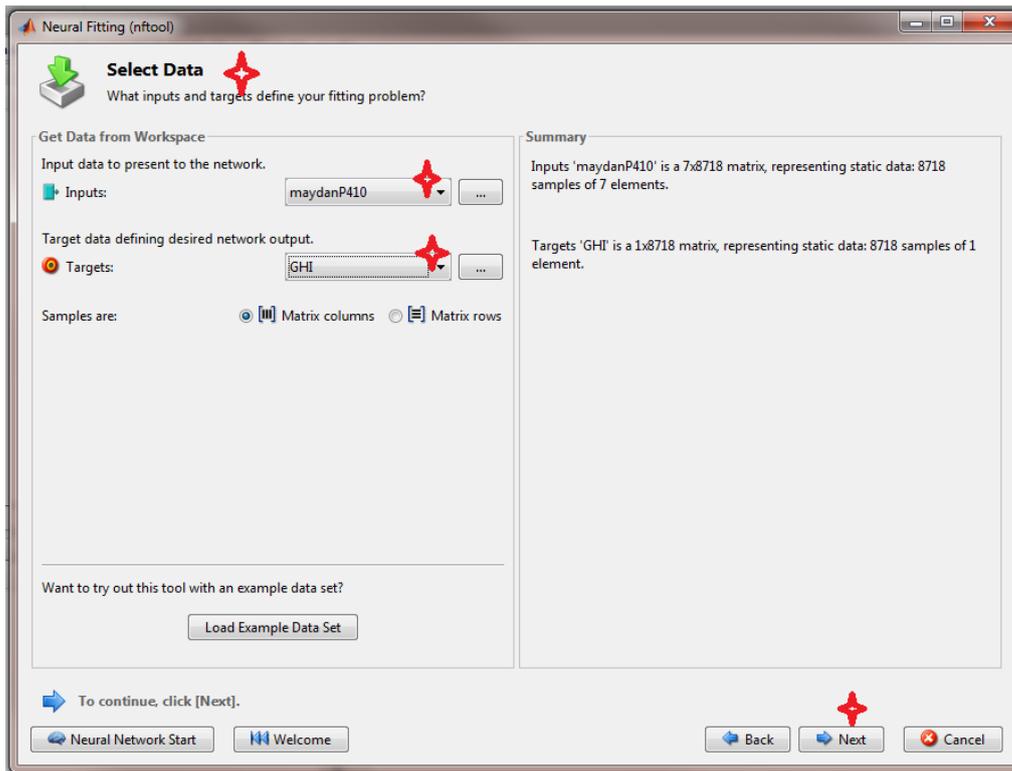
**Figure C5:** Select input and target data in step three in red symbols.
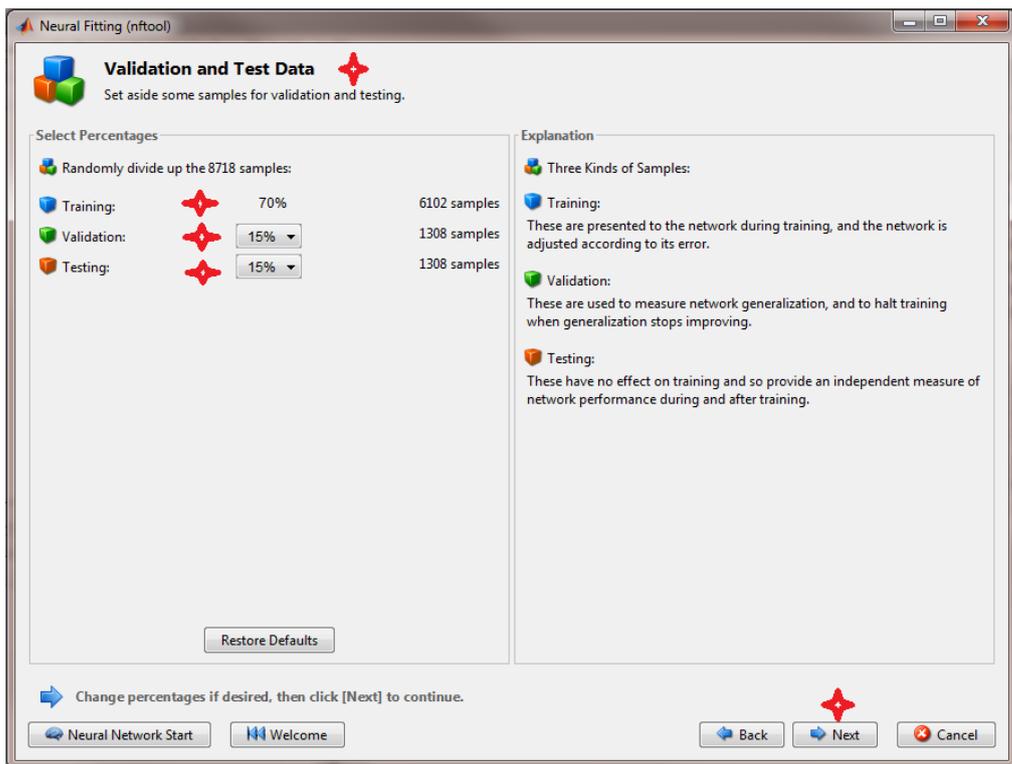


**Figure C6:** Determining the percentage of data for training, validation and testing in step four in red symbols.
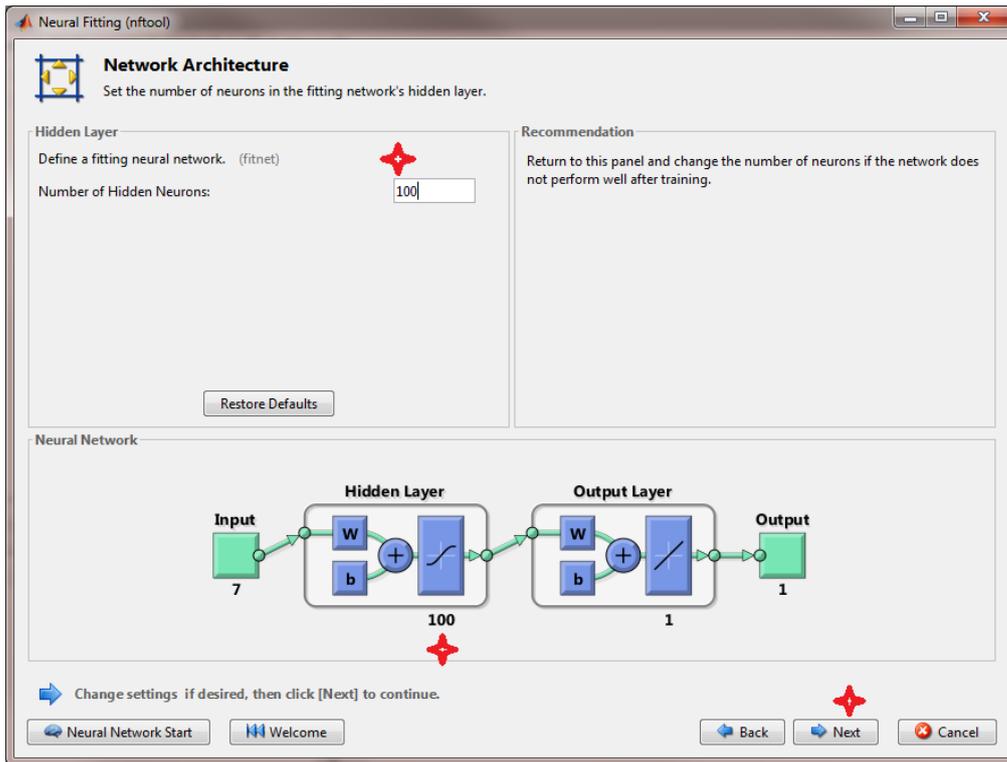
**Figure C7:** Determining the number of hidden node in the hidden layer in step four in red symbols.
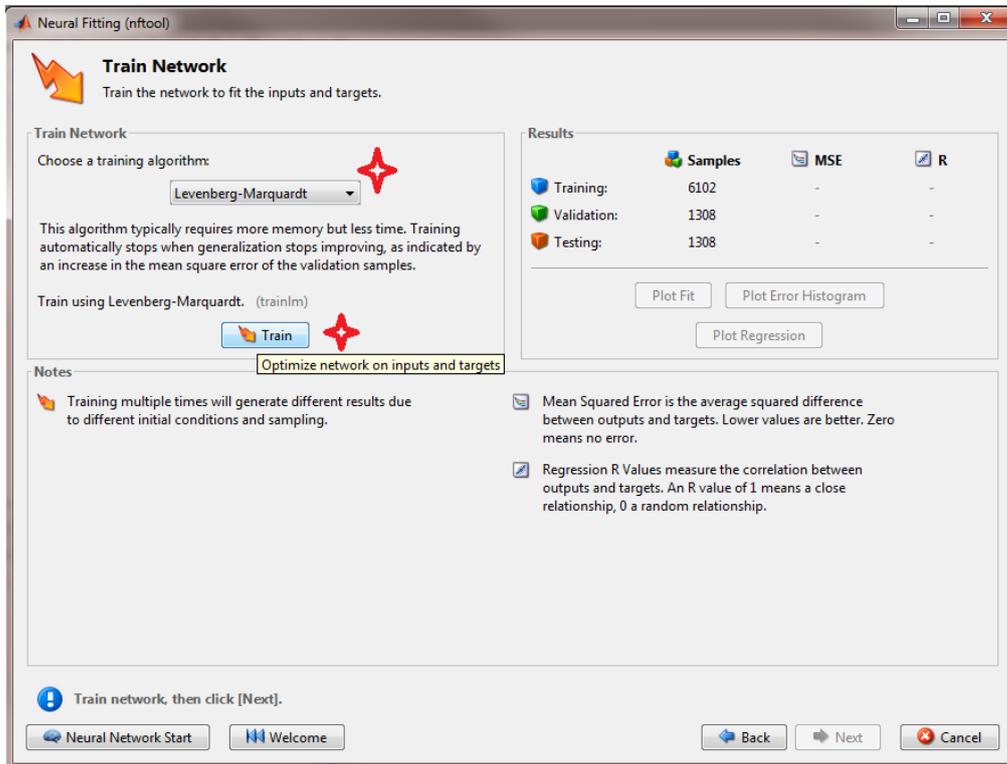


