ANALYSIS OF SATELLITE DERIVED SOLAR IRRADIANCE IN ISLANDS WITH SITE ADAPTATION TECHNIQUES FOR IMPROVING THE UNCERTAINTY

L. Mazorra Aguiar [a,*], J. Polo [b], J.M. Vindel [c], A. Oliver [a]

[a] University Institute for Intelligent Systems and Numerical Applications in Engineering, University of Las Palmas de Gran Canaria, Edificio Central del Parque Tecnológico, Campus de Tafira, 35017, Las Palmas de Gran Canaria, Spain.

[b] Photovoltaic Solar Energy Unit, Renewable Energy Division, CIEMAT, Avda Computense 40, 28040 Madrid , Spain.

[c] Renewable Energy Division, CIEMAT, Avda Computense 40, 28040 Madrid , Spain

[*] 0034928451986 / luis.mazorra@ulpgc.es

Abstract

Electrical energy production using renewable energies is one of the most important challenges in recent years. Among renewable energies, it is worth highlighting photovoltaic and thermoelectric systems due to their adaptation to the Canary Islands. One of the most important issues to ensure the stability for solar power systems, mostly in insular grids as Canary Islands, is the precise knowledge of solar radiation. In this paper, we focus in Gridded Satellite data suitability for modelling Global Horizontal Irradiation (GHI) in islands with complicated orography, as Canary Islands. Solar radiation data retrieved from CM SAF and McClear model were analysed and compared with 22 ground measurement stations in Canary Islands. Moreover, this analysis presents the results of including a site-adaptation methodology for improving satellite suitability. We used different procedures to perform this site adaptation depending on the solar radiation conditions (clear sky or cloudy sky hours), the location of the measurement station (we establish two clusters according to the climate conditions) and the season. This study could provide information about satellite models suitability in islands and a better knowledge of solar radiation behaviour. Furthermore, accurate satellite radiation data for wide spatial and temporal coverage could improve solar radiation modelling and forecasting.

Keywords

Solar Irradiance, Satellite images, Site adaptation, Clear sky

Highlights

- Assessment of CM SAF satellite derived solar radiation data and clear sky models in islands.
- Selection of clear sky hours and clustering of the region to obtain optimal site adaptation.
- Discussion of satellite assessment and proposal a site-adaptation

1. Introduction

Renewable electrical energy generation development is one of the most important issues for the next years. European Union (EU) institutions approved the renewable energies electrical generation must exceed the 20% of total generation in the EU for 2020 and EU is proposing a limit of 32% of renewable generation for 2030. According with this policy, the project "*Gorona del Viento*" in El Hierro island (Canary Islands, Spain) has reached 100% of renewable electrical generation for the whole island. In case of solar energy, both PV producers and CSP facilities, the knowledge of solar radiation at ground level is one of the most critical issues for increasing the percentage of electrical generation from this resource. Global horizontal irradiance (GHI), direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI) are the main solar radiation components. The assessment of these data from remote sensing retrievals in islands with complex orography is not frequently found in the evaluation of satellite-derived data elsewhere.

In Canary Islands, it is possible to find 22 ground measurement stations located in the seven islands and covering the whole territory. Complex orography of the islands and the effect of global winds, such as Trade Winds, generate a very local and changing climatological variations in the territory. Indeed, an accurate GHI mapping and forecasting in any location could provide better information for managing renewable electrical generation than single measurement datasets. In this sense, we can find solar radiation numerical models to obtain GHI data at any location [1]. This model takes into account the effects of geometrical, astrophysical and atmospheric considerations on the radiation at ground level. Surface GHI on the whole territory is calculated using factors as elevation, albedo, surface inclination and shadows casts [2]. This model uses an adaptive mesh of triangles to represent the terrain and its orography, a clear sky spatial model [3] and ground measurement station data.

Surface solar irradiance can also be obtained using satellite models from meteorological satellites images [4], such as Meteosat or NOAA (National Oceanic and Atmospheric Satellite). With this aim, satellites analysed clouds with different spatial and temporal resolutions using different sensors. NOAA and MetOp (Meteorological Operational Satellite) polar-orbiting satellites use the Advanced Very High-Resolution Radiometer (AVHRR) sensor. This family of sensors provide images of visible (VIS), near-infrared (NIR) and infrared (IR) bands. Climate data are derived from AVHRR sensor mounted by polar-orbiting NOAA and METOP satellites. This data record is known as "the CM SAF Cloud, Albedo And Surface Radiation dataset from AVHRR data" – second edition (CLARA-A2) [5]. CLARA-A2 products present different spatial resolution, from 0.05°x0.05° to 0.25°x0.25°, depending on the temporal resolution and the specific product. Moderate Resolution Imagin Spectroradiometer (MODIS) instruments have improved VIS/IR imager on board of polar-orbiting satellites. Terra and Aqua satellites mounted these devices. Earth's Radiant Energy System (CERES) is designed to improve the knowledge of the relation between cloud properties and solar and longwave radiation. CERES analyses data from the Tropical Rainfall Measuring Mission Visible and Infrared Scanner and the MODIS instruments covering period from 1998 to 2007 [6].

Polar orbiting satellites data presents a low temporal sampling rate. Geostationary instruments mounted in satellites improve this weakness obtaining higher temporal resolution information. Meteosat Second Generation (MSG) satellites mounted on board one of the most advanced imager in geostationary satellites, Spinning Enhanced Visible and InfraRed Imager (SEVIRI). Heliosat methods provide solar radiation data at ground level using Meteosat images. First Heliosat version provides solar radiation data using a linear relation between clearness index

and cloud index. This relation was later substituted by the clear sky index instead of clearness index [7]. In this way, Heliosat-2 estimates solar radiation using a clear sky model [8]. Moreover, Heliosat-3 introduces new formulations for clear sky transmittance and cloud index to improve the model using MSG visible channel [9]–[11]. Satellite Application Facility on Climate Monitoring (CM SAF) provides solar radiation and cloud characteristics processing images from MSG geostationary satellite network with SEVIRI sensor on board and NOAA polar satellites with AVHRR [12], [13]. Heliosat method and MAGIC approach converts satellite information and images in global and direct normal irradiance. The solar radiation data estimated with this method were validated with Baseline Surface Radiation Network (BSRN) ground stations and provided in Surface Solar Radiation Data Set – Heliosat (SARAH) database [14].

Solar radiation data retrieved from satellite models give high spatial and temporal resolutions depending on the geographical area. Satellite models provide both GHI, DNI and some information about clouds and atmosphere conditions depending on the model and geographical area. The accurateness of these satellite-derived data compared with ground data is different depending on the satellite model and location. In this way, Eissa studied the resolution of Helioclim [15], [16] data and surface downwelling solar irradiances estimated by the McClear [17] model under cloud-free skies in several stations in Egypt. They report worse errors (between 17 and 30% in general) in northern station closer to the sea. While Ineichen [18] worked with BSRN stations with hourly data a reported an assessment study with errors around 17% for global radiation and 34% for direct normal irradiance. On the other hand, using CMSAF database Antonanzas et al. [19] shows a satellite derived data assessment study for monthly and yearly time step. In this study, CMSAF was compared with a set of ground stations in Spain and obtained around 4% rRMSE. Vindel et al. [20] compared the accuracy of CMSAF and ERA-Interim reanalysis [21], provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), in three stations with different climates in Spain. The results obtained in this work are in agreement with the study of Bojanowski et al. [22] (rRMSE for CMSAF among 10%-15% in zones with Mediterranean climate and among 20%-30% in zones with oceanic climate).

Satellite solar radiation data often shows systematic errors, such as overestimation or underestimation trends, often due to regional inconsistencies in the external aerosols input data and other systematic errors. This bias between satellite and ground data could be reduced by finding a correction factor from the correlation of both time series [23], [24], [25]. It is also possible to reduce bias using a linear fitting of both datasets, as explained by Polo et al. [26] for stations in India depending on the season. Other methods include non-linear fittings, as Mieslinger et al. [27] minimizing a quadratic error expression, Schumann [28] using cumulative distribution function characteristics or Bender et al. [29] combining short-term ground measurements with longer-term satellite data with a multi-variate linear regression analysis. Vernay et al. [30], who proposed a new method using Fourier decomposition for calibrating daily global irradiation retrieved from HelioClim-3 database.

This paper reports an assessment of satellite-derived data using CMSAF hourly database for islands with complicated orography. The study divides results and discussion, taking into account the location, by clustering the region according to climatological conditions and sky conditions (distinguishing between clear sky and cloudy sky hours). The reader finds detail information about the level of certainty obtained with CMSAF and McClear model. Moreover, a method to improve solar radiation assessment by the combination of a clear sky model and a site adaptation model is proposed. In this way, this paper establish an easily

101 extensible methodology for satellite-derived radiation data adaptation, based on a physical
102 clustering of data and a linear regression model. Section 2 shows the data used in this paper,
103 both ground measurement and satellite, while Section 3 provides information about the
104 methodology performed to obtain the results. Section 4 and 5 present the results obtained for
105 satellite-derived data comparison with ground measurement and site adaptation results
106 respectively. Finally, Section 6 reports the conclusion of this paper.
107
108 **2. Solar radiation data**

109 In this study, ground data from measurement stations have been used to validate both
110 satellite-derived and clear sky models data.
111
112 **2.1 Ground Data Set**

113 Satellite-derived data provide information with high spatial and temporal resolution, so they
114 are an important source to study solar radiation behavior in many locations. However, satellite
115 models should be validated using ground measurement datasets in order to establish the
116 accuracy of these data. In this paper, we used 22 ground stations around Canary Islands to
117 compare and study satellite GHI data and clear sky data. Ground stations are handled by the
118 Agroclimatic Information System for Irrigation (SIAR) and the State Meteorological Agency
119 (AEMET). The raw data obtained for this work include the years 2010 and 2011.
120 .

