applied sciences

MDPI

Article

A Comprehensive Methodology for the Statistical Characterization of Solar Irradiation: Application to the Case of Morocco

Naoufal Bouhorma, Helena Martín *, Jordi de la Hoz and Sergio Coronas

Department of Electrical Engineering, Barcelona East School of Engineering (EEBE), Universitat Politècnica de Catalunya (UPC), C. Eduard Maristany, 10, 08019 Barcelona, Spain

* Correspondence: m.helena.martin@upc.edu

Featured Application: Determination and validation of solar irradiation data for the techno-economic valuation of solar projects.

Abstract: The prediction and characterization of solar irradiation relies mostly on either the use of complex models or on complicated mathematical techniques, such as artificial neural network (ANN)-based algorithms. This mathematical complexity might hamper their use by businesses and project developers when assessing the solar resource. In this study, a simple but comprehensive methodology for characterizing the solar resource for a project is presented. It is based on the determination of the best probability distribution function (PDF) of the solar irradiation for a specific location, assuming that the knowledge of statistical techniques may be more widely extended than other more complex mathematical methods. The presented methodology was tested on 23 cities across Morocco, given the high interest in solar investments in the country. As a result, a new database for solar irradiation values depending on historical data is provided for Morocco. The results show the great existing variety of PDFs for the solar irradiation data at the different months and cities, which demonstrates the need for undertaking a proper characterization of the irradiation when the assessment of solar energy projects is involved. When it is simply needed to embed the radiation uncertainty in the analysis, as is the case of the techno-economic valuation of solar energy assets, the presented methodology can reach this objective with much less complexity and less demanding input data. Moreover, its application is not limited to solar resource assessment, but can also be easily used in other fields, such as meteorology and climate change studies.

Keywords: solar energy; irradiation; satellite data; predictive models; probability distribution function; Morocco

Citation: Bouhorma, N.; Martín, H.; de la Hoz, J.; Coronas, S. A Comprehensive Methodology for the Statistical Characterization of Solar Irradiation: Application to the Case of Morocco. *Appl. Sci.* **2023**, *13*, 3365. https://doi.org/10.3390/app13053365

Academic Editors: Weidong Huang and Francesco Calise

Received: 5 October 2022
Revised: 23 February 2023
Accepted: 2 March 2023
Published: 6 March 2023

1. Introduction

Nowadays the world is heading towards an energetic transition phase where the main goal is to reduce the dependency on carbon-based sources of energy and increment the integration of new technologies that can utilize renewable sources as the main fuel to produce usable energy. A high interest is given to solar energy technologies, since it is a power source that is not limited to a few regions around the world. However, solar irradiation available on the Earth surface on solar power premises is not easily manageable due to the unpredictability of weather conditions, such as cloudiness and humidity.

One of the regions where solar energy is playing an important role in the energy sector is the MENA (Middle East and North Africa) region. The Middle East Solar Industry Association (MESIA) has reported that in the MENA region energy investments are going to reach USD 1 trillion in the period 2019–2023 [1]. Driven by pure market forces, many MENA regions have stated their intentions to increase renewable energy (RE) production, as these technologies have become cheaper than burning fossil fuels. Another reason for further augmenting the RE production in the region is the expected growth in the energy demand, along with the depletion of the fossil fuel reserves. In this line, due to the water scarcity issue that these countries experience, energy-intensive desalination operations are increasingly being adopted [2].

Due to their geographical location, the MENA countries have a huge potential in terms of solar energy projects, as solar energy is quite abundant in the region. As such, multiple countries in the region devised national energy plans, where an increase in solar energy production is a vital part of them. Concentrated solar power (CSP) capacity, for instance, has experienced a considerable growth in the MENA region. Some of the highlighted countries where CSP deployment has increased over the last years are Algeria, Egypt, Israel, Kuwait, Morocco, Saudi Arabia and the United Arab Emirates (UAE) (see Figure 1).