**Figure C8:** Choosing the training algorithm and train the model in step five in red symbols.
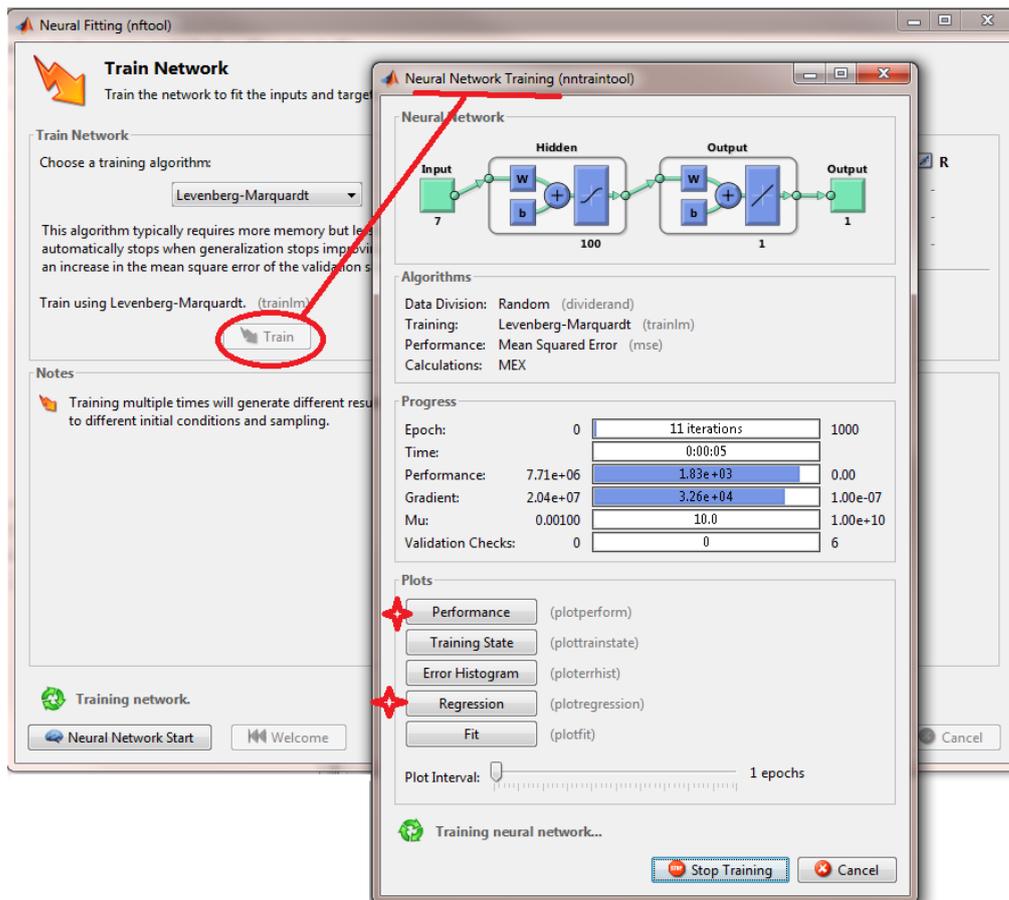
**Figure C9:** Checking the performance of the training model based on errors and regression plots to select a model or to retrain it, in step six red symbols.
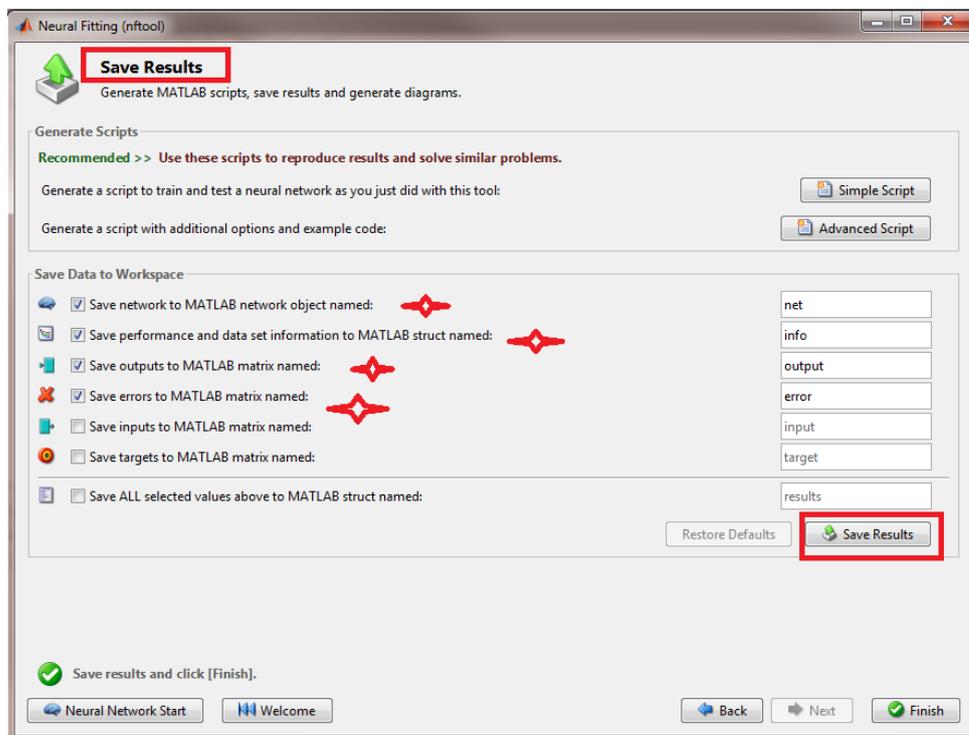


**Figure C10:** Save the results, errors and trained model in the final step.

**Table C10:** MATLAB codes for appling an ANN model as the same steps in Figures (C3–C9).

```matlab
% Solve an Input-Output Fitting problem with a Neural Network
% Script generated by Neural Fitting app
% Created 19-Aug-2018 17:56:24
%
% This script assumes these variables are defined:
%   maydanP31 - input data.
%   GHI1 - target data.
x = Climate.variables;
t = GHI;
% Choose a Training Function
% For a list of all training functions type: help nntrain
% 'trainlm' is usually fastest.
% 'trainbr' takes longer but may be better for challenging problems.
% 'trainscg' uses less memory. Suitable in low memory situations.
trainFcn = 'trainlm';  % Levenberg-Marquardt backpropagation.
% Create a Fitting Network
hiddenLayerSize = 100;
net = fitnet(hiddenLayerSize,trainFcn);
% Choose Input and Output Pre/Post-Processing Functions
% For a list of all processing functions type: help nnprocess
net.input.processFcns = {'removeconstantrows','mapminmax'};

net.output.processFcns = {'removeconstantrows','mapminmax'};

% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'sample';  % Divide up every sample
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
% Choose a Performance Function
% For a list of all performance functions type: help nnperformance
net.performFcn = 'mse';  % Mean Squared Error
% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...
    'plotregression', 'plotfit'};
% Train the Network
[net,tr] = train(net,x,t);
% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y)
% Recalculate Training, Validation and Test Performance
trainTargets = t .* tr.trainMask;
valTargets = t .* tr.valMask;
testTargets = t .* tr.testMask;
trainPerformance = perform(net,trainTargets,y)
valPerformance = perform(net,valTargets,y)
testPerformance = perform(net,testTargets,y)
% View the Network
view(net)
% Plots
% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, ploterrhist(e)
%figure, plotregression(t,y)
%figure, plotfit(net,x,t)
```

```matlab
% Deployment
% Change the (false) values to (true) to enable the following code blocks.
% See the help for each generation function for more information.
if (false)
  % Generate MATLAB function for neural network for application
  % deployment in MATLAB scripts or with MATLAB Compiler and Builder
  % tools, or simply to examine the calculations your trained neural
  % network performs.
  genFunction(net,'myNeuralNetworkFunction');
  y = myNeuralNetworkFunction(x);
end
if (false)
  % Generate a matrix-only MATLAB function for neural network code
  % generation with MATLAB Coder tools.
  genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
  y = myNeuralNetworkFunction(x);
end
if (false)
  % Generate a Simulink diagram for simulation or deployment with.
  % Simulink Coder tools.
  gensim(net);
end
```

# References

Aguiar, L. M., Pereira, B., David, M., Díaz, F. & Lauret, P. 2015. Use of satellite data to improve solar radiation forecasting with Bayesian Artificial Neural Networks. *Solar Energy,* 122(1309-1324.