Station	Island	Source	Lat (º)	Long. (º)	Alt. (m)	Period	Nº data (filter)	clear sky %	Mean W/m²	Std W/m²
Gáldar (GAL)	Gran Canaria	SIAR	28.16	-15.67	16	2010-11	7250	28	455.06	251.01
San Mateo (SMA)	Gran Canaria	SIAR	28.02	-15.53	785	2010-11	7250	32	447.72	283.77
Vecindario (VEC)	Gran Canaria	SIAR	27.84	-15.43	76	2010-11	7267	55	537.49	264.87
Maspalomas (MASP)	Gran Canaria	AEMET	27.76	-15.76	265	2010-11	7285	58	541.90	261.40
Izaña (IZA)	Tenerife	AEMET	28.31	-16.50	2371	2010-11	7265	72	634.24	295.68
Garimba (GAR)	Tenerife	SIAR	28.51	-16.39	493	2010-11	7266	13	389.94	243.71
Valle Guerra-Pajarillos (PAJ)	Tenerife	SIAR	28.53	-16.39	110	2010-11	7245	23	443.32	258.11
El Pico (PIC)	Tenerife	SIAR	28.52	-16.37	256	2010-11	7103	22	446.73	266.20
Puerto de la Cruz (PCR)	Tenerife	SIAR	28.41	-16.53	142	2011	3568	18	389.79	249.90
Guía de Isora (ISO)	Tenerife	SIAR	28.23	-16.83	48	2011	3568	42	515.93	250.79
La Fuente (LFU)	Tenerife	SIAR	28.37	-16.86	28	2010-11	7237	18	419.21	247.86
La Laguna – Güímar (LLA)	Tenerife	SIAR	28.32	-16.38	156	2011	3569	47	485.28	264.89
San Sebastián (SSB)	La Gomera	SIAR	28.10	-17.12	63	2010-11	7236	41	486.29	264.21
Hermigua (HER)	La Gomera	SIAR	28.17	-17.20	213	2010	3670	11	382.69	281.10
Barlovento (BAR)	La Palma	SIAR	28.83	-17.78	139	2010-11	7230	12	371.22	243.59
Tazacorte II (TAZII)	La Palma	SIAR	28.60	-17.92	94	2010-11	7248	31	482.84	276.41
Los Llanos de Aridane (ARI)	La Palma	SIAR	28.65	-17.92	281	2010-11	6730	39	492.18	258.64
Fuencaliente (FUE)	La Palma	SIAR	28.49	-17.87	65	2010-11	7249	12	443.91	276.72
Frontera (FRO)	El Hierro	SIAR	27.78	-18.01	54	2011	3572	18	439.54	262.06
Haría (HAR)	Lanzarote	SIAR	29.10	-13.48	105	2010-11	7124	29	478.77	256.68
Tinajo (TIN)	Lanzarote	SIAR	29.05	-13.66	271	2011	7260	17	435.90	243.88
Antigua (ANT)	Fuerteventura	SIAR	28.33	-13.94	120	2010-11	7266	44	501.45	256.75

Table 1.- Ground measurement stations and datasets information.

It is a common practice in solar radiation discussions to use a quality check procedure in order to avoid wrong measurement stations [31], [32], [16]. In this paper, we used the physically possible limits and maximum allowed variability to filter out raw GHI data. Moreover, in this paper the authors compared the whole data set with a clear sky model (McClear) day by day. In this way, we checked and filtered out any extremely rare situation. Indeed, only global solar radiation below exoatmospheric theoretical radiation and over zero are considered valid data. In case an hourly data is not valid, the whole day is rejected, so the final dataset only contains valid hourly data for complete days. Furthermore, to calculate errors and deviation between satellite and ground data, night values were not considered. The variable used to distinguish between night and daily data is the zenith angle, taking into account a limit of 80º. Table 1 shows the main information of each location used in this survey and the final valid data once we filtered out datasets.

2.2 Satellite-derived Data Set

As explained before, satellite-derived data used in this survey were retrieved from the Satellite Application Facility on Climate Monitoring (CM SAF). CM SAF datasets estimate solar radiation from images taken with Meteosat Second Generation (MSG) geostationary satellite network with SEVIRI sensor on board and NOAA polar satellites with AVHRR sensor. The model used by CMSAF to calculate global and direct normal irradiance from satellite information and images is based on the Heliosat method and the MAGICSOL approach. The radiation data estimated with this method are validated with BSRN ground stations and provided in SARAH database [14]. In this paper, the hourly global solar radiation data comes from the "*SARAH-1 database* [12], *specifically the v002, SIS - Surface incoming shortwave radiation, MVIRI/SEVIRI on METEOSAT, Hourly, Mean, MSG full disk (includes Europe, Africa, Atlantic Ocean)"*. The region of interest retrieved for this paper corresponds to the Canary Islands, from 27º N to 30º N in latitude and from -13º W to -19º W in longitude, with a spatial resolution of 0.5x0.5º for each pixel. The period considered is 01/012010 to 31/12/2011 for the whole area. The variable retrieved and studied in this paper is the Surface incoming shortwave radiation (SIS) from SARAH-1 database. SIS corresponds to the Global Solar Radiation, expressed in Wm^{-2}, and it is considered as the irradiance reaching a horizontal plane at the Earth surface in the 0.2 - 4 µm wavelength region.

This study consists of validating CM SAF hourly data for the Canary Islands ground measurement stations and provides a discussion about satellite data assessment in islands. Before calculating errors for each location with hourly data provided for the whole grid in years 2010 and 2011, we estimated daily average following CM SAF recommendations [14]:

$$SIS_{DA} = SIS_{CLSDA} \frac{\sum_{i=1}^{n} SIS_i}{\sum_{i=1}^{n} SIS_{CLSi}} \tag{1}$$

Where SIS_{CLSDA} is the Clear sky model daily average, SIS_i is ground IGH hourly data and SIS_{CLSi} is clear sky IGH hourly data. The monthly average is calculated from the daily means, SIS_{DA}. These daily and monthly averages allow us to overview the general approach of satellite-derived data to the Canary Islands climate conditions. Fig. 1 shows monthly means of CM SAF database global solar radiation in July and October for Canary Islands. Gridded satellite shows quite satisfactorily the general trend of the climatological conditions in the whole area. In summer months, a major presence of clouds in northern areas is easily

170 observed. On the other hand, during fall months solar radiation shows stable levels throughout
171 the whole grid. The presence of Trade winds, mostly during summer months, generates an
172 accumulation of clouds because of the complex orography of the northern parts of the Islands.
173 Satellite-derived data show consistent information according to the climatological knowledge
174 and ground measurement.
175
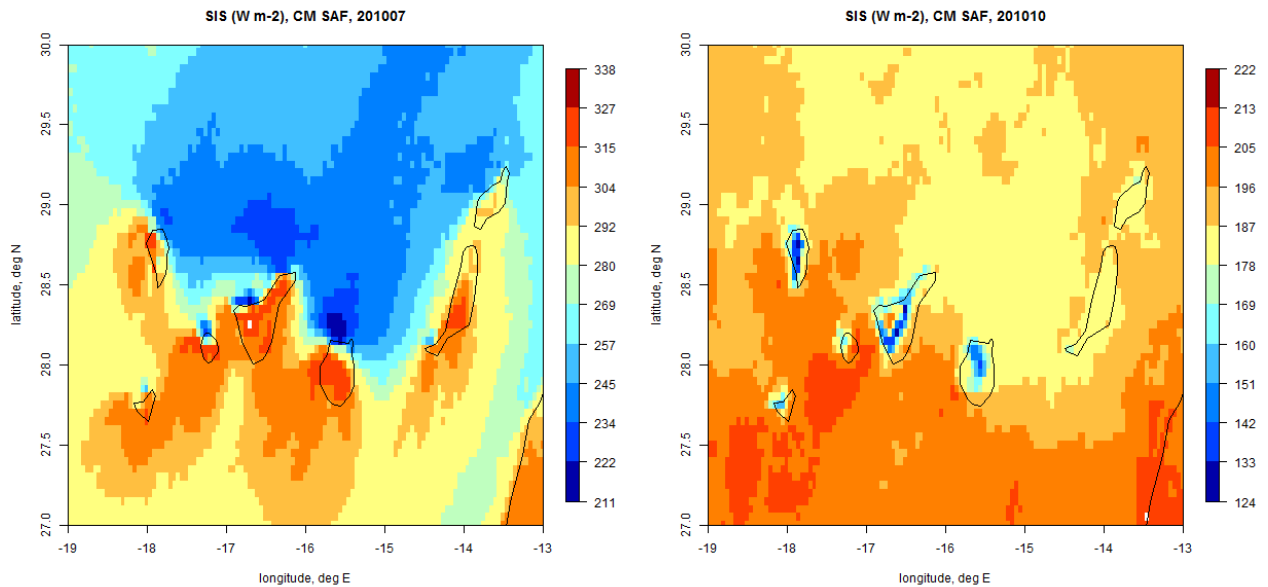
177 **Figure 1.-** SIS Monthly means of gridded satellite data for Canary Islands, July (left) and
178 October (right) 2010.

179
180

181   3.   Methodology of satellite data assessment and adaptation

182   Firstly, satellite hourly data are compared with ground data in every location individually.
183 In order to establish a rigorous comparison for all types of weather conditions, we have
184 distinguished between hours with clear skies and hours with a presence of clouds. In this way,
185 the precision of the model can be discerned in both conditions and also compared with clear-
186 sky models. As observed in Fig. 1, the climatological conditions in the northern zone of the
187 islands are completely different from southern one. Therefore, a division of the entire territory
188 of the islands has been made taking into account solar radiation data retrieved from the
189 satellite for the whole grid. We have divided both satellite and ground measurement station in
190 two clusters and studied the comparison of both clusters separately. Indeed, the next sections
191 explain the steps followed in this study.