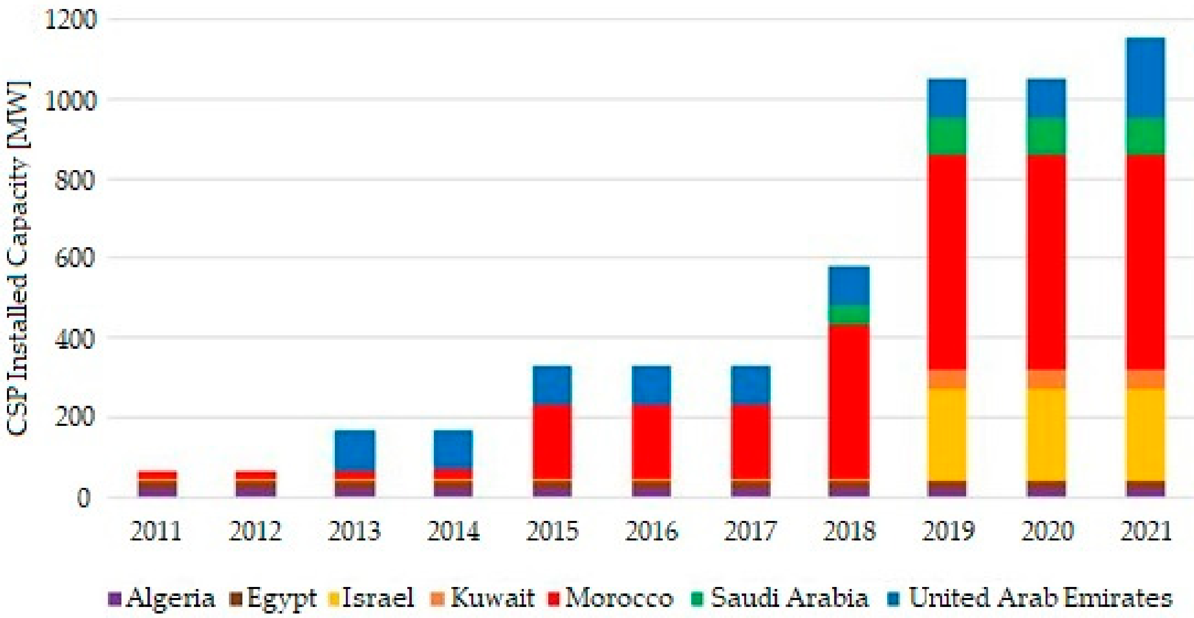

Figure 1. CSP capacity installed in the MENA region. Source: own elaboration based on [3].

Following this momentum, future plans about increasing solar investment are becoming more popular. Egypt, for instance, established the goal of increasing renewable production up to 20% by 2022. The country also expects a national capacity of solar energy of 3.5 GW by the year 2027. Saudi Arabia devised a plan in which by the year 2023, 9.5 GW of the country's energy production will come from clean energy. The United Arab Emirates (UAE) set an ambitious goal of 1 GW of solar capacity by 2020 and 5 GW by 2030.

In this regard, the Kingdom of Morocco stated, through the National Solar Plan in 2009, the goal of reaching 2 GW of installed capacity by 2020. Recently, this goal was updated to 4 GW by 2030. According to it, it is expected that up to 52% of electricity production will be based on renewable technology. As a result, Morocco is considered to be an important actor concerning solar energy development in the future of the MENA region and, consequently, it has been chosen as an object of study.

The Kingdom of Morocco enjoys a considerable solar energy resource due to its latitude. For a successful implementation of these solar projects, the use of reliable solar irradiation data is essential. In this regard, determination and validation of the irradiation is mandatory for the techno-economic valuation of solar projects—current [4–6] and future—since working on inaccurate data may lead to significant deviations between the predicted and the actually achieved results.

1.1. State of the Art

The scientific literature about solar irradiation is widely extended and covers all the relevant aspects, from those studies focusing on the determination and forecasting of the global horizontal irradiation (GHI) as well as the beam or direct normal irradiation (DNI) on a horizontal surface to those on a tilted surface, at different time frames (see Figure 2).

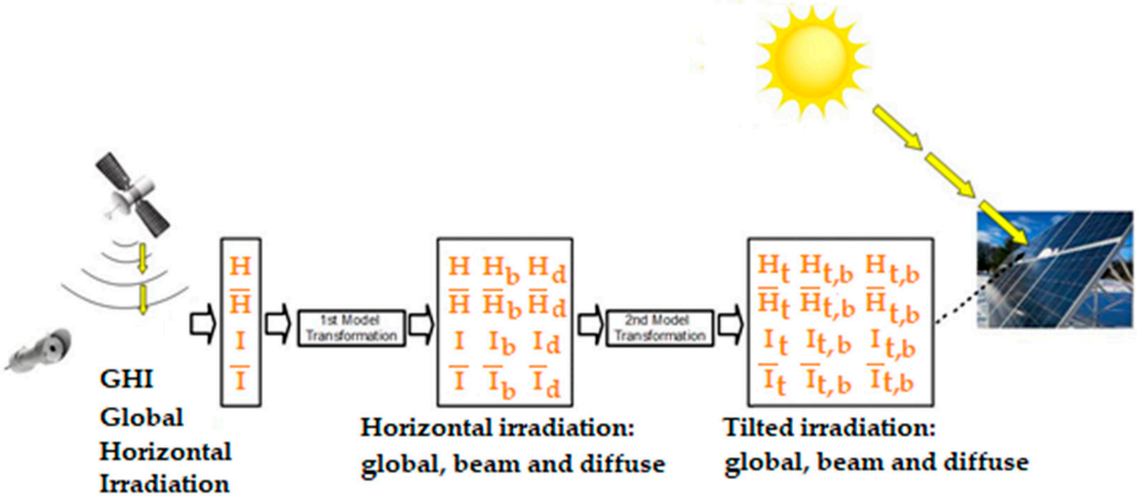

Figure 2. Characterization of solar irradiation in the scientific literature, dealing with either daily (H), average daily ($\overline{H}$), hourly (I) or average hourly ($\bar{I}$) global, beam ($_b$) and diffuse ($_d$) values. Source: own elaboration.