Aguiar, L. M., Pereira, B., Lauret, P., Diaz, F. & David, M. 2016. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy,* 97(599-610.

Ahmed, S. & Mahammed, H. 2012. A statistical analysis of wind power density based on the Weibull and Ralyeigh models of "Penjwen Region" Sulaimani/Iraq. *Jordan Journal of Mechanical and Industrial Engineering,* 6(2)**,** pp 135-140.

Al-Alawi, S. M. & Al-Hinai, H. A. 1998. An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. *Renewable Energy,* 14(1)**,** pp 199-204.

AL-Jumaily, K. J., AL-Salihi, A. M. & Al-Tai, O. T. 2010. Evaluation of Meteosat-8 measurements using daily global solar radiation for two stations in Iraq. *Energy and Environment,* 1(4)**,** pp 635-642.

Al-Rasheedi, M., Gueymard, C. A., Ismail, A. & Hussain, T. 2018. Comparison of two sensor technologies for solar irradiance measurement in a desert environment. *Solar Energy,* 161(194-206.

Alam, S., Kaushik, S. & Garg, S. 2009. Assessment of diffuse solar energy under general sky condition using artificial neural network. *Applied Energy,* 86(4)**,** pp 554-564.

Almorox, J. & Hontoria, C. 2004. Global solar radiation estimation using sunshine duration in Spain. *Energy Conversion and Management,* 45(9-10)**,** pp 1529-1535.

Alton, P., North, P., Kaduk, J. & Los, S. 2005. Radiative transfer modeling of direct and diffuse sunlight in a Siberian pine forest. *Journal of Geophysical Research: Atmospheres,* 110(D23)**,** pp.

Ameen, B., Balzter, H. & Jarvis, C. 2018a. Quality Control of Global Horizontal Irradiance Estimates through BSRN, TOACs and Air Temperature/Sunshine Duration Test Procedures. *Climate,* 6(3)**,** pp 69.

Ameen, B., Balzter, H., Jarvis, C., Wey, E., Thomas, C. & Marchand, M. 2018b. Validation of Hourly Global Horizontal Irradiance for Two Satellite-Derived Datasets in Northeast Iraq. *Remote Sensing,* 10(10)**,** pp 1651.

Amillo, A., Huld, T. & Müller, R. 2014. A New Database of Global and Direct Solar Radiation Using the Eastern Meteosat Satellite, Models and Validation. *Remote Sensing,* 6(9)**,** pp 8165-8189.

Amillo, A. G., Ntsangwane, L., Huld, T. & Trentmann, J. 2018. Comparison of satellite-retrieved high-resolution solar radiation datasets for South Africa. *Journal of Energy in Southern Africa,* 29(2)**,** pp 63-76.

Ampratwum, D. B. & Dorvlo, A. S. S. 1999. Estimation of solar radiation from the number of sunshine hours. *Applied Energy,* 63(161-167.

Angstrom, A. 1924. Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society,* 50(210)**,** pp 121-126.

Antonanzas-Torres, F., Urraca, R., Antonanzas, J., Fernandez-Ceniceros, J. & Martinez-de-Pison, F. J. 2015. Generation of daily global solar irradiation with support vector machines for regression. *Energy Conversion and Management,* 96(277-286.

Badescu, V. 2008. *Modeling Solar Radiation at the Earth's Surface: Recent Advances*: Springer Berlin Heidelberg.

Balzarolo, M., Anderson, K., Nichol, C., Rossini, M., Vescovo, L., Arriga, N., Wohlfahrt, G., Calvet, J.-C., Carrara, A., Cerasoli, S., Cogliati, S., Daumard, F., Eklundh, L., Elbers, J. A., Evrendilek, F., Handcock, R. N., Kaduk, J., Klumpp, K., Longdoz, B., Matteucci, G., Meroni, M., Montagnani, L., Ourcival, J.-M., Sánchez-Cañete, E. P., Pontailler, J.-Y., Juszczak, R., Scholes, B. & Martín, M. P. 2011. Ground-Based Optical Measurements at European Flux Sites: A Review of Methods, Instruments and Current Controversies. *Sensors,* 11(8)**,** pp 7954.

Behrang, M. A., Assareh, E., Ghanbarzadeh, A. & Noghrehabadi, A. R. 2010. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy,* 84(8)**,** pp 1468-1480.

Besharat, F., Dehghan, A. A. & Faghih, A. R. 2013. Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews,* 21(798-821.

Betts, T., Gottschalg, R. & Infield, D. Progress towards modelling solar spectral radiation for optimisation of amorphous silicon photovoltaic systems.  : REMIC, 2001 Belfast, May 2001, INTERNATIONAL SOLAR ENERGY SOCIETY UK SECTION-CONFERENCE-C (UK-ISES). 163-170.

Beyer, H. G., Costanzo, C. & Heinemann, D. 1996. Modifications of the Heliosat procedure for irradiance estimates from satellite images. *Solar Energy,* 56(3)**,** pp 207-212.

Bird, R. E. & Hulstrom, R. L. 1981. Simplified clear sky model for direct and diffuse insolation on horizontal surfaces,

Blanc, P., Gschwind, B., Lefèvre, M. & Wald, L. 2011a. The HelioClim project: Surface solar irradiance data for climate applications. *Remote Sensing,* 3(2)**,** pp 343-361.

Blanc, P., Gschwind, B., Lefèvre, M. & Wald, L. 2011b. The HelioClim Project: Surface Solar Irradiance Data for Climate Applications. *Remote Sensing,* 3(12)**,** pp 343-361.

Blanc, P. & Wald, L. 2016. On the effective solar zenith and azimuth angles to use with measurements of hourly irradiation. *Advances in Science and Research,* 13(1-6.

Blonquist, J. M., Tanner, B. D. & Bugbee, B. 2009. Evaluation of measurement accuracy and comparison of two new and three traditional net radiometers. *Agricultural and Forest Meteorology,* 149(10)**,** pp 1709-1721.

Bojanowski, J. S. 2014. *Quantifying solar radiation at the earth surface with meteorological and satellite data.* PhD, University of Twente.

Bojanowski, J. S., Vrieling, A. & Skidmore, A. K. 2013. Calibration of solar radiation models for Europe using Meteosat Second Generation and weather station data. *Agricultural and Forest Meteorology,* 176(1-9.

BOM. 2018. *Bureau of Meteorology, National Water Account 2015, Commonwealth of Australia* [Online]. Available: http://reg.bom.gov.au/climate/reg/oneminsolar/ [Accessed 02 February 2018].

Bouchouicha, K., Razagui, A., Bachari, N. E. I. & Aoun, N. 2016. Estimation of Hourly Global Solar Radiation Using MSG-HRV images. *International Journal of Applied Environmental Sciences,* 11(2)**,** pp 351-368.

Boussaada, Z., Curea, O., Remaci, A., Camblong, H. & Mrabet Bellaaj, N. 2018. A Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation. *Energies,* 11(3)**,** pp 620.

Box, M. A. 2017. *Physics of Radiation and Climate*: Taylor & Francis Group.

Brabec, M., Badescu, V., Dumitrescu, A. & Paulescu, M. 2016. A new point of view on the relationship between global solar irradiation and sunshine quantifiers. *Solar Energy,* 126(252-263.

Bristow, K. L. & Campbell, G. S. 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and forest meteorology,* 31(2)**,** pp 159-166.

Cano, D., Monget, J.-M., Albuisson, M., Guillard, H., Regas, N. & Wald, L. 1986. A method for the determination of the global solar radiation from meteorological satellite data. *Solar Energy,* 37(1)**,** pp 31-39.

Cebecauer, T. & Suri, M. Accuracy improvements of satellite-derived solar resource based on GEMS re-analysis aerosols.  Conference SolarPACES, 2010.

Chiras, D. 2016. *Power from the Sun: A Practical Guide to Solar Electricity Revised 2nd Edition*: New Society Publishers.

Chukwu, C. 2012. Analysis of some meteorological parameters using artificial neural network method for Makurdi, Nigeria. *African Journal of Environmental Science and Technology,* 6(3)**,** pp 182-188.

Chukwujindu, N. S. 2017. A comprehensive review of empirical models for estimating global solar radiation in Africa. *Renewable and sustainable energy reviews,* 78(955-995.

Chukwujindu, N. S. & Ogbulezie Julie, C. 2017. A critical review of theoretical models for estimating global solar radiation between 2012-2016 in Nigeria.

Crisosto, C., Hofmann, M., Mubarak, R. & Seckmeyer, G. 2018. One-Hour Prediction of the Global Solar Irradiance from All-Sky Images Using Artificial Neural Networks. *Energies,* 11(11)**,** pp 2906.