192

193 3.1 Clear sky models

194   A clear sky model estimates solar radiation components at ground level only taking into
195 account clear skies' instants, i.e. for cloudless conditions. In the bibliography, it is possible to
196 find several clear models with very good accuracy. Most of these models estimate global solar
197 radiation using different atmospheric parameters that represent the state of the atmosphere in
198 terms of attenuation of solar radiation. Most of these models use attenuation parameters as the
199 aerosol optical depths (AODs), water vapour, ozone, Linke turbidity factor or pressure.
200 AERONET measurement stations provide AODs and water vapour data for a wide variety of

locations and a broad period of time [33], while World Ozone Monitoring Mapping provided by the Canadian Government give the possibility of retrieving ozone values in several stations [34]. Monitoring Atmospheric Composition and Climate (MACC) project also offers AODs, water vapour, ozone and other data records on atmospheric composition for the whole world from 2004, available in [35]. In addition, the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2), is a long-term global reanalysis to assimilate space-based observations of aerosols and represent their interactions with other physical processes in the climate system. MERRA-2 provides data from 1980 [36]. The results obtained comparing different clear sky models with ground measurement data only for clear sky conditions offers very good results all around the world [37], [38].

One of the most well-known clear sky models in solar energy community is Bird and Hullstrom model [39], which considers several mechanisms of interaction with the atmosphere as Rayleigh dispersion, absorption of ozone, oxygen, CO_2 and water vapour or absorption and dispersion of aerosols. Another widely used in solar energy field is the REST2 model [40], a dual-band model based on the CPCR2 model including spectral distribution of exoatmospheric radiation, solar constant, Angstrom turbidity, water vapour and reduced NO_2. The clear sky model used in this paper is McClear sky model [41]. This model estimates data series for global horizontal clear sky irradiance (GHI_{cs}), direct normal clear sky irradiance (DNI_{cs}) and diffuse horizontal clear sky irradiance (DHI_{cs}) and it is available in [35] from 2004 to current day minus 2, *"d-2"*. This model is available worldwide with a temporal resolution of minute, hourly, daily or monthly time step. MACC project provides atmospheric composition parameters used in McClear model, such as AOD at 550 and 1240 nm, water vapour and ozone column. McClear model differs from the transmittance models mentioned above in the implementation approach. McClear is a Look-up table (LUT) model as a result of many radiative transfer calculations with libRadtran covering the whole range of values of the main involving parameters. The results of McClear are available from Copernicus Atmosphere Monitoring Service (CAMS, http://www.soda-pro.com/web-services/radiation/cams-mcclear).

3.2 Identification of clear sky days

A very important issue when studying a clear sky model is to assess the accuracy to estimate solar radiation comparing with ground measurements. Firstly, it is necessary to separate the clear skies (cloud free skies) from the cloudy skies in our ground dataset. In this was, it is possible to compare the clear sky model with clear skies ground data. It is possible to find different methods in the bibliography to find out only clear sky hours. A group of methods estimates clear sky hours using and relating global, beam and diffuse radiation. Ineichen [37], [42] proposes a method to detect clear skies studying the stability of clearness index and the broad band AOD and relating diffuse, global and beam solar radiation. Lefevre et al. [38], in the same way, propose studying clearness index, corrected clearness index, direct normal radiation clearness index and diffuse fraction.

In this paper, the method proposed detects clear sky individual hours from the ground datasets only using GHI [43], [44]. The model separates clear and cloudy skies comparing for each period a clear sky model with the ground data,. To evaluate the comparison of both data series for each day, we calculate the correlation coefficients matrix, C, and its determinant, Eq. (2). If this determinant is lower than a threshold this period is considered clear sky. The threshold should be established experimentally once we have observed the data.

$$C = \begin{bmatrix} \rho_{GHI,GHI} & \rho_{GHI,GHI_{cs}} \\ \rho_{GHI_{cs},GHI} & \rho_{GHI_{cs},GHI_{cs}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{GHI,GHI_{cs}} \\ \rho_{GHI_{cs},GHI} & 1 \end{bmatrix}, \; F = \det(C) \qquad (2)$$

$$\rho_{GHI_{cs},GHI} = \frac{Cov(GHI_{cs},GHI)}{\sigma_{GHI_{cs}}\sigma_{GHI}} \qquad (3)$$

Where cov denotes the covariance, σ means the standard deviation, GHI and GHIcs denote the hourly time series for global irradiance measured and that of clear sky estimated, respectively. This method separates daily periods of clear and cloudy skies, so we obtained days completely cloud-free to evaluate the clear sky models. However, to obtain a higher number of clear sky hours, Reno and Hansen [38][45] use a moving window of period of times with 1 min. data series to detect individual cloud-free conditions. This method separates clear and cloudy skies using several conditions and thresholds. These conditionsare the maximum GHI value in each period, the mean GHI value in each period and three different parameters to study the variability of each period. As in the previous case, the threshold of each condition should be considered experimentally once we have observed ground hourly data. If the considered period meets all the conditions, the hours of this period are considered clear sky.

In this paper, a methodology using both previous methods to detect clear skies individual hours is proposed. First, if one single day meets the daily correlation determinant condition, F, the whole hours of this day are considered clear skies [43]. If not, we used the moving window method to detect possible individual cloud-free hours. As, Reno and Hansen [45] works with 1 min. data series, we should stablish new threshold for each condition. Each moving window, with four hours each window, should meet five conditions to be considered clear sky. If one hour belongs at least to a window considered clear sky, this hour is also considered cloud-free. The conditions and thresholds used in this paper are obtained experimentally studying the data:

- **Daily correlation determinant (F) between clear sky model and ground data**, Eq. (2), should be lower than 0.02.
- **Mean value of GHI** in the window should be lower than 30 W/m^2.
- **Maximum value of GHI** in the window should be lower than 30 W/m^2.
- **Line length (L) of GHI** in the window, Eq. (4), should be between -19 and 19, with the GHI length calculated with hourly data in W/m^2. Where "t" is the time step, in this case hourly data.

$$L = \sum_{i=1}^{n-1} \sqrt{(GHI_{i+1} - GHI_i)^2 + (t_{i+1} - t_i)^2} \qquad (4)$$

- **Standard deviation of rate of change (s), Eq, (5), in GHI** in the window should be lower than 0.24 calculated with hourly data in W/m^2.

$$s_i = \frac{GHI_{i+1} - GHI_i}{t_{i+1} - t_i} \; \forall \, i \in \{1, 2, \dots, n-1\} \qquad (5)$$

- **Maximum difference (X) between changes in GHI and clear sky time series** in the window, Eq. (6) should be lower than 26 W/m^2.

$$X = max\{|(GHI_{i+1} - GHI_{i+1,cs}) - (GHI_i - GHI_{i,cs})|\} \; \forall \, i \in \{1, 2, \dots, n\} \qquad (6)$$

## 3.3 Clustering using satellite solar radiation data

As explained in Section 2.2, satellite-derived data show a clear difference between northern and southern areas of the islands. The northern area apparently presents lower solar radiation values and major presence of clouds. This fact is consistent with the previous knowledge of the Canary Islands general climatological conditions. Trade Winds from Northeast direction induce a major presence of clouds in the north area because of the complex and high orography. This issue is also observable in ground measurement data for all stations. Northern stations mostly present less than 40% of clear sky hours in general, while southern stations present more than 40% of clear sky hours, Table 1. Moreover, the yearly mean of the hourly solar radiation data for all northern stations is higher than for southern stations, Tables 1 and 2.

In this study, we proposed a comparative study between satellite and ground data series separating northern and southern stations. As a site-adaptation for satellite data is proposed, it is necessary to stablish the border between the two geographical areas using the whole grid information. In this case, we propose to separate both regions using the GHI (W/m^2) satellite-derived data series for the whole year. The clustering analysis will allow the identification of regions with different radiation patterns, where we can analyze the specific behavior of this variable. K-means algorithm [46], [47] is the most widely used technique for clustering, and thus, it has been the method chosen for our analysis. This technique employs an algorithm to minimize the sum of squared distances between the objects of each group and the centroid of this group. The algorithm is implemented as follows: first, initial clusters are selected randomly; the distances between centroids and data of these initial clusters are obtained; each data is allocated within the cluster in which its distance to the centroid is the lowest; from these new data, new centroids are calculated. This process must be repeated until the sum of distances between cluster centroids and data converges.
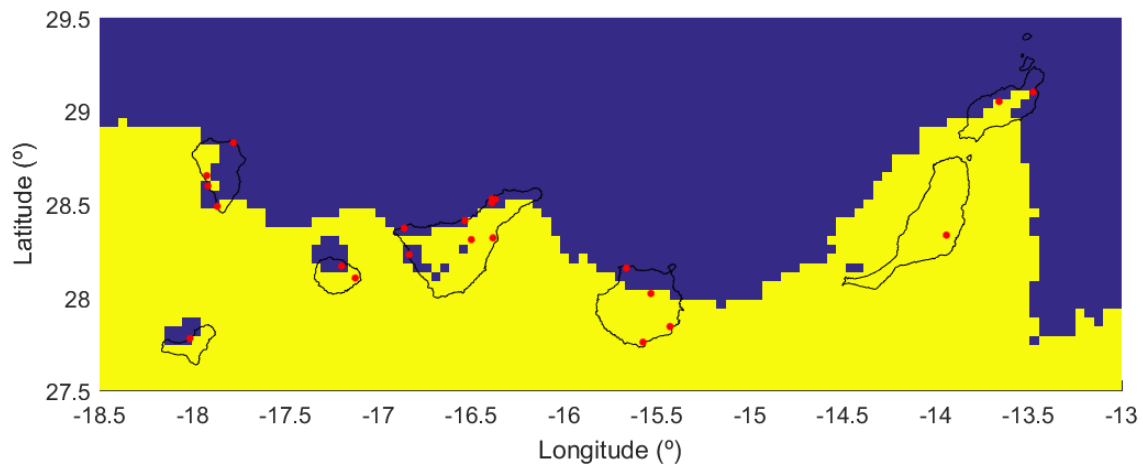
**Figure 2.-** Cluster north (blue) and south (yellow) according to climate conditions observed in satellite data. The ground measurement stations used in this study in red.