The existing literature aimed at forecasting solar irradiation and/or irradiance is quite extensive. Likewise, so are the different methods used to carry it out. As extracted from [7], which presents an in-depth analysis of the performance of different solar irradiance forecasting models, the predominant forecasted variable in most studies is the GHI, followed by the DNI. Moreover, the most frequent forecasting horizon is the short term, i.e., intra-hour, intra-day or, to a lesser degree, day-ahead predictions. This same trend is observed in [8], where an overview focused on solar irradiation forecasting methods using machine learning approaches is given.

In this context, some references can be highlighted dealing with short-term solar irradiance forecasting [9–35] and, to a lesser extent, some focused on monthly solar predictions [36–41].

Intra-hour GHI forecast employing a cloud retrieval technique to develop a physics-based smart persistence model is improved in [9], and an algorithm using cloud physical properties for intra-day GHI and DNI forecasting with time horizons of 0–4 h at a 15 min temporal resolution is developed in [10]. A GHI forecasting model based on satellite data from Finland with a forecast horizon of 4 h and a 15 min temporary resolution is developed and validated in [11], while the error obtained from the Japan meteorological agency mesoscale model in the hourly-averaged GHI forecasts from 2008 to 2012 is assessed in [12]. The predictions of 15 clear-sky irradiance models by comparison with the RRTMG physical radiative transfer model, acting as a benchmarking reference, for hourly GHI and DNI over a whole year are evaluated in [13], while the differences induced in the hourly and daily GHI predictions by the mesoscale atmospheric weather research forecasting model in Greece when using different shortwave radiation are assessed in [14]. Likewise, hourly GHI in the Arabian Peninsula using a three-dimensional meteorology–chemistry model including a state-of-the-art prognostic treatment of aerosols is simulated in [15]. In another prediction methodology vision, the performance of an exponential smoothing model with decomposition methods to improve its hourly GHI forecasting accuracy and computational efficiency is analyzed in [16], while prediction intervals for DNI estimates from GHI observations at the minute scale over the Korean Peninsula using the Engerer model under a probabilistic approach are calculated in [17]. Furthermore, a hybrid convolutional neural network–long short-term memory model with spatiotemporal correlations to improve the accuracy of short-term GHI prediction for ensuring the optimum utilization of photovoltaic power generation sources is proposed in [18].

In turn, the precision of short-term forecasts for GHI and DNI of a global numerical weather prediction (NWP) model in Portugal is evaluated in [19], while the performance of three NWP models in forecasting daily GHI for Australia is assessed in [20]. In another forecasting methodology approach, the use of different machine learning techniques for deterministic and probabilistic GHI and DNI forecasts using local irradiance data and sky images with forecasting time horizons ranging from 5 min up to 30 min is evaluated in [21]. Likewise, the integration of different forecasting models, by means of machine learning techniques, to improve the short-term predictions of GHI and DNI with a forecast horizon of 6 h and a temporary resolution of 15 min in the Iberian Peninsula is studied in [22], while a benchmarking of different machine learning techniques for intra-day GHI forecasting from 1 h to 6 h ahead in an insular context is proposed in [23]. An artificial neural network (ANN)-based algorithm to improve GHI forecasts obtained from the NWP model of the European Centre for Medium-range Weather Forecasts (ECMWF) with a time horizon of 72 h and a time-step of 30 min is developed in [24]. Similarly, ANN models to produce hourly GHI forecasts from 1 to 6 h ahead are designed in [25,26], using exogenous (satellite and NWP model of the ECMWF) and ground data, and ground measurement and satellite data, respectively. In the same way, six ANN models to estimate the monthly mean daily GHI in different locations of the UAE are developed in [27]. A nonparametric method, based on k-means algorithm, for ultra-short-term forecasting of GHI to deliver predictions with a forecast horizon from 500 ms to 5 min under a probabilistic perspective is assessed in [28]. Similarly, a forecast methodology based on the k-nearest neighbours algorithm for intra-hour GHI and DNI with horizons ranging from 5 min up to 30 min using ground telemetry and sky images is proposed in [29]. In turn, a Gaussian process regression method for GHI forecast horizons from 30 min to 5 h is modelled in [30].

With a larger forecasting term, the daily GHI is predicted using ANN models for 25 Moroccan cities in [31], with empirical and machine learning models for 5 Moroccan cities in [32] and with hybrid ARIMA–ANN model for 3 cities in Morocco in [33]. The daily GHI is also forecasted with ANN models for 35 Moroccan, Algerian, Spanish and Mauritian cities in [34] and the monthly mean daily GHI using time series models in [35].