Dahmani, K., Notton, G., Voyant, C., Dizene, R., Nivet, M. L., Paoli, C. & Tamas, W. 2016. Multilayer Perceptron approach for estimating 5-min and hourly horizontal global irradiation from exogenous meteorological data in locations without solar measurements. *Renewable Energy,* 90(267-282.

Damiani, A., Irie, H., Horio, T., Takamura, T., Khatri, P., Takenaka, H., Nagao, T., Nakajima, T. Y. & Cordero, R. R. 2018. Evaluation of Himawari-8 surface downwelling solar radiation by ground-based measurements. *Atmospheric Measurement Techniques,* 11(4)**,** pp.

Davy, R. J., Huang, J. R. & Troccoli, A. 2016. Improving the accuracy of hourly satellite-derived solar irradiance by combining with dynamically downscaled estimates using generalised additive models. *Solar Energy,* 135(854-863.

De Souza, J. L., Lyra, G. B., Dos Santos, C. M., Ferreira Junior, R. A., Tiba, C., Lyra, G. B. & Lemes, M. A. M. 2016. Empirical models of daily and monthly global solar irradiation using sunshine duration for Alagoas State, Northeastern Brazil. *Sustainable Energy Technologies and Assessments,* 14(35-45.

Diallo, M., Albarelo, T., Primerose, A. & Linguet, L. 2018. Improving the Heliosat-2 method for surface solar irradiation estimation under cloudy sky areas. *Solar Energy,* 169(565-576.

Doorga, J. R. S., Rughooputh, S. D. D. V. & Boojhawon, R. 2019. Modelling the global solar radiation climate of Mauritius using regression techniques. *Renewable Energy,* 131(861-878.

Dubayah, R. 1992. Estimating Net Solar Radiation Using Landsat Thematic Mapper And Digital Elevation Data. *Water Resources Research,* 28(9)**,** pp 2469-2484.

Eissa, Y., Korany, M., Aoun, Y., Boraiy, M., Abdel Wahab, M., Alfaro, S., Blanc, P., El-Metwally, M., Ghedira, H., Hungershoefer, K. & Wald, L. 2015a. Validation of the Surface Downwelling Solar Irradiance Estimates of the HelioClim-3 Database in Egypt. *Remote Sensing,* 7(7)**,** pp 9269-9291.

Eissa, Y., Marpu, P. R., Gherboudj, I., Ghedira, H., Ouarda, T. B. M. J. & Chiesa, M. 2013. Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using SEVIRI images. *Solar Energy,* 89(1-16.

Eissa, Y., Munawwar, S., Oumbe, A., Blanc, P., Ghedira, H., Wald, L., Bru, H. & Goffe, D. 2015b. Validating surface downwelling solar irradiances estimated by the McClear model under cloud-free skies in the United Arab Emirates. *Solar Energy,* 114(17-31.

Eke, R., Betts, T. R. & Gottschalg, R. 2017. Spectral irradiance effects on the outdoor performance of photovoltaic modules. *Renewable & Sustainable Energy Reviews,* 69(429-434.

El-Metwally, M. 2004. Simple new methods to estimate global solar radiation based on meteorological data in Egypt. *Atmospheric Research,* 69(3)**,** pp 217-239.

Fadare, D. A. 2009. Modelling of solar energy potential in Nigeria using an artificial neural network model. *Applied Energy,* 86(9)**,** pp 1410-1422.

Fallahi, S., Amanollahi, J., Tzanis, C. G. & Ramli, M. F. 2018. Estimating solar radiation using NOAA/AVHRR and ground measurement data. *Atmospheric Research,* 199(93-102.

Fan, J., Wang, X., Wu, L., Zhang, F., Bai, H., Lu, X. & Xiang, Y. 2018. New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: A case study in South China. *Energy Conversion and Management,* 156(618-625.

Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X. & Zou, H. 2019. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. *Renewable and Sustainable Energy Reviews,* 100(186-212.

FAO. 2017. Chapter 3 - Meteorological data        *Food and Agriculture Organization.* Español Food and Agriculture Organization.

Feng, F. & Wang, K. 2018a. Merging Satellite Retrievals and Reanalyses to Produce Global Long-Term and Consistent Surface Incident Solar Radiation Datasets. *Remote Sensing,* 10(1)**,** pp 115.

Feng, J., Wang, W. & Li, J. 2018b. An LM-BP Neural Network Approach to Estimate Monthly-Mean Daily Global Solar Radiation Using MODIS Atmospheric Products. *Energies,* 11(12)**,** pp 3510.

Frank, C. W., Wahl, S., Keller, J. D., Pospichal, B., Hense, A. & Crewell, S. 2018. Bias correction of a novel European reanalysis data set for solar energy applications. *Solar Energy,* 164(12-24.

García Cabrera, R. D., Cuevas Agulló, E., Ramos López, R., Cachorro, V. E., Redondas, A. & Moreno Ruiz, J. A. 2018. Description of the Baseline Surface Radiation Network (BSRN) station at the Izaňa Observatory (2009–2017): measurements and quality control/assurance procedures.

Gašparović, I., Gašparović, M. & Medak, D. 2018. Determining and analysing solar irradiation based on freely available data: A case study from Croatia. *Environmental Development*.

Gautier, C., Diak, G. & Masse, S. 1980. A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. *Journal of Applied Meteorology,* 19(8)**,** pp 1005-1012.

GDMS. 2017. *Ministry of Transport and Communications 2016. General  Directorate  of Meteorology  & Seismology* [Online]. Available: http://gdms-krg.org/ku/ [Accessed 14 May 2017].

Geiger, M., Diabaté, L., Ménard, L. & Wald, L. 2002. A web service for controlling the quality of measurements of global solar irradiation. *Solar energy,* 73(6)**,** pp 475-480.

Geuder, N., Affolter, R., Kraas, B. & Wilbert, S. 2014. Long-term behavior, accuracy and drift of LI-200 pyranometers as radiation sensors in Rotating Shadowband Irradiometers (RSI). *Proceedings of the Solarpaces 2013 International Conference,* 49(2330-2339.

Gherboudj, I. & Ghedira, H. 2016. Assessment of solar energy potential over the United Arab Emirates using remote sensing and weather forecast data. *Renewable and Sustainable Energy Reviews,* 55(1210-1224.

Goswami, D. Y., Kreith, F. & Kreider, J. F. 2000. *Principles of Solar Engineering, Second Edition*: Taylor & Francis.

Grüter, W., Guillard, H., Communities, C. o. t. E., Möser, W., Monget, J. M., Palz, W., Raschke, E., Reinhardt, R. E., Schwarzmann, P. & Wald, L. 1986. *Solar Radiation Data from Satellite Images: Determination of Solar Radiation at Ground Level from Images of the Earth Transmitted by Meteorological Satellites - An Assessment Study*: Springer Netherlands.

Gueymard, C. A., Gustafson, W. T., Bender, G., Etringer, A. & Storck, P. Evaluation of procedures to improve solar resource assessments: optimum use of short-term data from a local weather station to correct bias in long-term satellite derived solar radiation time series.  World Renewable Energy Forum Conference Proceedings, 2012. 13-17.

Gueymard, C. A. & Myers, D. R. 2008. Validation and ranking methodologies for solar radiation models. *Modeling solar radiation at the earth's surface.* Springer.

Gueymard, C. A. & Wilcox, S. M. 2011. Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data. *Solar Energy,* 85(5)**,** pp 1068-1084.

Gutierrez-Corea, F. V., Manso-Callejo, M. A., Moreno-Regidor, M. P. & Velasco-Gomez, J. 2014. Spatial estimation of sub-hour Global Horizontal Irradiance based on official observations and remote sensors. *Sensors (Basel),* 14(4)**,** pp 6758-87.

Habte, A., Sengupta, M. & Lopez, A. 2017. Evaluation of the national solar radiation database (nsrdb): 1998-2015,

Habte, A., Sengupta, M. & Wilcox, S. 2012. Comparing measured and satellite-derived surface irradiance. *In Proceedings of the ASME 2012 6th International Conference on Energy Sustainability collocated with the ASME 2012 10th International Conference on Fuel Cell Science, Engineering and Technology, San Diego, CA, USA, 23–26 July.* American Society of Mechanical Engineers: New York, NY, USA, 2012.

Habte, A., Wilcox, S. & Stoffel, T. 2015. Evaluation of Radiometers Deployed at the National Renewable Energy Laboratory's Solar Radiation Research Laboratory,

Haigh, J. D. & Cargill, P. 2015. *The Sun's Influence on Climate*: Princeton University Press.

Harrison, G. & Harrison, R. G. 2014. *Meteorological Measurements and Instrumentation*, Hoboken, UNITED KINGDOM: John Wiley & Sons, Incorporated.

Hassan, G. E., Youssef, M. E., Mohamed, Z. E., Ali, M. A. & Hanafy, A. A. 2016. New Temperature-based Models for Predicting Global Solar Radiation. *Applied Energy,* 179(437-450.

Hassan, M. A., Khalil, A., Kaseb, S. & Kassem, M. A. 2017a. Potential of four different machine-learning algorithms in modeling daily global solar radiation. *Renewable Energy,* 111(52-62.

Hassan, M. A., Khalil, A., Kaseb, S. & Kassem, M. A. 2017b. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Applied Energy,* 203(897-916.