Fig. 2 shows the two clusters obtained for the whole Canary Islands region according to k-means algorithm. Both regions are completely consistent with previous knowledge of climatological condition. Fig. 3 shows images taken from the Meteosat Second Generation

9

332   (MSG) at several days, where a very similar pattern of clouds can be observed. Besides, the
333   separation of North and South regions is observed in many days throughout the year.
334   Moreover, in general, most of the measurement stations situated in Cluster South present a
335   higher mean and more percentage of clear sky days than stations in Cluster North. Only, TIN
336   (south) presents less than 40% of clear days and ISO (north) presents more than 40%, being
337   both placed in the border of both regions.
338
339
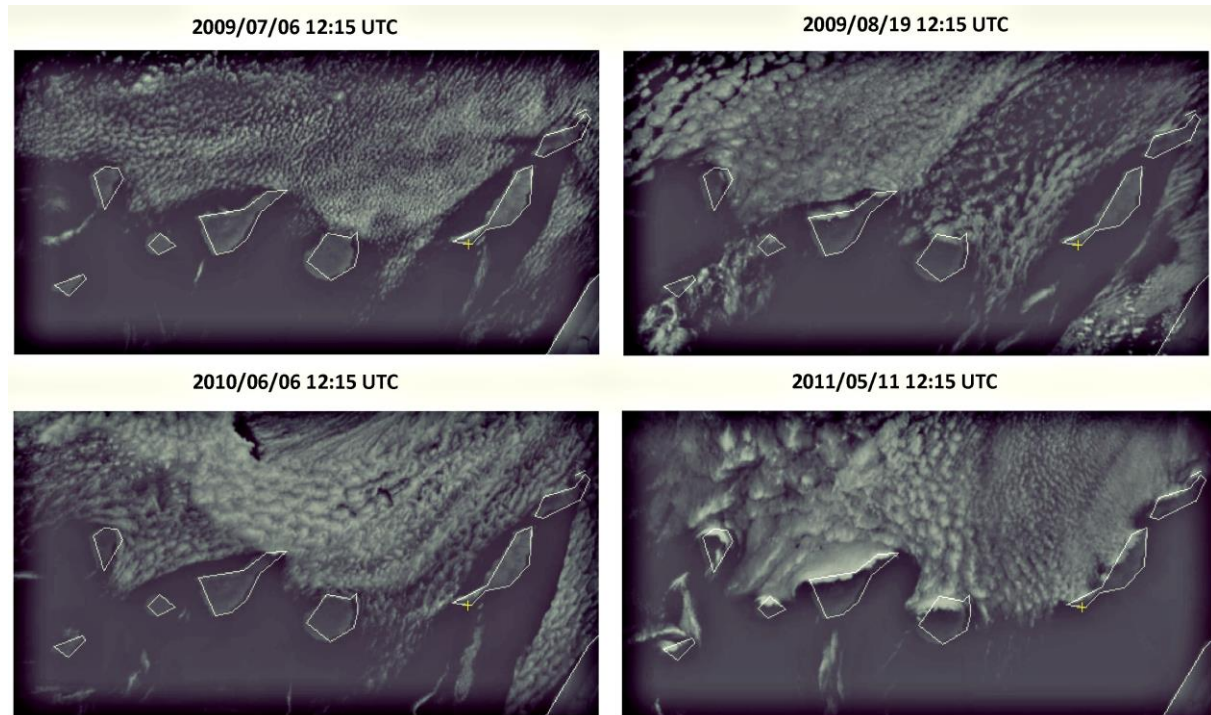
340
341   **Figure 3.-** Satellite images of Canary Islands in summer months obtained with MSG. It is
342   possible to view the presence of clouds in northern area.
343
344
345   **3.4 Site-adaptation methodology**

346   Satellite-derived data series are basically based on the estimation of clouds' properties and
347   their influence in sky conditions. Satellite models need information of some atmospheric
348   components to calculate the sky transmittance under cloudless conditions. Indeed, the
349   accuracy of these models can be affected by several uncertainty sources. Satellite data series
350   provide information of solar radiation of almost every location on Earth, however they cannot
351   estimate some local effects, as complex terrain effects or snow albedo. Nevertheless, many
352   improvements have been performed in satellite models in recent years [48]–[55]. Despite the
353   improvements achieved recently, the use of simultaneous ground data can help reduce even
354   more the uncertainty by correcting some systematic errors on the retrieval; these techniques
355   are commonly referred to as site-adaptation [56].
356
357   In this paper, a linear method is used to correct the bias with ground data estimating a
358   correction factor [26]. It is possible to study the bias between two series using scatter plots
359   between satellite and ground data, shown in **Fig. 4** and **Fig. 5**. Clouds of points linear fitting
360   shows a deviation with ideal linear fitting. Linear fitting obtained in the scatter plot between
361   satellite and ground data is estimated using least squares method. Both linear regression

362 coefficients, *{a,b}*, are linear equation coefficients. Linear site adaptation tries to move
363 satellite linear fitting to the ideal fitting with ground data using the following expression:
364

$$y_{sat,new} = y_{sat} - \left[(a-1)x_{ground} + b\right] \qquad (7)$$

366

367 Where y_{sat} represents the original CM SAF solar radiation, x_{ground} is the ground dataset
368 and *{a,b}* are the linear regression coefficients obtained in the scatter plot between satellite
369 and ground data. Hence, expression in Eq (7) gives us a new GHI satellite-derived dataset,
370 $y_{sat,new}$, using the original satellite data and ground data at each location or cluster. The new
371 dataset has been adapted to ground observed data and presents a smaller bias and dispersion.
372 Eq. (7) is based on both satellite original data and ground data, so it is only useful at locations
373 with available ground measurement. In order to establish an expression extended to the whole
374 Canary Islands' grid, we calculated a new comparison between new satellite data corrected
375 with Eq. (7) and original satellite data. The following linear expression provides the
376 possibility of calculating corrected satellite data for the rest of the grid, Eq. (8).
377

$$\cdot \, y_{sat,new} = a^* \cdot y_{sat} + b^* \qquad (8)$$

379

380 Again, *{a*,b*}* are the linear regression coefficients obtained in the scatter plot between
381 satellite original and adapted satellite data. So, these two new coefficients relate directly
382 original satellite data with adapted satellite data. As Eq (8) only use satellite data and not any
383 ground data, this equation allow us to estimate adapted data for the whole grid. As explained,
384 when observing ground and satellite data, a meaningful clear difference between south and
385 north data is observed. Indeed, in order to find the optimal bias removal, we separated Canary
386 Islands in two clusters according to climatological conditions. Both clusters contain all data
387 series for ground stations belonging geographically to this cluster, so *{a*,b*}* linear
388 coefficients allow us to estimate adapted satellite data for each whole cluster respectively. We
389 used a training dataset to establish the model to find optimal correction factor. This model
390 will be later validated with another independent set, testing dataset. In this way, it is possible
391 to avoid overfitting and study the accuracy of the model when new data are presented for the
392 whole grid. Table 2 presents all cluster dataset. Cluster north dataset offers a lower mean
393 value than south dataset as obtained in individual ground stations data. Moreover, the
394 presence of clear skies in cluster north is lower than 30% and in cluster south it is over 40%.
395 The division between training and testing datasets (60 and 40% of the total respectively)
396 preserves these conditions.
397
398

CLUSTER	Nº data (filter)	Mean W/m²	clear sky (%)	Std W/m²
NORTH	72206	436.37	23	260.50
SOUTH	39747	507.18	46	262.43
NORTH (TRAINING)	43324	437.47	23	260.68
SOUTH (TRAINING)	23848	506.44	46	261.68
NORTH (TESTING)	28882	434.72	23	260.22
SOUTH (TESTIING)	15899	508.28	46	263.58

399 **Table 2.-** Cluster datasets divided in training and validating.
400

11

401
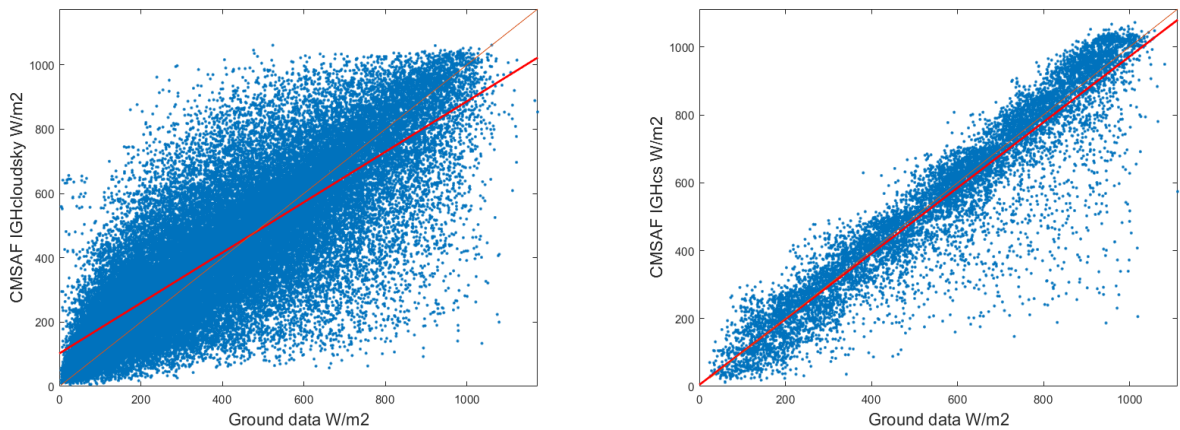
403 **Figure 4.-** SIS hourly datasets comparison between ground data and CMSAF data for all
404 ground stations in Cluster North, cloudy sky
405 (left) and clear sky (right) for 2010 and
2011.