Although hourly and sub-hourly solar irradiation data are essential for an accurate techno-economical assessment of a CSP or PV project, a pre-feasibility study using monthly solar irradiation data is a common practice, and it is usually performed for the selection of potential sites where the power plant is expected to be located [36–38]. GHI and DNI are the types of data used for evaluating these projects. For the assessment of monthly data, ANN models are used to estimate it in Saudi Arabia [39,40] and in Uganda [41].

When it comes to the design of energy projects that involve GHI and DNI, it is important to test their model for different solar irradiation profiles to study the performance variations that the solar system could have when it is implemented in real life. Using only one set of data has the limitation that the model works completely fine if the studied conditions are met. These conditions often represent the mean values that the solar installations will experience. However, when the solar system is subjected to different input conditions, the outcome results may vary greatly from the ones projected during its planning and design phase. That is, there will be a chance that the projected profitability is not reached. As such, the variability of solar irradiation can be considered a risk since it can jeopardize the project's profitability. Since solar systems are going to be implemented increasingly in the future, this variability should be properly taken into account.

Frequently, however, the availability of enough measured data is limited. For this reason, satellite-based data are used to assess the solar resource when the measured data are scarce. References [42,43] compared the performance of several satellite data sets with ground measurements data in Morocco and North Africa, respectively. Statistical methods for measuring errors were used in various papers as a way to validate the solar irradiation data. For instance, Aguiar et al. [44] compared satellite data and ground measurements for various sites in the Canary Islands. Urraca et al. [45] employed the same validation method for various sites in Europe, detecting operational errors for some Baseline Surface Radiation Network (BSRN) stations. Schumann et al. [46] undertook data validation through statistical methods for Tamanrasset, Algeria, and Meyer et al. [47] applied it in Spain. Ineichen et al. [48] concluded that for European and Mediterranean sites, the irradiance data retrieved from various satellite databases had low uncertainty with a negligible bias when compared to ground measurements for the same locations. For African sites however, the lack of meteorological stations in the region could prove fatal for solar projects evaluation. Because of that, satellite data are one option that most projects decide on. References [49–51] showed that the SARAH database is good enough for monitoring and analysis of solar conditions in many sites, especially for Africa. Additionally, Huld et al. [52] concluded that the PVGIS database is of high quality for PV performance estimates for both Europe and Africa. Finally, in references [53,54] it was stated that solar irradiance values have high variability in areas with variable landforms, such as those with mountains and coastal areas.

The main characteristics of interest of the several references considered here are summarized in Table 1. The references have been listed either according to the long- or short-term character of the solar irradiance forecasting or according to their use of satellite-based data.

Table 1. Classification of the considered references. Source: own elaboration based on [9–54].

Term of Solar Irradiance Forecasting	
Short-term irradiance forecasting [9–35]	[9] Intra-hour GHI cloud retrieval technique to develop a physics-based smart persistence model [10] Intra-day GHI and DNI algorithm using cloud physical properties [11] A 15 min GHI forecasting model [12] Hourly-averaged GHI forecasts [13] Hourly GHI and DNI clear-sky irradiance vs. RRTMG physical radiative transfer model [14] Hourly and daily GHI from mesoscale atmospheric weather research forecasting model [15] Hourly GHI with a three-dimensional meteorology–chemistry model including a treatment of aerosols [16] Hourly GHI exponential smoothing model with decomposition methods [17] A 1 min DNI under a probabilistic approach [18] Short-term GHI with hybrid convolutional ANN model with spatiotemporal correlations [19] Short-term GHI and DNI forecasts of a global numerical weather model [20,21] A 5–30 min GHI and DNI with machine learning techniques [22] A 15 min GHI and DNI with machine learning techniques [23] Intra-day GHI with machine learning techniques [24] A 30 min GHI with ANN algorithm [25,26] Hourly GHI ANN models [27] Mean daily GHI with ANN models [28] A 500 ms–5 min GHI based on k-means algorithm [29] A 5–30 min GHI and DNI based on the k-nearest neighbours algorithm [30] A 30 min–5 h GHI Gaussian process regression method [31] Daily GHI with ANN models for for 25 Moroccan cities [32] Daily GHI with empirical and machine learning models for 5 Moroccan cities [33] Monthly mean daily GHI using time series models [34] Daily GHI with hybrid ARIMA–ANN model for 3 cities in Morocco [35] Daily GHI with ANN models for 35 Moroccan, Algerian, Spanish and Mauritian cities