Ibrahim, I. A. & Khatib, T. 2017. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Conversion and Management,* 138(413-425.

Ineichen, P. 2013. Solar radiation resource in Geneva: measurements, modeling, data quality control, format and accessibility.

Ineichen, P., Barroso, C. S., Geiger, B., Hollmann, R., Marsouin, A. & Mueller, R. 2009. Satellite Application Facilities irradiance products: hourly time step comparison and validation over Europe. *International Journal of Remote Sensing,* 30(21)**, pp 5549-5571.

Iqbal, M. 2012. *An Introduction To Solar Radiation*: Elsevier Science.

ISO, I. 1990. 9060: 1990: Specification and classification of instruments for measuring hemispherical solar and direct solar radiation. *Geneva, Switzerland*.

Jadidi, A., Menezes, R., de Souza, N. & de Castro Lima, A. 2018. A Hybrid GA–MLPNN Model for One-Hour-Ahead Forecasting of the Global Horizontal Irradiance in Elizabeth City, North Carolina. *Energies,* 11(10)**, pp 2641.

Jahani, B. & Mohammadi, B. 2018. A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. *Theoretical and Applied Climatology*.

Janjai, S. 2010. A method for estimating direct normal solar irradiation from satellite data for a tropical environment. *Solar Energy,* 84(9)**, pp 1685-1695.

Janjai, S., Pankaew, P. & Laksanaboonsong, J. 2009. A model for calculating hourly global solar radiation from satellite data in the tropics. *Applied Energy,* 86(9)**, pp 1450-1457.

Janjai, S., Pankaew, P., Laksanaboonsong, J. & Kitichantaropas, P. 2011. Estimation of solar radiation over Cambodia from long-term satellite data. *Renewable Energy,* 36(4)**, pp 1214-1220.

Ji, Q. & Tsay, S. C. 2010. A novel nonintrusive method to resolve the thermal dome effect of pyranometers: Instrumentation and observational basis. *Journal of Geophysical Research: Atmospheres,* 115(D7)**, pp.

Jiang, Y. 2009. Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. *Energy,* 34(9)**,** pp 1276-1283.

Jimenez, V. A., Barrionuevo, A., Will, A. & Rodríguez, S. 2016. Neural Network for Estimating Daily Global Solar Radiation Using Temperature, Humidity and Pressure as Unique Climatic Input Variables. *Smart Grid and Renewable Energy,* 07(03)**,** pp 94-103.

Jong, R. D. & Stewart, D. 1993. Estimating global solar radiation from common meteorological observations in western Canada. *Canadian Journal of Plant Science,* 73(2)**,** pp 509-518.

Journée, M. & Bertrand, C. 2011a. Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy,* 85(1)**,** pp 72-86.

Journée, M. & Bertrand, C. 2011b. Geostatistical merging of ground-based and satellite-derived data of surface solar radiation. *Advances in Science and Research,* 6(1-5.

Journée, M., Müller, R. & Bertrand, C. 2012. Solar resource assessment in the Benelux by merging Meteosat-derived climate data and ground measurements. *Solar Energy,* 86(12)**,** pp 3561-3574.

Kaba, K., Sarıgül, M., Avcı, M. & Kandırmaz, H. M. 2018. Estimation of daily global solar radiation using deep learning model. *Energy,* 162(126-135.

Kariuki, B. W. & Sato, T. 2018. Interannual and spatial variability of solar radiation energy potential in Kenya using Meteosat satellite. *Renewable energy,* 116(88-96.

Kaskaoutis, D. G., Kambezidis, H. D., Dumka, U. C. & Psiloglou, B. E. 2016. Dependence of the spectral diffuse-direct irradiance ratio on aerosol spectral distribution and single scattering albedo. *Atmospheric Research,* 178-179(84-94.

KAZ. 2018. *Kipp and Zonen, Solar Instruments* [Online]. Available: http://www.kippzonen.com/ProductGroup/1/Solar-Instruments [Accessed 15 May 2018].

Keshtegar, B., Mert, C. & Kisi, O. 2018. Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. *Renewable and Sustainable Energy Reviews,* 81(330-341.

Khaliliaqdam, N. & Soltani, A. 2012. Quality control and Methods for modeling daily global solar radiation (Case study: Gorgan, Iran). *International Journal of Agriculture and Crop Sciences.*

Kheradmanda, S., Nematollahi, O. & Ayoobia, A. R. 2016. Clearness index predicting using an integrated artificial neural network (ANN) approach. *Renewable and Sustainable Energy Reviews,* 58(1357-1365.

Khosravi, A., Koury, R. N. N., Machado, L. & Pabon, J. J. G. 2018a. Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production,* 176(63-75.

Khosravi, A., Nunes, R. O., Assad, M. E. H. & Machado, L. 2018b. Comparison of artificial intelligence methods in estimation of daily global solar radiation. *Journal of Cleaner Production,* 194(342-358.

Kim, B.-Y., Lee, K.-T., Zo, I.-S., Lee, S.-H., Jung, H.-S., Rim, S.-H. & Jang, J.-P. 2018. Calibration of the Pyranometer Sensitivity Using the Integrating Sphere. *Asia-Pacific Journal of Atmospheric Sciences,* 54(4)**,** pp 639-648.

Kisi, O. 2014. Modeling solar radiation of Mediterranean region in Turkey by using fuzzy genetic approach. *Energy,* 64(429-436.

Klüser, L., Killius, N. & Gesell, G. 2015. APOLLO_NG–a probabilistic interpretation of the APOLLO legacy for AVHRR heritage channels. *Atmospheric Measurement Techniques,* 8(10)**,** pp 4155-4170.

König-Langlo, G., Sieger, R., Schmithüsen, H., Bücker, A., Richter, F. & Dutton, E. 2013. The baseline surface radiation network and its World radiation monitoring centre at the Alfred Wegener Institute. World Meteorological Organization.

Koo, C., Li, W., Cha, S. H. & Zhang, S. 2019. A novel estimation approach for the solar radiation potential with its complex spatial pattern via machine-learning techniques. *Renewable Energy,* 133(575-592.

Korany, M., Boraiy, M., Eissa, Y., Aoun, Y., Abdel Wahab, M. M., Alfaro, S. C., Blanc, P., El-Metwally, M., Ghedira, H., Hungershoefer, K. & Wald, L. 2016. A database of multi-year (2004–2010) quality-assured surface solar hourly irradiation measurements for the Egyptian territory. *Earth System Science Data,* 8(1)**,** pp 105-113.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. 2006. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift,* 15(3)**,** pp 259-263.

Kumar, R., Aggarwal, R. & Sharma, J. 2015. Comparison of regression and artificial neural network models for estimation of global solar radiations. *Renewable and Sustainable Energy Reviews,* 52(1294-1299.

Kumler, A., Xie, Y. & Zhang, Y. 2019. A Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique. *Solar Energy,* 177(494-500.

Lange, S. 2018. Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. *Earth System Dynamics,* 9(2)**,** pp 627-645.

Lave, M. & Weekley, A. 2016. Comparison of high-frequency solar irradiance: Ground measured vs. satellite-derived. *In Proceedings of the 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), Portland, OR, USA, 5–10 June 2016.* IEEE: Piscataway, NJ, USA, 2016.

Lee, H.-J., Kim, S.-Y. & Yun, C.-Y. 2017. Comparison of Solar Radiation Models to Estimate Direct Normal Irradiance for Korea. *Energies,* 10(5)**,** pp 594.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J. W. & Morcrette, J. J. 2013. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques,* 6(9)**,** pp 2403-2418.

Lefèvre, M. & Wald, L. 2016. Validation of the McClear clear-sky model in desert conditions with three stations in Israel. *Advances in Science and Research,* 13(21-26.

Lemos, L. F., Starke, A. R., Boland, J., Cardemil, J. M., Machado, R. D. & Colle, S. 2017. Assessment of solar radiation components in Brazil using the BRL model. *Renewable Energy,* 108(569-580.

Lima, F. J. L., Martins, F. R., Pereira, E. B., Lorenz, E. & Heinemann, D. 2016. Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks. *Renewable Energy,* 87, Part 1(807-818.

Lin, C., Yang, K., Huang, J., Tang, W., Qin, J., Niu, X., Chen, Y., Chen, D., Lu, N. & Fu, R. 2015. Impacts of wind stilling on solar radiation variability in China. *Scientific reports,* 5(15135.

Linares-Rodriguez, A., Ruiz-Arias, J. A., Pozo-Vazquez, D. & Tovar-Pescador, J. 2013. An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. *Energy,* 61(636-645.

Liu, X., Mei, X., Li, Y., Zhang, Y., Wang, Q., Jensen, J. R. & Porter, J. R. 2009. Calibration of the Ångström–Prescott coefficients (a, b) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin. *Agricultural and forest meteorology,* 149(3-4)**,** pp 697-710.

Loghmari, I., Timoumi, Y. & Messadi, A. 2018. Performance comparison of two global solar radiation models for spatial interpolation purposes. *Renewable and Sustainable Energy Reviews,* 82(837-844.

Long, C. & Shi, Y. 2008. An automated quality assessment and control algorithm for surface radiation measurements. *Open Atmos. Sci. J,* 2(1)**,** pp 23-37.