407
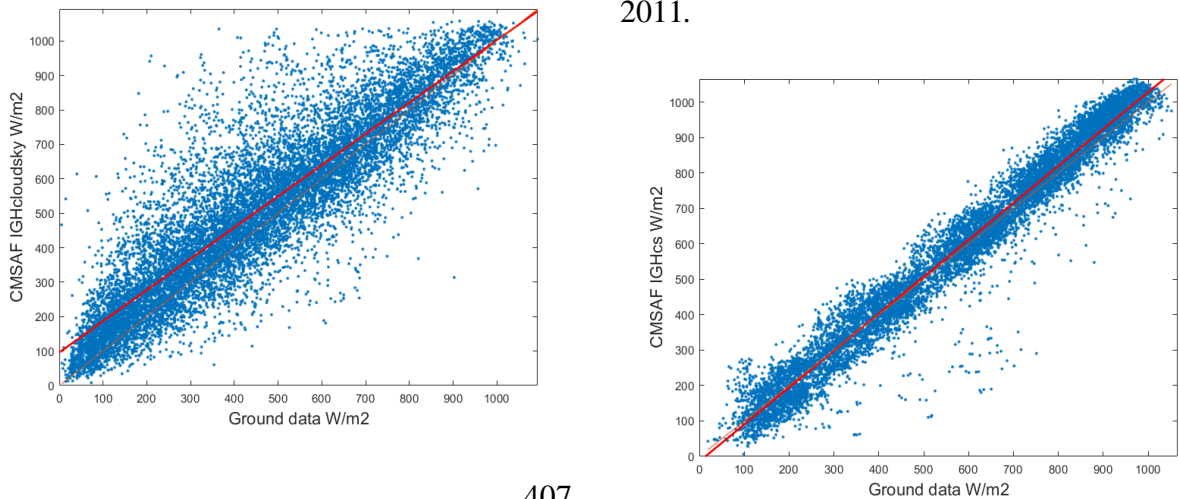
408
409 **Figure 5.-** SIS hourly datasets comparison between ground data and CMSAF data for all
410 ground stations in Cluster South, cloudy sky (left) and clear sky (right) for 2010 and 2011.

411   In cluster north, Fig. 4, CM SAF data presents an overestimation for low solar radiation
412 values and an underestimation for high radiation conditions in cloudy sky hours, while a bias
413 deviation from ideal linear fitting is clearly observable. On the other hand, for clear sky
414 conditions bias deviation is quite lower, only for high radiation conditions. For cluster south,
415 Fig. 5, comparison between CM SAF and ground data give much better results, as a
416 consequence of a major presence of cloud-free conditions. For cloudy hours, Fig. 5 shows an
417 overfitting mostly for low solar radiation values, while for clear sky hours deviation is almost
418 non-existent.

419
420   Errors between ground and satellite data also present different trends if we study datasets
421 divided by months. In Fig. 6, cluster north shows higher residual values for summer months
422 because of the stronger effect of Trade winds during this season. Indeed, it is recommendable
423 to estimate a site adaptation model for bias removal taking into account these differences. In
424 this paper, several correction factors were obtained and discussed for cluster north and south,
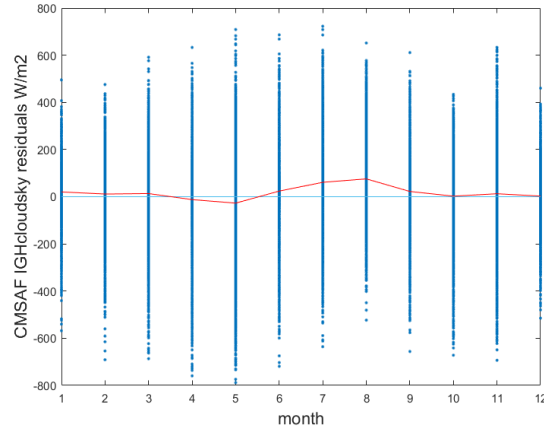425 cloudy and clear sky hours and for each season.

Figure 6.- SIS residuals between CM SAF data and ground data for cluster north and cloudy sky conditions by months.

Finally, to obtain the optimal site-adaption for both Canary Islands clusters we tested different linear fitting using all the separated datasets. For each group of data, the regression is estimated with the training dataset and then is validated with testing dataset. During this survey, we verified the following regressions to generate new satellite data:

- **Op-1.-** Clear sky linear regression using CM SAF original data and ground data.
- **Op-2.-** Substitute clear sky CM SAF for a clear sky model (McClear) and adapt this model using a linear regression with ground measurement.
- **Op-3.-** Cloudy sky linear regression using CM SAF original data and ground data.
- **Op-4.-** Cloudy sky linear regression using CM SAF original data and ground data for each season individually.
- **Op-5.-** All sky (both clear and cloudy skies) linear regression using CM SAF original data and ground data.

The best option for both clusters in Canary Islands is using the **Op-2** for clear sky days and **Op-4** for cloudy days. In Cluster north, where the difference between sky conditions in each season is more visible, the improvement for using **Op-4** instead of **Op-3** is higher. In the next section, we show the results with these two models.

4. Satellite data assessment results

For each station, we compared GHI ground measurement data with CM SAF GHI data and with McClear data (only in case of clear skies) in order to evaluate the accuracy of the models. In the same way, we compared Cluster North and South datasets.

4.1 Error metrics

In order to evaluate the performance of each method, we used two standard error metrics widely used in the solar forecasting community: the Root Mean Square Error (RMSE) and Mean Bias Error (MBE). Dividing both absolute error by the average of the hourly GHI data, we compute their relative metrics (% rRMSE and % rMBE).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(GHI_{forecast,i} - GHI_{measured,i}\right)^2} \tag{9}$$

13

$$MBE = \frac{1}{N}\left[\sum_{i=1}^{N}\left(GHI_{forecast,i} - GHI_{measured,i}\right)\right] \tag{10}$$

4.2 Clear sky days

Table 3 shows the results from the comparison of CM SAF satellite data and McClear model with ground data for clear sky conditions. Errors in terms of MBE and RMSE are presented for all the stations and for both clusters.

Station & Cluster	MBE		RMSE		%MBE		%RMSE	
	MC	CMSAF	MC	CMSAF	MC	CMSAF	MC	CMSAF
Gáldar (GAL)	**11.52**	-12.49	**30.51**	72.59	**2.19**	-2.37	**5.80**	13.80
San Mateo (SMA)	26.83	**-5.41**	**52.16**	65.67	4.42	**-0.89**	**8.59**	10.82
Vecindario (VEC)	**1.51**	7.42	**30.01**	60.04	**0.25**	1.24	**5.03**	10.07
Maspalomas (MASP)	**5.06**	-7.97	**30.43**	53.38	**0.87**	-1.37	**5.22**	9.16
Izaña (IZA)	**-4.50**	-200.64	**26.13**	289.26	**-0.65**	-28.89	**3.76**	41.65
Garimba (GAR)	11.30	**-9.54**	**49.66**	78.75	2.02	**-1.70**	**8.87**	14.06
Valle Guerra-Pajarillos (PAJ)	**3.37**	-9.06	**32.91**	65.90	**0.61**	-1.65	**5.98**	11.97
El Pico (PIC)	**-6.60**	-32.87	**36.19**	89.02	**-1.16**	-5.78	**6.37**	15.66
Puerto de la Cruz (PCR)	13.87	**3.37**	**34.42**	83.18	2.45	**0.59**	**6.08**	14.69
Guía de Isora (ISO)	17.47	**-3.55**	**42.03**	87.25	3.05	**-0.62**	**7.34**	15.23
La Fuente (LFU)	9.31	**-5.40**	**34.35**	74.76	1.69	**-0.98**	**6.24**	13.58
La Laguna – Güímar (LLA)	20.65	**13.02**	**36.69**	50.15	3.51	**2.21**	**6.23**	8.51
San Sebastián (SSB)	26.24	26.39	**43.13**	63.06	4.62	4.65	**7.60**	11.11
Hermigua (HER)	**6.85**	10.22	**38.11**	55.25	**1.11**	1.65	**6.15**	8.92
Barlovento (BAR)	26.99	**18.01**	**44.70**	56.44	5.14	**3.43**	**8.51**	10.74
Tazacorte II (TAZII)	28.18	**-20.55**	**54.56**	143.93	4.75	**-3.46**	**9.20**	24.26
Los Llanos de Aridane (ARI)	**28.93**	-39.85	**49.36**	121.56	**5.04**	-6.94	**8.60**	21.17
Fuencaliente (FUE)	24.16	**-1.79**	**55.79**	109.48	4.02	**-0.30**	**9.28**	18.22
Frontera (FRO)	30.25	**12.51**	**48.42**	84.85	5.49	**2.27**	**8.79**	15.41
Haría (HAR)	**5.29**	8.92	**30.76**	59.86	**0.95**	1.60	**5.51**	10.72
Tinajo (TIN)	20.74	20.17	**57.83**	72.15	3.84	3.73	**10.70**	13.35
Antigua (ANT)	**20.04**	23.91	**40.44**	53.51	**3.52**	4.20	**7.11**	9.41
CLUSTER NORTH	16.11	**-14.58**	**42.81**	97.04	2.86	**-2.59**	**7.61**	17.25
CLUSTER SOUTH	11.83	**10.50**	**35.19**	57.06	2.05	**1.82**	**6.08**	9.86

Table 3.- Errors between Satellite or Clear sky model and ground measurement for clear sky hours.