Table 1. *Cont.*

Term of Solar Irradiance Forecasting	
Monthly irradiance forecasting [36–41]	[36] Best Practices Handbook for the Collection and Use of Solar Resource Data, selection of potential sites [37] Steps for solar resource assessment, selection of potential sites [38] Solar resource assessment, selection of potential sites [39] Monthly data, ANN models are used to estimate it in Saudi Arabia [40] ANN models are used to estimate it in Saudi Arabia [41] ANN models are used to estimate it in Uganda
Use of satellite-based data for solar resource assessment	
Use of satellite-based data for solar resource assessment [42–54]	[42] Satellite data comparison with ground measurements in Morocco [43] Satellite data comparison with ground measurements in North Africa [44] Satellite data comparison with ground measurements in the Canary Islands [45] Satellite data comparison with ground measurements for sites in Europe [46] Satellite data validation through statistical methods in Algeria [47] Satellite data validation through statistical methods in Spain [48] Satellite data comparison with ground measurements for European and Mediterranean sites [49–51] SARAH satellite database validation for several sites, especially for Africa [52] PVGIS satellite database validation for both Europe and Africa [53,54] Variability of irradiance values in areas with variable landforms

1.2. Justification and Objectives

As mentioned earlier, Morocco is currently a key country for solar development, and as a consequence some studies dealing with irradiation prediction and characterization focused on that country can be found in the literature. For instance, El Mghouchi et al. [55] carried out an assessment on the different solar irradiation prediction models for the northern Moroccan city of Tetouan. In reference [56], two empirical models were analyzed for 24 different cities across Morocco. Marchand et al. [42] conducted a study on the variability of solar irradiation data on different locations in the country. Proving the above-mentioned trend of employing satellite data for the solar resource validation, Wahab et al. [43] compared satellite and ground measured data for northern African countries, including Morocco. Likewise, references [31–35] have also been focused on the irradiation forecasting in several cities of Morocco, employing different methodologies.

Given the high potential for solar energy development in the MENA region and, particularly in Morocco, this paper will focus on this country to ease the deployment of solar energy projects.

As concluded in the above paragraphs, the prediction and characterization of solar irradiation using satellite data provides accurate and satisfactory results. Moreover, as highlighted in Section 1.1, state of the art, the prediction and characterization of solar irradiation rely mostly on either the use of complex models (i.e., Collares-Pereira) or on complex mathematical techniques (such as ANN-based algorithms). This mathematical complexity might hamper their use by businesses and project developers when assessing the solar resource.

In this regard, the first objective of the paper is to introduce a comprehensive and simple methodology to assess the solar resource for a project, based on the determination of its probability distribution function (PDF) for a specific location, assuming that the knowledge of statistical techniques may be more widely extended than other more complex mathematical techniques.

A second objective of the paper is to illustrate the selection of the best satellite database available for a determined location, focusing on the case of several cities across Morocco.

To the best of the authors' knowledge, no studies have been found where a new database for solar irradiation values that depends on historical data is provided for Morocco. Thus, it is considered to be the first study covering this gap in the scientific literature. This methodology is easily replicable to other locations, and it might be a useful tool to provide reliable solar irradiation data for future solar project assessments.

The paper is organized as follows. First, a general overview of the employed methodology is given in Section 2. Then, the selected case study is defined, and the data collection process is explained, as well as the validation and quality control phase, in order to use the best possible set of satellite data. Once the definitive set of data is prepared, distribution -fitting tests are used and the frequency of the irradiation values for every hour in a typical day of each month is presented in Section 3. In Section 4, the discussion on the results obtained is undertaken followed by a discussion on the potential applications that this methodology has. Finally, conclusions are raised in Section 4.

2. Materials and Methods

2.1. Methodology

The present study was conducted following different stages. Figure 3 provides a flowchart with the applied methodology.

As can be seen in Figure 3, in the first stage, solar radiation data from different sources is retrieved. Satellite-based data are used, as reliable and validated ground-measured data are not easily available. This will allow to move to the next step, which is the validation of the data among the different databases that have been retrieved by using statistical indicators. The database with the lowest statistical error will then be chosen to apply the distribution fitting test, which will consist of a two-step process. The first step will determine whether the data set to be analyzed follows a unimodal distribution or not, by applying Hartigan's dip test. The second step is applied if the unimodality is confirmed, determining the parametric distribution fitting the best to the set of data by the Anderson–Darling test. A more detailed explanation of these steps will be given in the following sections.