Long, C. N. & Dutton, E. G. 2002. *BSRN Global Network Recommended QC Tests,* v2.0; PANGAEA: Bremerhaven Germany, 2002.

Lopez, G., Batlles, F. & Tovar-Pescador, J. 2005. Selection of input parameters to model direct solar irradiance by using artificial neural networks. *Energy,* 30(9)**,** pp 1675-1684.

López, G., Gueymard, C. A., Bosch, J. L., Rapp-Arrarás, I., Alonso-Montesinos, J., Pulido-Calvo, I., Ballestrín, J., Polo, J. & Barbero, J. 2018. Modeling water vapor impacts on the solar irradiance reaching the receiver of a solar tower plant by means of artificial neural networks. *Solar Energy,* 169(34-39.

Loutfi, H., Bernatchou, A., Raoui, Y. & Tadili, R. 2017. Learning Processes to Predict the Hourly Global, Direct, and Diffuse Solar Irradiance from Daily Global Radiation with Artificial Neural Networks. *International Journal of Photoenergy,* 2017(1-13.

Lu, N., Qin, J., Yang, K. & Sun, J. 2011. A simple and efficient algorithm to estimate daily global solar radiation from geostationary satellite data. *Energy,* 36(5)**,** pp 3179-3188.

Marchand, M., Al-Azri, N., Ombe-Ndeffotsing, A., Wey, E. & Wald, L. 2017. Evaluating meso-scale change in performance of several databases of hourly surface irradiation in South-eastern Arabic Pensinsula. *Advances in Science and Research,* 14(7-15.

Marchand, M., Ghennioui, A., Wey, E. & Wald, L. 2018. Comparison of several satellite-derived databases of surface solar radiation against ground measurement in Morocco. *Advances in Science and Research,* 15(21-29.

Marquez, R., Pedro, H. T. C. & Coimbra, C. F. M. 2013. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy,* 92(176-188.

Mathijssen, T., Bijma, J. & Knap, W. 2018. Traceability of CM-11 pyranometer calibrations at KNMI, (De Bilt).

MATLAB. 2018. The Neural Network Toolbox .The MathWorks, Inc. R2016a ed. Natick, Massachusetts, United States.

Maxwell, E., Wilcox, S. & Rymes, M. 1993. Users Manual for SERI QC Software-Assessing the Quality of Solar Radiation Data,  (Report No. NREL-lP-463-5608, 1617 Cole Boulevard Golden, Colorado).

Meyer, R., Hoyer, C., Schillings, C., Trieb, F., Diedrich, E. & Schroedter, M. SOLEMI: A new satellite-based service for high-resolution and precision solar radiation data for Europe, Africa and Asia.  Proceedings of the 2003 ISES Solar World Congress, GÃuteborg, Sweden, 2003. 14-19.

Michalsky, J., Harrison, L. & Berkheiser III, W. 1995. Cosine response characteristics of some radiometric and photometric sensors. *Solar Energy,* 54(6)**,** pp 397-402.

Mohanty, S., Patra, P. K. & Sahoo, S. S. 2016. Prediction and application of solar radiation with soft computing over traditional and conventional approach – A comprehensive review. *Renewable and Sustainable Energy Reviews,* 56(778-796.

Moradi, I. 2009. Quality control of global solar radiation using sunshine duration hours. *Energy,* 34(1)**,** pp 1-6.

Moradi, I., Mueller, R., Alijani, B. & Kamali, G. A. 2009. Evaluation of the Heliosat-II method using daily irradiation data for four stations in Iran. *Solar Energy,* 83(2)**,** pp 150-156.

Moré, J. J. 1978. The Levenberg-Marquardt algorithm: implementation and theory. *Numerical analysis.* Springer.

Moreno-Tejera, S., Ramírez-Santigosa, L. & Silva-Pérez, M. A. 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. *Renewable Energy,* 78(531-537.

Moreno-Tejera, S., Silva-Pérez, M., Lillo-Bravo, I. & Ramírez-Santigosa, L. 2016. Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. *Solar Energy,* 132(430-441.

Moreno, A., Gilabert, M. & Martínez, B. 2011. Mapping daily global solar irradiation over Spain: a comparative study of selected approaches. *Solar Energy,* 85(9)**,** pp 2072-2084.

Mueller, R., Behrendt, T., Hammer, A. & Kemper, A. 2012. A New Algorithm for the Satellite-Based Retrieval of Solar Surface Irradiance in Spectral Bands. *Remote Sensing,* 4(12)**,** pp 622-647.

Mueller, R., Dagestad, K.-F., Ineichen, P., Schroedter-Homscheidt, M., Cros, S., Dumortier, D., Kuhlemann, R., Olseth, J., Piernavieja, G. & Reise, C. 2004. Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module. *Remote sensing of Environment,* 91(2)**,** pp 160-174.

Mueller, R., Matsoukas, C., Gratzki, A., Behr, H. & Hollmann, R. 2009. The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance—A LUT based eigenvector hybrid approach. *Remote Sensing of Environment,* 113(5)**,** pp 1012-1024.

Mueller, R., Trentmann, J., Träger-Chatterjee, C., Posselt, R. & Stöckli, R. 2011. The Role of the Effective Cloud Albedo for Climate Monitoring and Analysis. *Remote Sensing,* 3(12)**,** pp 2305-2320.

Muller, R. 2014. Calibration and Verification of Remote Sensing Instruments and Observations. *Remote Sensing,* 6(6)**,** pp 5692-5695.

Muneer, T. 2007. *Solar Radiation and Daylight Models*: CRC Press.

Muneer, T. & Fairooz, F. 2002. Quality control of solar radiation and sunshine measurements–lessons learnt from processing worldwide databases. *Building Services Engineering Research and Technology,* 23(3)**,** pp 151-166.

Munkhammar, J. & Widén, J. 2018. An N-state Markov-chain mixture distribution model of the clear-sky index. *Solar Energy,* 173(487-495.

Najmaddin, P. M. 2017. *Simulating river runoff and terrestrial water storage variability in data-scarce semi-arid catchments using remote sensing.* PhD thesis, University of Leicester. Department of Geography, Available at: http://hdl.handle.net/2381/40771, .

Nguyen, B. T. & Pryor, T. 1997. The relationship between global solar radiation and sunshine duration in Vietnam. *Renewable Energy,* 11(1)**,** pp 47-60.

Njau, E. C. 1996. Generalised derivation of the Angstrom and Angstrom-Prescott equations. *Renewable energy,* 7(1)**,** pp 105-108.

Nottrott, A. & Kleissl, J. 2010. Validation of the NSRDB–SUNY global horizontal irradiance in California. *Solar Energy,* 84(10)**,** pp 1816-1827.

NREL. 2017. *National Renewable Energy Laboratory, SOLAR and LUNAR POSITION CALCULATORS* [Online]. Available: https://midcdmz.nrel.gov/apps/go2url.pl?site=BMS [Accessed 7 June 2017].

Ntsangwane, L., Sivakumar, V., Mabasa, B., Zwane, N., Ncongwane, K. & Botai, J. 2018. Quality Control of Solar Radiation Data within the South African Weather Service Solar Radiometric Network.

Ododo, J., Sulaiman, A., Aidan, J., Yuguda, M. & Ogbu, F. 1995. The importance of maximum air temperature in the parameterisation of solar radiation in Nigeria. *Renewable Energy,* 6(7)**,** pp 751-763.

Olano, X., Sallaberry, F., de Jalón, A. G. & Gastón, M. 2015. The influence of sky conditions on the standardized calibration of pyranometers and on the measurement of global solar irradiation. *Solar Energy,* 121(116-122.

Ozgoren, M., Bilgili, M. & Sahin, B. 2012. Estimation of global solar radiation using ANN over Turkey. *Expert Systems with Applications,* 39(5)**,** pp 5043-5051.

Palmer, D., Cole, I., Betts, T. & Gottschalg, R. 2017. Interpolating and Estimating Horizontal Diffuse Solar Irradiation to Provide UK-Wide Coverage: Selection of the Best Performing Models. *Energies,* 10(2)**,** pp 181.

Palmer, D., Koubli, E., Cole, I., Betts, T. & Gottschalg, R. 2018. Satellite or ground-based measurements for production of site specific hourly irradiance data: Which is most accurate and where? *Solar Energy,* 165(240-255.

Pashiardis, S. & Kalogirou, S. 2016. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. *Renewable Energy,* 96(1015-1033.

Paulescu, M., Stefu, N., Calinoiu, D., Paulescu, E., Pop, N., Boata, R. & Mares, O. 2016b. Ångström–Prescott equation: Physical basis, empirical models and sensitivity analysis. *Renewable and Sustainable Energy Reviews,* 62(495-506.

Perez-Astudillo, D., Bachour, D. & Martin-Pomares, L. 2018. Improved quality control protocols on solar radiation measurements. *Solar Energy,* 169(425-433.

Pérez-Burgos, A., Díez-Mediavilla, M., Alonso-Tristán, C. & Dieste-Velasco, M. 2018. Retrieval of monthly average hourly values of direct and diffuse solar irradiance from measurements of global radiation in Spain. *Journal of Renewable and Sustainable Energy,* 10(2)**,** pp 023707.