For ground stations in cluster south, we obtained errors below 11% rRMSE for clear sky conditions calculated with CM SAF, while McClear model provide errors between 5-7% rRMSE. Only (TIN) station, located in the border of both clusters, presents error over these limits. In terms of % rMBE, both models show very similar results with errors lower than 5%. Almost in all cases, both CM SAF and McClear overestimate comparing with ground data. Only (MASP) station presents underestimation with CM SAF data. For Cluster South dataset,

both models give around 2% rMBE, while in terms of % rRMSE McClear present better results than CM SAF, 6% and 10% respectively.

In case of cluster north, we obtained in general higher errors than in cluster north. In terms of % rRMSE, CM SAF estimates solar radiation for clear sky with errors between 10 and 18%, while McClear offers results between 5 and 10%. Only in (TAZII) and (ARI) CM SAF obtained errors higher than 21%. As it happens with cluster south, in terms of % rMBE both models obtained similar results, always below 5%. For Cluster North dataset, both models give between 2.5-3% rMBE, but CM SAF underestimate radiation and Mc Clear overestimate radiation compared with ground data. On the other hand, in terms of % rRMSE McClear present 7.5% errors results and CM SAF satellite errors are over 17%.

In particular, it is worth mentioning the results obtained in Izaña station (IZA). In terms of both % rMBE and % rRMSE, CM SAF offers higher errors than McClear. CM SAF errors are around 29% and 41% in terms of % rMBE and % rRMSE respectively, while McClear results are both below 4%. (IZA) station is located over 2000 meters high in Tenerife Island and McClear fits ground measurement. This station is not included in any cluster because of its particular conditions.

4.3 Rest of the days

As for clear sky conditions, cloudy hours present different results for each cluster. In cluster north, CM SAF offers results over 30 or 40% rRMSE for cloudy skies at almost all locations. In terms of % rMBE, CM SAF radiation errors compared with ground stations state between 10 and 30% both over and underestimate. For cluster north dataset, CM SAF provides errors of 4% rMBE and 38% rRMSE for cloudy sky conditions, while for all sky it conditions obtains 2% rMBE and 32% rRMSE.

On the other hand, in cluster south for cloudy sky conditions CM, SAF obtains results below 25% in terms of RMSE at mostly locations. In terms of % rMBE, for all ground stations in cluster south CM SAF overestimate ground measurements. Most of the stations present results below 12% rMBE. Taking into account whole clusters' datasets, it is observable that CM SAF obtains better results for cluster south than for cluster north in terms of % rRMSE, while in terms of % rMBE cluster north presents better results. That is because in cluster north MBE presents both over and underestimate results, so when all stations are used for the whole cluster we obtained lower errors. Individually, errors for each station are higher than in cluster south.

For cluster south dataset, CM SAF provides errors of 12% rMBE and 28% rRMSE for cloudy sky conditions, while for all sky conditions it obtains 7% rMBE and 20% rRMSE. As for clear sky conditions, (IZA) station present higher errors because of its altitude. In terms of MBE, CM SAF underestimates solar radiation both for cloudy sky conditions and for all sky conditions, with errors of -22% and 27% respectively. Furthermore, for RMSE, CM SAF offers worse results at this location, with errors over 45% both for cloudy sky and all sky conditions. (PAJ), (PIC) and (GAR) stations are located almost in the same latitude and longitude but with different altitudes. It is observable that errors increase for cloudy and all sky conditions with altitude, both for MBE and RMSE.

Station & Cluster	CLOUD SKY				ALL DATA (cloud & clear)			
	MBE	RMSE	%MBE	%RMSE	MBE	RMSE	%MBE	%RMSE

Gáldar (GAL)	-28.75	121.88	-6.72	28.48	-24.25	110.47	-5.33	24.28
San Mateo (SMA)	57.69	165.13	15.50	44.37	37.40	141.01	8.35	31.50
Vecindario (VEC)	28.62	106.24	6.15	22.82	16.95	84.03	3.15	15.63
Maspalomas (MASP)	15.55	98.95	3.20	20.37	1.91	75.93	0.35	14.01
Izaña (IZA)	-108.48	239.64	-22.55	49.82	-174.61	276.15	-27.53	43.54
Garimba (GAR)	84.20	167.14	23.12	45.90	71.86	158.35	18.43	40.61
Valle Guerra-Pajarillos (PAJ)	3.08	118.63	0.75	28.89	0.24	108.63	0.05	24.50
El Pico (PIC)	20.17	172.37	4.89	41.77	8.56	157.94	1.92	35.35
Puerto de la Cruz (PCR)	25.59	151.07	7.31	43.14	21.52	141.10	5.52	36.20
Guía de Isora (ISO)	-72.94	174.01	-15.36	36.63	-43.93	144.23	-8.51	27.95
La Fuente (LFU)	-25.94	132.93	-6.63	33.97	-22.34	124.72	-5.33	29.75
La Laguna – Güímar (LLA)	82.45	151.02	21.03	38.52	49.58	114.88	10.22	23.67
San Sebastián (SSB)	111.18	177.36	25.86	41.26	76.43	142.10	15.72	29.22
Hermigua (HER)	98.67	167.36	28.00	47.49	88.64	158.68	23.16	41.46
Barlovento (BAR)	71.75	132.59	20.49	37.86	65.31	125.92	17.59	33.92
Tazacorte II (TAZII)	-6.43	176.18	-1.49	40.74	-10.85	166.76	-2.25	34.54
Los Llanos de Aridane (ARI)	-54.62	174.03	-12.44	39.65	-48.80	155.50	-9.92	31.59
Fuencaliente (FUE)	51.01	174.01	12.07	41.17	44.72	167.63	10.07	37.76
Frontera (FRO)	43.55	159.16	10.49	38.35	37.94	148.52	8.63	33.79
Haría (HAR)	36.17	101.28	8.12	22.73	28.17	91.10	5.88	19.03
Tinajo (TIN)	90.33	162.23	21.82	39.18	78.20	150.56	17.94	34.54
Antigua (ANT)	52.95	113.63	11.77	25.26	40.32	92.41	8.04	18.43
CLUSTER NORTH	16.40	152.82	4.12	38.35	9.25	141.90	2.12	32.52
CLUSTER SOUTH	54.14	127.00	12.10	28.39	34.23	101.28	6.75	19.97

524 **Table 4.-** Errors between Satellite model and ground measurement for cloudy sky hours and
525 all hours.

526

527 ## 5. Site-adaptation results

528 Site-adaptation's best option for both clusters in Canary Islands is using linear regression
529 for McClear model in case of clear sky days, Op-2, and linear regression for each season
530 individually for cloudy days, Op-4. Tables 5 and 6 show errors' results for both clusters
531 testing dataset in terms of RMSE and MBE in clear sky, cloudy sky and all sky conditions.

532

533 In general, site-adaptation proposed in this paper obtained non-bias results for clear sky,
534 cloudy sky and all sky conditions in terms of MBE for cluster north. In terms of RMSE, site-
535 adaptation offers 7% (39.51 W/m^2) results for clear skies, 41% (165.47 W/m^2) for cloudy
536 skies and a mean of 33% (146.67 W/m^2) for the whole cluster north dataset. Linear site-
537 adaptation reduce almost to zero MBE errors compared with initial CM SAF for clear, cloudy
538 and all sky conditions, while in terms of RMSE, the proposed model improves only results in
539 clear sky condition. For cloudy and all sky conditions, site-adaptation reduces bias but
540 slightly increases RMSE results. Fig. 7 shows a bias removal compared with the previous
541 figure for the same cluster, both in clear and cloud sky conditions.

542

543 Cluster south results have also been improved using site-adaptation proposed in this paper.
544 In terms of MBE, new dataset general bias has disappeared compared with previous CM SAF
545 results, with errors below 1% rMBE. Furthermore, unlike cluster north, in terms of RMSE

546 site-adaptation also improves initial errors. CM SAF corrected dataset obtains 5% (31.33
547 W/m^2) results for clear skies, 27% (119.46 W/m^2) for cloudy skies and a mean of 18%
548 (92.23W/m^2) for the whole cluster north dataset. Fig. 8 shows a bias removal and a lower
549 dispersion compared with the previous figure for the same cluster both in clear and cloud sky
550 conditions.
551
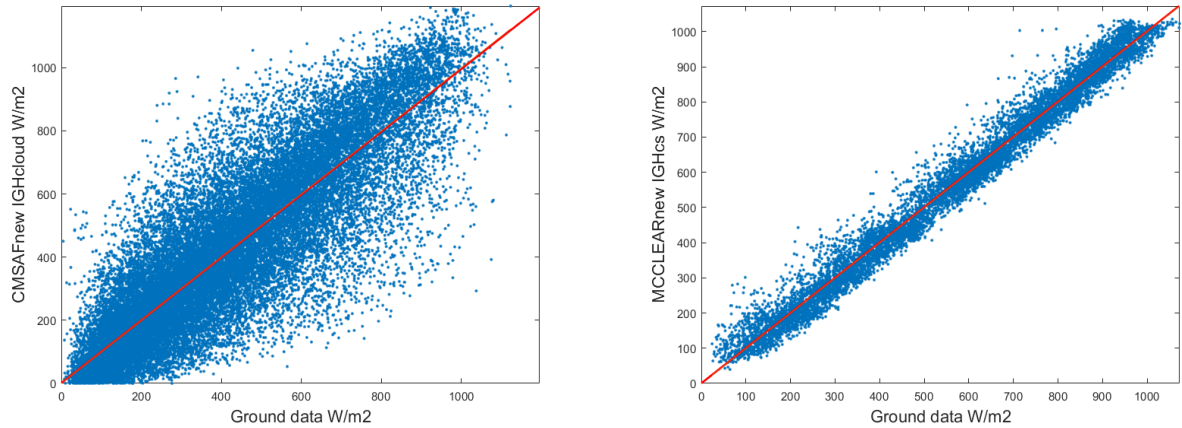
553