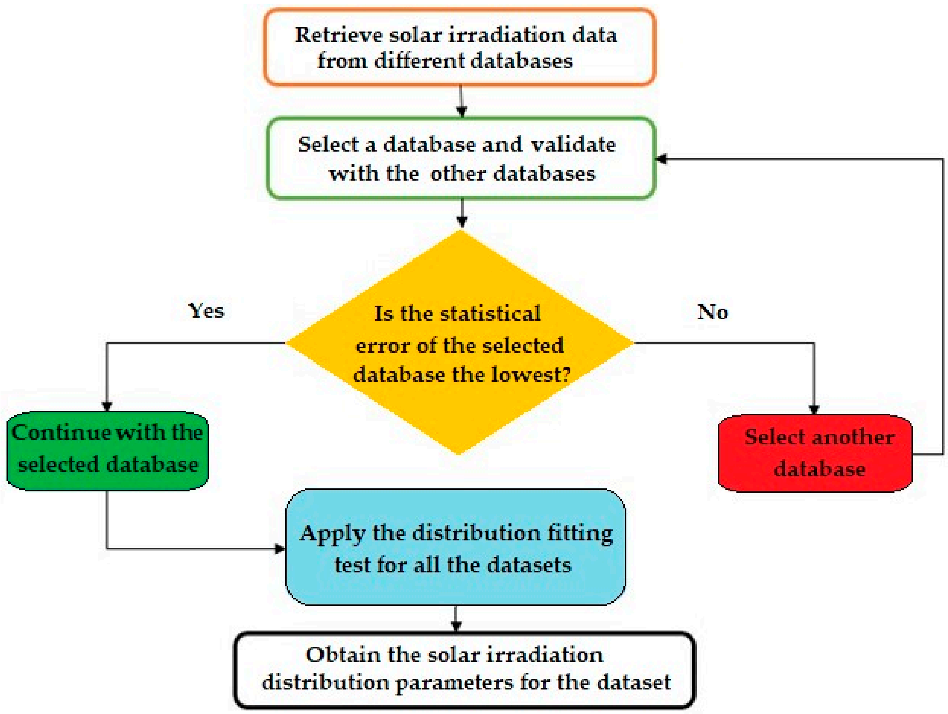

Figure 3. Flow chart of the proposed methodology. Source: own elaboration.

2.2. Case Study Definition and Data Collection and Validation

In order to cover a significant extent of the Moroccan territory, an extensive set of 23 representative cities of this country was chosen to carry out the analysis, listed at Table 2. Each one of the cities pertains to a different climatic zone, as defined in the map presented in [57,58], shown at Figure 4. The map was created using the degree day criteria and has been adopted for the Thermal Regulation of Construction in Morocco [59].

Table 2. Distribution of the analyzed cities in the different climatic zones of Morocco. Source: own elaboration based on [59,60].

Zone	City	Latitude [Degrees]	Longitude [Degrees]
1	Agadir	30.383	−9.567
	Casablanca	33.567	−7.667
	Essaouira	31.517	−9.783
	Kénitra	34.300	−6.600
	Laâyoune	27.160	−13.210
	Rabat	34.050	−6.767
	Safi	32.283	−9.233
	Sidi Ifni	29.360	−10.180
2	Al Hoceima	35.180	−3.850
	Larache	35.180	−6.130
	Nador	35.150	−2.910
	Tànger	35.733	−5.900
	Tétouan	35.580	−5.330
3	Beni Mellal	32.360	−6.400
	Fes	33.933	−4.983
	Meknes	33.883	−5.533
	Oujda	34.793	−1.933
	Taza	34.217	−4.000
4	Ifrane	33.500	−5.167
	Midelt	32.683	−4.733
5	Marrakech	31.617	−8.033
6	Er-Rachidia	31.930	−4.400
	Ouarzazate	30.933	−6.900

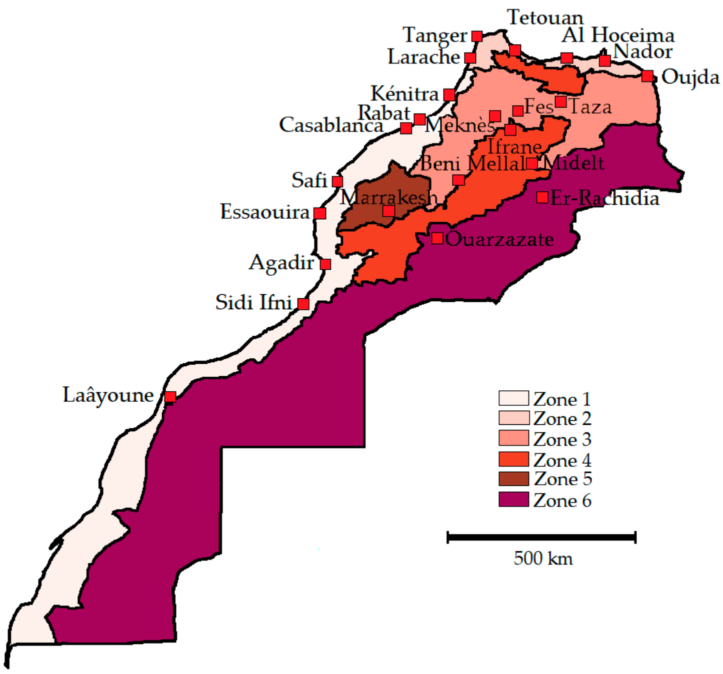

Figure 4. Climatic zones of Morocco. Source: own elaboration based on [57].