Petty, G. W. 2006. *A first course in atmospheric radiation*, Second Edition Sundog Publishing.

Pinker, R. & Ewing, J. 1985. Modeling surface solar radiation: Model formulation and validation. *Journal of Climate and Applied Meteorology,* 24(5)**,** pp 389-401.

Polo, J. 2015. Solar global horizontal and direct normal irradiation maps in Spain derived from geostationary satellites. *Journal of Atmospheric and Solar-Terrestrial Physics,* 130-131(81-88.

Polo, J., Martín, L. & Vindel, J. M. 2015. Correcting satellite derived DNI with systematic and seasonal deviations: Application to India. *Renewable Energy,* 80(238-243.

Polo, J., Wilbert, S., Ruiz-Arias, J. A., Meyer, R., Gueymard, C., Súri, M., Martín, L., Mieslinger, T., Blanc, P., Grant, I., Boland, J., Ineichen, P., Remund, J., Escobar, R., Troccoli, A., Sengupta, M., Nielsen, K. P., Renne, D., Geuder, N. & Cebecauer, T. 2016. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. *Solar Energy,* 132(25-37.

Prieto, J., Martínez-García, J. & Garcia, D. 2009. Correlation between global solar irradiation and air temperature in Asturias, Spain. *Solar Energy,* 83(7)**,** pp 1076-1085.

Qin, J., Chen, Z., Yang, K., Liang, S. & Tang, W. 2011. Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products. *Applied Energy,* 88(7)**,** pp 2480-2489.

Qu, Z., Gschwind, B., Lefevre, M. & Wald, L. 2014. Improving HelioClim-3 estimates of surface solar irradiance using the McClear clear-sky model and recent advances in atmosphere composition. *Atmospheric Measurement Techniques,* 7(11)**,** pp 3927-3933.

Quaiyum, S., Rahman, S. & Rahman, S. 2011. Application of Artificial Neural Network in Forecasting Solar Irradiance and Sizing of Photovoltaic Cell for Standalone Systems in Bangladesh. *International Journal of Computer Application,* 32(10)**,** pp 0975-8887.

Quej, V. H., Almorox, J., Arnaldo, J. A. & Saito, L. 2017. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *Journal of Atmospheric and Solar-Terrestrial Physics,* 155(62-70.

Quesada-Ruiz, S., Linares-Rodríguez, A., Ruiz-Arias, J. A., Pozo-Vázquez, D. & Tovar-Pescador, J. 2015. An advanced ANN-based method to estimate hourly solar radiation from multi-spectral MSG imagery. *Solar Energy,* 115(494-504.

Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S. & Khoshnevisan, B. 2014. Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews,* 39(1005-1011.

Rao K, D. V. S. K., Premalatha, M. & Naveen, C. 2018. Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: A case study. *Renewable and Sustainable Energy Reviews,* 91(248-258.

Rasul, A. O. 2016. *Remote Sensing of Surface Urban Cool and Heat Island Dynamics in Erbil, Iraq, between 1992 and 2013.* PhD thesis. University of Leicester. Department of Geography, Available at: http://hdl.handle.net/2381/38508, (Accessed: 04 November 2018).

Renno, C., Petito, F. & Gatto, A. 2016. ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building. *Journal of Cleaner Production,* 135(1298-1316.

Reno, M. J., Hansen, C. W. & Stein, J. S. 2012. Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis, Laboratories, S. N. (Albuquerque, New Mexico 87185 and Livermore, California 94550).

REW. 2003. *Renewable Energy World , The Sun as an Energy Resource* [Online]. Renewable Energy World. Available: https://www.volker-quaschning.de/articles/fundamentals1/index.php [Accessed 20 December 2018].

Rigollier, C., Bauer, O. & Wald, L. 2000. On the clear sky model of the ESRA—European Solar Radiation Atlas—with respect to the Heliosat method. *Solar energy,* 68(1)**,** pp 33-48.

Rigollier, C., Lefevre, M., Cros, S. & Wald, L. Heliosat 2: an improved method for the mapping of the solar radiation from Meteosat imagery.  2002 EUMETSAT Meteorological Satellite Conference, Dublin, Ireland, 1–6 September 2002. EUMETSAT, Darmstadt, Germany, 585-592.

Rigollier, C., Lefèvre, M. & Wald, L. 2004. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. *Solar Energy,* 77(2)**,** pp 159-169.

Rigollier, C. & Wald, L. The HelioClim Project: from satellite images to solar radiation maps.  In Proceedings of the ISES Solar World Congress, 1999. Citeseer.

Riihelä, A., Carlund, T., Trentmann, J., Müller, R. & Lindfors, A. 2015. Validation of CM SAF Surface Solar Radiation Datasets over Finland and Sweden. *Remote Sensing,* 7(6)**,** pp 6663-6682.

Riihelä, A., Kallio, V., Devraj, S., Sharma, A. & Lindfors, A. 2018. Validation of the SARAH-E Satellite-Based Surface Solar Radiation Estimates over India. *Remote Sensing,* 10(3)**,** pp 392.

Rincón, A., Jorba, O., Frutos, M., Alvarez, L., Barrios, F. P. & González, J. A. 2018. Bias correction of global irradiance modelled with weather and research forecasting model over Paraguay. *Solar Energy,* 170(201-211.

Roerink, G. J., Bojanowski, J. S., de Wit, A. J. W., Eerens, H., Supit, I., Leo, O. & Boogaard, H. L. 2012. Evaluation of MSG-derived global radiation estimates for application in a regional crop model. *Agricultural and Forest Meteorology,* 160(36-47.

Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N. & Zhang, T. 2011. Assessment of BSRN radiation records for the computation of monthly means. *Atmospheric Measurement Techniques,* 4(2)**,** pp 339-354.

Saberian, A., Hizam, H., Radzi, M. A. M., Ab Kadir, M. Z. A. & Mirzaei, M. 2014. Modelling and Prediction of Photovoltaic Power Output Using Artificial Neural Networks. *International Journal of Photoenergy,* 2014(1-10.

Sanchez-Lorenzo, A., Wild, M., Brunetti, M., Guijarro, J. A., Hakuba, M. Z., Calbó, J., Mystakidis, S. & Bartok, B. 2015. Reassessment and update of long-term trends in downward surface shortwave radiation over Europe (1939-2012). *Journal of Geophysical Research: Atmospheres,* 120(18)**,** pp 9555-9569.

Sanchez-Lorenzo, A., Wild, M. & Trentmann, J. 2013. Validation and stability assessment of the monthly mean CM SAF surface solar radiation dataset over Europe against a homogenized surface dataset (1983–2005). *Remote Sensing of Environment,* 134(355-366.

Sanchez, G., Serrano, A., Cancillo, M. & Garcia, J. 2015. Pyranometer thermal offset: Measurement and analysis. *Journal of Atmospheric and Oceanic Technology,* 32(2)**,** pp 234-246.

Scharmer, K. & Greif, J. 2000. *The European Solar Radiation Atlas: Fundamentals and Maps*: Presses des MINES.

Schillings, C., Meyer, R. & Mannstein, H. 2004. Validation of a method for deriving high resolution direct normal irradiance from satellite data and application for the Arabian Peninsula. *Solar Energy,* 76(4)**,** pp 485-497.

Schroedter-Homscheidt, M., Hoyer-Klick, C., Killius, N., Lefèvre, M., Wald, L., Wey, E. & Saboret, L. 2017. User's Guide to the CAMS Radiation Service, Copernicus Atmosphere Monitoring Service, DLR,  SoDa,,

Schwandt, M., Chhatbar, K., Meyer, R., Mitra, I., Vashistha, R., Giridhar, G., Gomathinayagam, S. & Kumar, A. 2014. Quality check procedures and statistics for the Indian SRRA solar radiation measurement network. *Energy Procedia,* 57(1227-1236.

Sengupta, M., Weekley, A., Habte, A., Lopez, A. & Molling, C. 2015. Validation of the National Solar Radiation Database (NSRDB)(2005-2012): Preprint,

Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G. & Shelby, J. 2018. The national solar radiation data base (NSRDB). *Renewable and Sustainable Energy Reviews,* 89(51-60.

Şenkal, O. 2010. Modeling of solar radiation using remote sensing and artificial neural network in Turkey. *Energy,* 35(12)**,** pp 4795-4801.

Şenkal, O. & Kuleli, T. 2009. Estimation of solar radiation over Turkey using artificial neural network and satellite data. *Applied Energy,* 86(7-8)**,** pp 1222-1228.

Sharifi, S. S., Rezaverdinejad, V. & Nourani, V. 2016. Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: A comparative study of selected temperature-based approaches. *Journal of Atmospheric and Solar-Terrestrial Physics,* 149(131-145.

Shi, G.-Y., Hayasaka, T., Ohmura, A., Chen, Z.-H., Wang, B., Zhao, J.-Q., Che, H.-Z. & Xu, L. 2008. Data Quality Assessment and the Long-Term Trend of Ground Solar Radiation in China. *Journal of Applied Meteorology and Climatology,* 47(4)**,** pp 1006-1016.