554 **Figure 7.-** SIS hourly datasets comparison between ground data and CMSAF new dataset for
555 all ground stations in Cluster North, cloudy
556 sky (left) and clear sky (right) for 2010 and
2011.
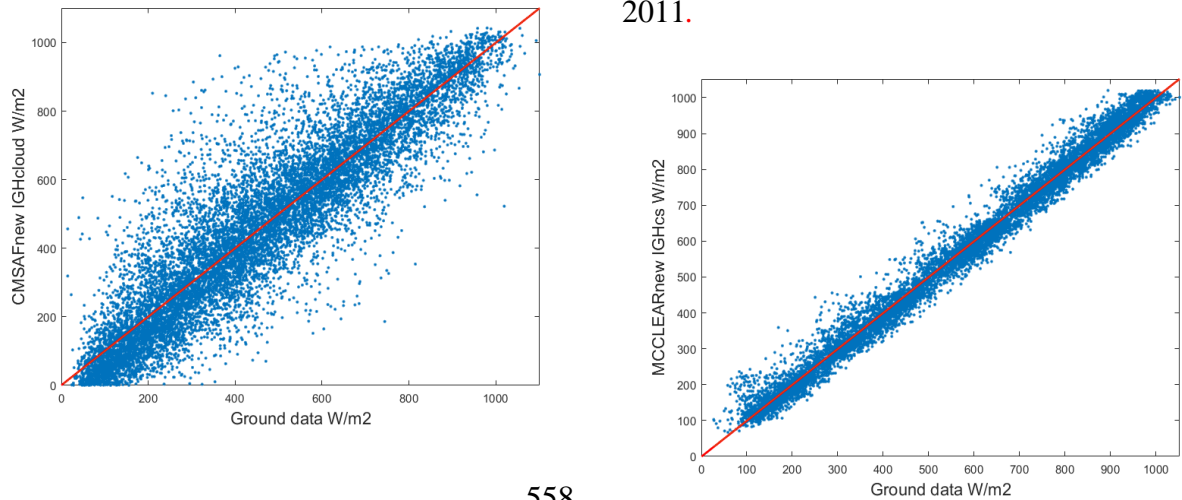
558

559 **Figure 8.-** SIS hourly datasets comparison between ground data and CMSAF new dataset for
560 all ground stations in Cluster South, cloudy sky (left) and clear sky (right) for 2010 and 2011.

561

562

563

Station & Cluster	CLEAR SKY			
	MBE	RMSE	%MBE	%RMSE
CLUSTER NORTH (training)	0.98	39.79	0.17	7.02
CLUSTER SOUTH (training)	-0.13	31.75	-0.02	5.49

CLUSTER NORTH (testing)	0.80	39.51	0.14	7.10
CLUSTER SOUTH (testing)	0.14	31.33	0.02	5.41

Table 5.- Site adaptation final results for clear sky hours, both for training and testing datasets.

Station & Cluster	CLOUD SKY				ALL DATA (cloud & clear)			
	MBE	**RMSE**	**%MBE**	**%RMSE**	**MBE**	**RMSE**	**%MBE**	**%RMSE**
CLUSTER NORTH (training)	0.43	165.82	0.11	41.62	0.99	146.45	0.23	33.51
CLUSTER SOUTH (training)	1.12	122.18	0.25	27.33	0.23	91.51	0.05	18.08
CLUSTER NORTH (testing)	0.52	165.47	0.13	41.52	-0.27	146.67	-0.06	33.68
CLUSTER SOUTH (testing)	1.60	119.46	0.36	26.68	1.46	92.23	0.29	18.13

Table 6.- Site adaptation final results for cloudy and all hours, both for training and testing datasets.

6. Conclusions

This work studies CM SAF and McClear global solar radiation assessment in several stations in Canary Islands. This survey used clear and cloudy sky conditions separately in order to establish the best option in all conditions.

CM SAF data reproduce quite satisfactorily GHI results, even taking into account complex local climatological variations in Canary Islands. However, some areas and locations show improvable bias and errors. Northern areas present a higher occurrence of clouds especially in the summer months when Trade Winds meet with high mountains. Canary Islands satellite gridded information was separated in two clusters, South and North, using GHI datasets for 2010 and 2011 years. Cluster datasets include all ground GHI values for measurement stations belonging to each cluster. In this way, CM SAF and McClear data accuracy in comparison with ground data were studied for each ground station individually and using both cluster datasets.

For clear sky conditions, McClear model obtained better results for both clusters in terms of % rRMSE, while in terms of % MBE both models offer similar results. Cluster south presents errors between 5-7%, while stations in cluster north errors are situated between 5-10%. Normally, McClear and CM SAF clear sky data tend to overestimate GHI data compared with ground stations. For cloudy skies, CM SAF GHI data offer higher errors for both clusters. In terms of % rRMSE, cluster north presents errors normally over 30%, while for cluster south general results show errors below 25%. Cloud presence in cluster north, both generated or coming from trade winds direction, provoke a lower accuracy in GHI satellite radiation. In terms of % MBE, cluster south results provide always positive errors due to an overestimation of satellite data. While in cluster north for cloudy sky conditions we obtained both positive and negative MBE results depending on the measurement station. Thus, combining both McClear for clear skies and CM SAF for rest of the days offers good results.

Finally, this paper proposes a site adaptation linear correction of satellite derived data to obtain a bias removal. Different site adaptation regression was studied, taking into account the cluster, clear or cloudy sky conditions and season of the year. The optimal adaption proposed

601 in this paper is using linear regression for McClear model in the case of clear sky days, Op-2,
602 and linear regression for each season individually for cloudy days, Op-4. For cluster north,
603 clear skies give errors below 0.2% rMBE and 7% rRMSE, improving initial results in 2% and
604 10% respectively. In case of cloudy skies, new satellite dataset improves results in terms of %
605 rMBE, almost 4%, but % rRMSE initial results are 3% lower. For all sky conditions, site
606 adaptation provide a bias removal in terms of MBE and similar errors in terms of RMSE. In
607 cluster south, the linear regression proposed in this paper obtained better results for both clear
608 and cloudy sky conditions in terms of MBE and RMBE. For clear sky conditions, site
609 adaptation improves errors in 2% rMBE and 5% rRMSE, while for cloudy sky conditions
610 errors are almost 12% rMBE and 1.5% rRMSE lower than initial CM SAF data. Furthermore,
611 individually at each location, it is possible to get a bias removal using a linear site adaptation.
612
613 Both CM SAF and McClear models provide GHI data with great spatial and temporal
614 resolution for Canary Islands and offer quite satisfactory accurate data. At locations and areas
615 where satellite models provide data with an important bias or errors, these site adaptation
616 techniques could improve final uncertainty. These data could improve forecasting and
617 downscaling models as a means of obtaining better knowledge on solar radiation in these
618 islands.
619

633 ## 8. References

634 [1] F. Díaz, G. Montero, and L. Mazorra-Aguiar, "Wind Field and Solar Radiation
635 Characterization and Forecasting. Solar radiation maps," in *Green Energy and
636 Technology*, 2018, pp. 229–257.
637 [2] F. Díaz, G. Montero, J. M. Escobar, E. Rodríguez, and R. Montenegro, "An adaptive
638 solar radiation numerical model," *J. Comput. Appl. Math.*, vol. 236, no. 18, pp. 4611–
639 4622, Dec. 2012.
640 [3] M. Ŝúri and J. Hofierka, "A new GIS-based solar radiation model and its application to
641 photovoltaic assessments," *Trans. GIS*, vol. 8, no. 2, pp. 175–190, Apr. 2004.
642 [4] A. Zelenka, R. Perez, R. Seals, and D. Renné, "Effective Accuracy of Satellite-Derived
643 Hourly Irradiances," *Theor. Appl. Climatol.*, vol. 62, no. 3–4, pp. 199–207, Apr. 1999.
644 [5] K. G. Karlsson *et al.*, "CLARA-A2: The second edition of the CM SAF cloud and
645 radiation data record from 34 years of global AVHRR data," *Atmos. Chem. Phys.*, vol.
646 17, no. 9, pp. 5809–5828, 2017.
647 [6] P. Minnis *et al.*, "CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and
648 Terra and Aqua MODIS Data—Part I: Algorithms," *IEEE Trans. Geosci. Remote
649 Sens.*, vol. 49, no. 11, pp. 4374–4400, Nov. 2011.