Measured solar irradiation data at ground level are not available for all cities and since there are no stations that belong to the BSRN in Morocco, only satellite data from different data sets are used [61]. Therefore, all data was collected through the PVGIS website [62]. The web application allows to obtain data using high spatial time and resolution. It allows to know the solar conditions from different regions around the world. The data can be obtained from different sets. Three out of four data sets are used, all of which have been validated using ground station measurements. The data sets used are PVGIS-SARAH, PVGIS-ERA5 and PVGIS-CMSAF.

For PV applications, the GHI is used to assess the feasibility of solar projects that use this technology, since these use both diffuse and direct radiation. For CSP applications, however, the diffuse irradiation is not relevant, which implies that only the DNI is considered for this technology [63]. An hourly time scale is necessary when analyzing specific projects for a duration that takes no longer than a year. This paper intends to present data for the assessment of solar projects for their whole lifetime, as carried out when evaluating their bankability in long-term planning. Thus, monthly rather than hourly intervals are used [64]. With all the above, the type of data used in this paper are the monthly irradiation data from each city, for the cases of global horizontal irradiation and direct normal irradiation.

A quality control (QC) was carried out to check whether the data are good for use or not. Because of the unavailability of ground solar irradiation data for some of the cities chosen in this study, the QC will be carried out between the three different data sets mentioned before. The QC will consist of computing the relative deviation between the solar irradiation measurements of each data set. There are statistical tools that are used to compute these deviations, and it is given by a single number. The higher the number, the higher the deviation between the measured values. The statistics used for the validation process are the mean absolute deviation (MAD), the mean bias deviation (MBD) and the root mean square deviation (RMSD), which are widely used in the validation process of satellite-based data with ground-based data [52]. These are calculated using the following expressions:

The mean absolute deviation (MAD):

$$MAD = (1/n) \sum_{i=1}^{n} \left| x_i^a - x_i^b \right|, \tag{1}$$

The mean bias deviation (MBD):

$$MBD = (1/n) \sum_{i=1}^{n} \left(x_i^a - x_i^b \right), \tag{2}$$

The root mean square deviation (RMSD):

$$RMSD = \sqrt{(1/n) \sum_{i=1}^{n} \left(x_i^a - x_i^b \right)^2}, \tag{3}$$

where n is the amount of data in the set, x_i^a is the data from set a and x_i^b is the data from set b.

It should be noted that each data set has different intervals. These are 11 years for the PVGIS-SARAH data set, 6 years for the PVGIS-ERA5 and 9 years for the PVGIS-CMSAF data set. Therefore, the value of n will be the lower one when computing the MAD, the MBD and the RMSD between two data sets.

The results of calculating the former statistics for the different pairs of data sets are shown in Tables 3 and 4 for the DNI and the GHI cases, respectively.

As it can be seen in Tables 3 and 4, the deviation between the data sets is lower in the case of GHI than in the case of DNI. Additionally, the blue dots in each of the subplots in Figure 5 correspond to the various months within the analyzed time intervals for selected cities, and their coordinates are the radiation values retrieved from the databases represented in the horizontal and the vertical axis, respectively. The subplots located at the left side of Figure 5 refer to DNI values while those at the right side relate to GHI values. The position of the represented data and its dispersion illustrates the degree of mismatch between the compared databases. Additionally, the corresponding linear regression has been determined, helping to visualize the existing relationship between the databases.

Table 3. MAD, MBD and RMSD for DNI [Wh/m^2]. Source: own elaboration.

	CM-SAF and ERA5			SARAH and ERA5			SARAH and CM-SAF		
City	**MAD**	**MBD**	**RMSD**	**MAD**	**MBD**	**RMSD**	**MAD**	**MBD**	**RMSD**
Agadir	17.57	−11.62	23.53	23.18	−19.75	29.28	12.91	−5.01	16.09
Al Hoceima	13.02	4.45	16.70	27.26	−24.88	31.79	29.03	−27.39	33.59
Beni Mellal	20.98	−20.33	24.05	20.52	−19.43	24.26	13.14	1.78	16.24
Casablanca	14.23	−0.18	17.09	21.63	−17.89	25.49	20.96	−12.96	25.34
Er−Rachidia	39.11	−39.11	41.09	19.48	−15.24	21.83	26.33	25.13	30.61
Essaouira	17.98	17.71	21.04	18.80	−17.32	22.90	33.18	−32.71	36.88
Fes	23.23	−21.11	27.88	17.14	−11.45	20.28	24.40	11.62	30.82
Ifrane	23.82	−23.01	26.82	26.35	−25.51	29.89	13.02	−1.68	16.97
Kenitra	15.15	11.13	18.76	19.15	−16.30	22.22	24.58	−23.50	29.05
Laayoune	14.55	−12.71	17.80	11.92	−8.08	15.02	12.38	6.21	15.81
Larache	13.17	−5.86	16.85	19.19	−16.93	21.95	16.74	−14.04	20.66
Marrakech	10.61	−7.06	13.47	14.30	−8.06	16.70	12.88	0.02	15.75
Meknes	12.61	−7.89	15.56	16.73	−13.23	19.22	13.46	−3.79	16.84
Midelt	24.60	−24.36	27.14	25.74	−23.79	30.58	15.78	1.58	19.79
Nador	16.93	4.57	21.47	25.88	−24.26	30.56	27.78	−25.55	33.88
Ouarzazate	21.79	−21.71	24.18	15.92	−9.87	21.35	19.81	14.58	24.82
Oujda	14.81	−2.52	17.49	19.23	−18.68	22.31	19.54	−13.91	24.61
Rabat	15.24	8.14	17.83	21.76	−19.38	24.95	25.31	−22.44	29.88
Safi	14.83	12.33	17.52	15.50	−11.98	19.71	24.73	−22.82	29.06
Sidi Ifni	26.04	−13.87	33.89	25.83	−25.29	32.75	25.82	−16.33	29.85
Tanger	14.23	−3.23	17.78	17.67	−15.10	21.62	18.20	−10.85	21.76
Taza	10.98	0.43	13.66	15.42	−10.35	18.40	15.78	−8.96	19.57
Tetouan	36.69	−36.10	44.23	38.43	−37.95	46.31	13.97	−1.22	18.15