Shi, H., Li, W., Fan, X., Zhang, J., Hu, B., Husi, L., Shang, H., Han, X., Song, Z. & Zhang, Y. 2018. First assessment of surface solar irradiance derived from Himawari-8 across China. *Solar Energy,* 174(164-170.

Soares, J., Oliveira, A. P., Božnar, M. Z., Mlakar, P., Escobedo, J. F. & Machado, A. J. 2004. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique. *Applied energy,* 79(2)**,** pp 201-214.

SoDa. 2018. *Solar Radiation Data* [Online]. Available: http://www.soda-pro.com/ [Accessed 20 March 2018].

Soulayman, S. 2017. *Economical and Technical Considerations for Solar Tracking: Methodologies and Opportunities for Energy Management: Methodologies and Opportunities for Energy Management*: IGI Global.

Spencer, J. 1982. A comparison of methods for estimating hourly diffuse solar radiation from global solar radiation. *Solar Energy,* 29(1)**,** pp 19-32.

Suehrcke, H., Bowden, R. S. & Hollands, K. G. T. 2013. Relationship between sunshine duration and solar radiation. *Solar Energy,* 92(160-171.

Supit, I. & Kapper, R. V. 1998. A simple method to estimate global radiation. *Solar Energy,* 63(3)**,** pp 147-160.

Tang, W., Yang, K., He, J. & Qin, J. 2010. Quality control and estimation of global solar radiation in China. *Solar Energy,* 84(3)**,** pp 466-475.

Thomas, C., Saboret, L., Wey, E., Blanc, P. & Wald, L. 2016c. Validation of the new HelioClim-3 version 4 real-time and short-term forecast service using 14 BSRN stations. *Advances in Science and Research,* 13(129-136.

Thomas, C., Wey, E., Blanc, P. & Wald, L. 2016a. Validation of three satellite-derived databases of surface solar radiation using measurements performed at 42 stations in Brazil. *Advances in Science and Research,* 13(81-86.

Thomas, C., Wey, E., Blanc, P., Wald, L. & Lefèvre, M. 2016b. Validation of HelioClim-3 Version 4, HelioClim-3 Version 5 and MACC-RAD Using 14 BSRN Stations. *Energy Procedia,* 91(1059-1069.

Urraca, R., Antonanzas, J., Sanz-Garcia, A., Aldama, A. & Martinez-de-Pison, F. J. An Algorithm Based on Satellite Observations to Quality Control Ground Solar Sensors: Analysis of Spanish Meteorological Networks. International Conference on Hybrid Artificial Intelligence Systems, 2018a. Springer, 609-621.

Urraca, R., Gracia-Amillo, A. M., Huld, T., Martinez-de-Pison, F. J., Trentmann, J., Lindfors, A. V., Riihelä, A. & Sanz-Garcia, A. 2017a. Quality control of global solar radiation data with satellite-based products. *Solar Energy,* 158(49-62.

Urraca, R., Gracia-Amillo, A. M., Koubli, E., Huld, T., Trentmann, J., Riihela, A., Lindfors, A. V., Palmer, D., Gottschalg, R. & Antonanzas-Torres, F. 2017b. Extensive validation of CM SAF surface radiation products over Europe. *Remote Sens Environ,* 199(171-186.

Urraca, R., Huld, T., Gracia-Amillo, A., Martinez-de-Pison, F. J., Kaspar, F. & Sanz-Garcia, A. 2018b. Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data. *Solar Energy,* 164(339-354.

Urraca, R., Huld, T., Martinez-de-Pison, F. J. & Sanz-Garcia, A. 2018c. Sources of uncertainty in annual global horizontal irradiance data. *Solar Energy,* 170(873-884.

Urraca, R., Martinez-de-Pison, E., Sanz-Garcia, A., Antonanzas, J. & Antonanzas-Torres, F. 2017c. Estimation methods for global solar radiation: Case study evaluation of five different approaches in central Spain. *Renewable and Sustainable Energy Reviews,* 77(1098-1113.

USGS, U. S. G. S. 2018. *digital elevation model (DEM)* [Online]. Available: https://earthexplorer.usgs.gov/ [Accessed 20 May 2018].

VAISALA. 2019. *Vaisala Sensors,*

*Covering every key parameter of surface weather observation for Vaisala MAWS101/201 Automatic Weather Stations* [Online]. Available: http://www.livedata.se/images/Vaisala/MAWS/MAWS_sensors.pdf [Accessed 19 March 2019].

Vernay, C., Blanc, P. & Pitaval, S. 2013. Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3. *Renewable Energy,* 57(339-347.

Vignola, F., Derocher, Z., Peterson, J., Vuilleumier, L., Félix, C., Gröbner, J. & Kouremeti, N. 2016. Effects of changing spectral radiation distribution on the performance of photodiode pyranometers. *Solar Energy,* 129(224-235.

Voyant, C., Muselli, M., Paoli, C. & Nivet, M.-L. 2011. Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. *Energy,* 36(1)**,** pp 348-359.

Walter-Shea, E. A., Hubbard, K. G., Mesarch, M. A. & Roebke, G. 2018. Improving the calibration of silicon photodiode pyranometers. *Meteorology and Atmospheric Physics,* 1-10.

Wang, P., van Westrhenen, R., Meirink, J. F., van der Veen, S. & Knap, W. 2019. Surface solar radiation forecasts by advecting cloud physical properties derived from Meteosat Second Generation observations. *Solar Energy,* 177(47-58.

Wilbert, S., Kleindiek, S., Nouri, B., Geuder, N., Habte, A., Schwandt, M. & Vignola, F. 2016. Uncertainty of Rotating Shadowband Irradiometers and Si-Pyranometers Including the Spectral Irradiance Error. *Solarpaces 2015: International Conference on Concentrating Solar Power and Chemical Energy Systems,* 1734(

WMO. 2008. World Meteorological Organization, Guide to Meteorological Instruments and Methods of Observation, WMO

Xia, S., Mestas-Nuñez, A., Xie, H. & Vega, R. 2017. An Evaluation of Satellite Estimates of Solar Surface Irradiance Using Ground Observations in San Antonio, Texas, USA. *Remote Sensing,* 9(12)**,** pp 1268.

Xie, Y., Yu, T., Gu, X., Zhao, L., Zhang, L. & Wang, G. The estimation of surface daily reflected solar radiation using Landsat-7 ETM+ and empirical models.  In Proceedings of the 2012 5th International Congress on Image and Signal Processing (CISP), Chongqing, China, 16-18 October, 2012. IEEE: Piscataway, NJ, USA,2012, 1109-1113.

Yadav, A. K. & Chandel, S. S. 2014a. Solar radiation prediction using Artificial Neural Network techniques: A review. *Renewable and Sustainable Energy Reviews,* 33(772-781.

Yadav, A. K., Malik, H. & Chandel, S. S. 2014b. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. *Renewable and Sustainable Energy Reviews,* 31(509-519.

Yang, D., Yagli, G. M. & Quan, H. Quality Control for Solar Irradiance Data.  2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), 2018. IEEE, 208-213.

Younes, S., Claywell, R. & Muneer, T. 2005. Quality control of solar radiation data: Present status and proposed new approaches. *Energy,* 30(9)**,** pp 1533-1549.

Yu, Y., Shi, J., Wang, T., Letu, H., Yuan, P., Zhou, W. & Hu, L. 2018. Evaluation of the Himawari-8 Shortwave Downward Radiation (SWDR) Product and its Comparison With the CERES-SYN, MERRA-2, and ERA-Interim Datasets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 99)**,** pp 1-14.

Zahumenský, I. 2004. Guidelines on quality control procedures for data from automatic weather stations. *World Meteorological Organization, Switzerland*.

Zhang, H., Huang, C., Yu, S., Li, L., Xin, X. & Liu, Q. 2018a. A Lookup-Table-Based Approach to Estimating Surface Solar Irradiance from Geostationary and Polar-Orbiting Satellite Data. *Remote Sensing,* 10(3)**,** pp 411.

Zhang, J., Zhao, L., Deng, S., Xu, W. & Zhang, Y. 2017. A critical review of the models used to estimate solar radiation. *Renewable and Sustainable Energy Reviews,* 70(314-329.

Zhang, T., Stackhouse Jr, P. W., Cox, S. J., Mikovitz, J. C. & Long, C. N. 2018b. Clear-Sky Shortwave Downward Flux at the Earth's Surface: Ground-Based Data vs. Satellite-Based Data. *Journal of Quantitative Spectroscopy and Radiative Transfer*.

Zhang, X., Liang, S., Wang, G., Yao, Y., Jiang, B. & Cheng, J. 2016. Evaluation of the Reanalysis Surface Incident Shortwave Radiation Products from NCEP, ECMWF, GSFC, and JMA Using Satellite and Surface Observations. *Remote Sensing,* 8(3)**,** pp 225.

Zo, I.-S., Jee, J.-B., Kim, B.-Y. & Lee, K.-T. 2017. 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)**,** pp 11-19.

Zou, L., Wang, L., Lin, A., Zhu, H., Peng, Y. & Zhao, Z. 2016. Estimation of global solar radiation using an artificial neural network based on an interpolation technique in southeast China. *Journal of Atmospheric and Solar-Terrestrial Physics,* 146(110-122.