650 [7]   H. G. Beyer, C. Costanzo, and D. Heinemann, "Modifications of the Heliosat
651   procedure for irradiance estimates from satellite images," *Sol. Energy*, vol. 56, no. 3,
652   pp. 207–212, Mar. 1996.
653 [8]   C. Rigollier, M. Lefèvre, and L. Wald, "The method Heliosat-2 for deriving shortwave
654   solar radiation from satellite images," *Sol. Energy*, vol. 77, no. 2, pp. 159–169, Jan.
655   2004.
656 [9]   R. W. Mueller *et al.*, "Rethinking satellite-based solar irradiance modelling: The
657   SOLIS clear-sky module," *Remote Sens. Environ.*, vol. 91, no. 2, pp. 160–174, May
658   2004.
659 [10] K.-F. Dagestad and J. A. Olseth, "A modified algorithm for calculating the cloud
660   index," *Sol. Energy*, vol. 81, no. 2, pp. 280–289, Feb. 2007.
661 [11] K.-F. Dagestad, "Mean bias deviation of the Heliosat algorithm for varying cloud
662   properties and sun-ground-satellite geometry," *Theor. Appl. Climatol.*, vol. 79, no. 3–4,
663   pp. 215–224, Dec. 2004.
664 [12] R. Müller, U. Pfeifroth, and C. Träger-Chatterjee, "Surface Solar Radiation Data Set-
665   Heliosat (SARAH)-Edition 1. Satellite Application Facility on Climate Monitoring,"
666   2015.
667 [13] U. Pfeifroth, S. Kothe, R. Müller, and J. Trentmann, "Surface Radiation Data Set–
668   Heliosat (SARAH)–Edition 2, Satellite Application Facility on Climate Monitoring,"
669   2017.
670 [14] R. H. Acker *et al.*, "EUMETSAT Satellite Application Facility on Climate Monitoring
671   CM SAF Cloud , Albedo , Radiation dataset , Surface Albedo Product User Manual,"
672   *Sol. Energy*, vol. 35, no. 1, pp. 1–24, 2010.
673 [15] Y. Eissa, M. Chiesa, and H. Ghedira, "Assessment and recalibration of the Heliosat-2
674   method in global horizontal irradiance modeling over the desert environment of the
675   UAE," *Sol. Energy*, vol. 86, no. 6, pp. 1816–1825, 2012.
676 [16] Y. Eissa *et al.*, "Validation of the surface downwelling solar irradiance estimates of the
677   HelioClim-3 database in Egypt," *Remote Sens.*, vol. 7, no. 7, pp. 9269–9291, 2015.
678 [17] Y. Eissa *et al.*, "Validating surface downwelling solar irradiances estimated by the
679   McClear model under cloud-free skies in the United Arab Emirates," *Sol. Energy*, vol.
680   114, pp. 17–31, 2015.
681 [18] P. Ineichen, "Long term satellite global, beam and diffuse irradiance validation,"
682   *Energy Procedia*, vol. 48, pp. 1586–1596, 2014.
683 [19] F. Antonanzas-Torres, F. Cañizares, and O. Perpiñán, "Comparative assessment of
684   global irradiation from a satellite estimate model (CM SAF) and on-ground
685   measurements (SIAR): A Spanish case study," *Renew. Sustain. Energy Rev.*, vol. 21,
686   pp. 248–261, 2013.
687 [20] J. Vindel, A. Navarro, R. Valenzuela, L. R.-A. Research, and undefined 2016,
688   "Temporal scaling analysis of irradiance estimated from daily satellite data and
689   numerical modelling," *Elsevier*.
690 [21] S. M. Uppala *et al.*, "The ERA-40 re-analysis," *Q. J. R. Meteorol. Soc.*, vol. 131, no.
691   612, pp. 2961–3012, Oct. 2005.
692 [22] J. Bojanowski, A. Vrieling, A. S.-S. Energy, and undefined 2014, "A comparison of
693   data sources for creating a long-term time series of daily gridded solar radiation for
694   Europe," *Elsevier*.
695 [23] J. Polo *et al.*, "Integration of ground measurements to model derived data," no.
696   October, 2015.
697 [24] T. Cebecauer and M. Suri, "Accuracy improvements of satellite-derived solar resource
698   based on GEMS re-analysis aerosols," *SolarPACES*, vol. 2009, pp. 1–4, 2010.
699 [25] J. M. Vindel, J. Polo, and F. Antonanzas-Torres, "Improving daily output of global to

direct solar irradiance models with ground measurements," *J. Renew. Sustain. Energy*, vol. 5, no. 6, p. 063123, Nov. 2013.

[26] J. Polo, L. Martín, and J. M. Vindel, "Correcting satellite derived DNI with systematic and seasonal deviations: Application to India," *Renew. Energy*, vol. 80, pp. 238–243, 2015.

[27] T. Mieslinger, F. Ament, K. Chhatbar, and R. Meyer, "A new method for fusion of measured and model-derived solar radiation time-series," *Energy Procedia*, vol. 48, pp. 1617–1626, 2014.

[28] K. Schumann, H. G. Beyer, K. Chhatbar, and R. Meyer, "Improving Satellite-Derived Solar Resource Analysis with Parallel Ground-Based Measurements," *ISES Sol. World Congr.*, 2011.

[29] G. Bender, F. Davidson, F. Eichelberger, and C. A. Gueymard, "the Road To Bankability: Improving Assessments for More Accurate Financial Planning ," *Sol. 2011 Conf.American Sol. Energy Soc.*, no. May, 2011.

[30] C. Vernay, S. Pitaval, and P. Blanc, "Calibration of long-term global horizontal irradiation estimated by HelioClim-3 through short-term local measurement campaigns: extending of the results to European and African sites," *World Renew. Energy Forum*, no. May 2012, pp. 1943–1950, 2014.

[31] a. Roesch, M. Wild, a. Ohmura, E. G. Dutton, C. N. Long, and T. Zhang, "Assessment of BSRN radiation records for the computation of monthly means," *Atmos. Meas. Tech.*, vol. 4, no. 2, pp. 339–354, 2011.

[32] I. Zahumenský, "Guidelines on quality control procedures for data from automatic weather stations," *World Meteorol. Organ.*, no. 955, pp. 2–6, 2004.

[33] B. N. Holben *et al.*, "AERONET - A federated instrument network and data archive for aerosol characterization," *Remote Sens. Environ.*, vol. 66, pp. 1–16, 1998.

[34] Government of Canada, "World Ozone Monitoring and Mapping." [Online]. Available: http://es-ee.tor.ec.gc.ca/e/ozone/ozoneworld.htm. [Accessed: 18-Jun-2018].

[35] MINES ParisTech / Transvalor, "Web Services - www.soda-pro.com." [Online]. Available: http://www.soda-pro.com/es/web-services#radiation. [Accessed: 18-Jun-2018].

[36] NASA, "MERRA-2." [Online]. Available: https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. [Accessed: 18-Jun-2018].

[37] P. Ineichen, "Validation of models that estimate the clear sky global and beam solar irradiance," *Sol. Energy*, vol. 132, pp. 332–344, 2016.

[38] M. J. Reno, C. W. Hansen, and J. S. Stein, "Global horizontal irradiance clear sky models: Implementation and analysis," *SANDIA Rep. SAND2012-2389*, 2012.

[39] R. Bird and R. Hulstrom, "Simplified clear sky model for direct and diffuse insolation on horizontal surfaces," 1981.

[40] C. A. Gueymard, "REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation - Validation with a benchmark dataset," *Sol. Energy*, vol. 82, no. 3, pp. 272–285, 2008.

[41] M. Lefèvre *et al.*, "McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmos. Meas. Tech.*, vol. 6, no. 9, pp. 2403–2418, 2013.

[42] P. Ineichen, "Comparison of eight clear sky broadband models against 16 independent data banks," *Sol. Energy*, vol. 80, no. 4, pp. 468–478, 2006.

[43] J. Petruccelli, B. Nandram, and M. Chen, "Applied statistics for engineers and scientists," 1999.

[44] J. Polo, L. F. Zarzalejo, P. Salvador, and L. Ramírez, "Angstrom turbidity and ozone column estimations from spectral solar irradiance in a semi-desertic environment in

Spain," *Sol. Energy*, vol. 83, no. 2, pp. 257–263, 2009.

[45] M. J. Reno and C. W. Hansen, "Identification of periods of clear sky irradiance in time series of GHI measurements," *Renew. Energy*, vol. 90, pp. 520–531, 2016.

[46] M. Anderberg, "Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks," 2014.

[47] A. Fausett and M. E. Celebi, "An Accelerated Nearest Neighbor Search Method for the K-Means Clustering Algorithm.," in *Florida Artificial Intelligence Research Society Conference; The Twenty-Sixth International FLAIRS Conference*, 2013.

[48] A. Porfirio and J. Ceballos, "A method for estimating direct normal irradiation from GOES geostationary satellite imagery: Validation and application over Northeast Brazil," *Sol. Energy*, vol. 155, pp. 178–190, 2017.

[49] H. Beyer and F. Habyarimana, "Assessment of the quality of satellite derived irradiance data for Rwanda by comparison with data from a fleet of automated ground stations," *researchgate.net*.

[50] W. Tang, J. Qin, K. Yang, S. Liu, N. Lu, and X. Niu, "Retrieving high-resolution surface solar radiation with cloud parameters derived by combining MODIS and MTSAT data," *Atmos. Chem. Phys.*, vol. 16, no. 4, pp. 2543–2557, Mar. 2016.

[51] R. Perez, J. Schlemmer, K. Hemker, S. Kivalov, A. Kankiewicz, and C. Gueymard, "Satellite-to-irradiance modeling - a new version of the SUNY model," in *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*, 2015, pp. 1–7.

[52] A. Riihelä, T. Carlund, J. Trentmann, R. Müller, and A. Lindfors, "Validation of CM SAF Surface Solar Radiation Datasets over Finland and Sweden," *Remote Sens.*, vol. 7, no. 6, pp. 6663–6682, May 2015.

[53] P. Ineichen, "Satellite Irradiance Based on MACC Aerosols: Helioclim 4 and SolarGIS, Global and Beam Components Validation," 2014.

[54] J. Polo, F. Antonanzas-Torres, J. M. Vindel, and L. Ramirez, "Sensitivity of satellite-based methods for deriving solar radiation to different choice of aerosol input and models," *Renew. Energy*, vol. 68, pp. 785–792, Aug. 2014.

[55] J. Polo, J. M. Vindel, and L. Martín, "Angular dependence of the albedo estimated in models for solar radiation derived from geostationary satellites," *Sol. Energy*, vol. 93, pp. 256–266, Jul. 2013.

[56] J. Polo *et al.*, "Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets," *Sol. Energy*, vol. 132, pp. 25–37, 2016.