Table 4. MAD, MBD and RMSD for GHI [Wh/m^2]. Source: own elaboration.

	CM-SAF and ERA5			SARAH and ERA5			SARAH and CM-SAF		
City	**MAD**	**MBD**	**RMSD**	**MAD**	**MBD**	**RMSD**	**MAD**	**MBD**	**RMSD**
Agadir	6.51	1.30	8.30	6.06	0.89	7.53	3.87	0.55	5.35
Al Hoceima	7.54	4.80	8.93	6.05	−2.51	8.96	6.55	−6.20	8.20
Beni Mellal	4.65	−1.93	6.24	5.07	−1.93	6.98	3.73	0.46	4.64
Casablanca	7.23	5.53	8.51	5.70	2.44	6.88	5.29	−1.27	6.99
Er-Rachidia	11.00	−11.00	12.24	4.38	1.77	6.05	14.16	13.81	15.11
Essaouira	12.07	12.07	13.07	5.19	2.61	6.60	8.19	−8.10	10.05
Fes	4.97	−1.24	6.23	5.62	3.88	6.86	6.91	5.90	8.40
Ifrane	7.36	−6.10	8.77	6.15	−3.61	7.97	5.10	3.05	6.44
Kenitra	10.45	10.34	11.96	5.49	3.73	6.57	5.54	−4.89	7.55
Laayoune	4.89	2.75	6.29	7.00	5.36	8.03	6.33	3.79	7.76
Larache	3.95	−0.55	5.17	4.59	2.38	5.51	3.30	2.33	4.21
Marrakech	4.86	3.03	5.73	5.36	3.76	6.52	3.31	1.14	4.22
Meknes	4.70	2.71	6.17	4.82	2.78	5.85	3.27	0.71	4.07
Midelt	8.79	−8.62	10.60	7.50	−5.91	10.34	6.95	3.59	8.93
Nador	7.50	2.81	9.61	5.75	−1.49	8.63	5.12	−2.92	6.61
Ouarzazate	3.88	−2.00	4.82	6.85	2.23	9.33	8.53	5.77	10.05
Oujda	4.69	1.20	6.04	4.21	0.24	5.50	4.04	0.15	5.13
Rabat	9.91	9.56	11.46	4.98	1.80	6.02	6.67	−5.50	8.94
Safi	9.06	8.88	10.20	5.09	3.03	6.38	5.75	−4.79	7.78
Sidi Ifni	12.48	0.31	15.45	8.17	−2.60	11.66	11.22	−4.97	13.01
Tanger	5.61	3.63	6.86	5.26	3.00	6.31	3.89	−0.30	4.68
Taza	6.70	5.77	8.36	5.97	4.04	7.26	4.13	−0.55	5.85
Tetouan	13.08	−12.41	18.69	10.75	−7.97	16.04	5.42	4.87	7.87

It is possible to observe that the GHI data have less dispersion than the DNI, suggesting that for all the databases, the GHI values seem more reliable than the DNI ones. Using the numerical results from Tables 3 and 4, it can be stated that those pairs of databases involving PVGIS-SARAH for the GHI and PVGIS-CMSAF for the DNI accumulate the higher number of error metrics with lower average values for all the analyzed cities. Consequently, it seems convenient to use the PVGIS-CMSAF in the case of the DNI and the PVGIS-SARAH in the case of GHI.

As can be seen in Figure 6, the selected DNI and GHI databases are then categorized into several sets, according to the month and the considered city. A total number of sets corresponding to 12 months by 23 cities is created for both DNI and GHI radiation data. The size of each of the sets amounts to the number of years collected in the CMSAF and SARAH databases, i.e., 9 and 11 years, respectively.

Once the classification of the data is complete, the process to determine the probability distribution function (PDF) will be explained next.

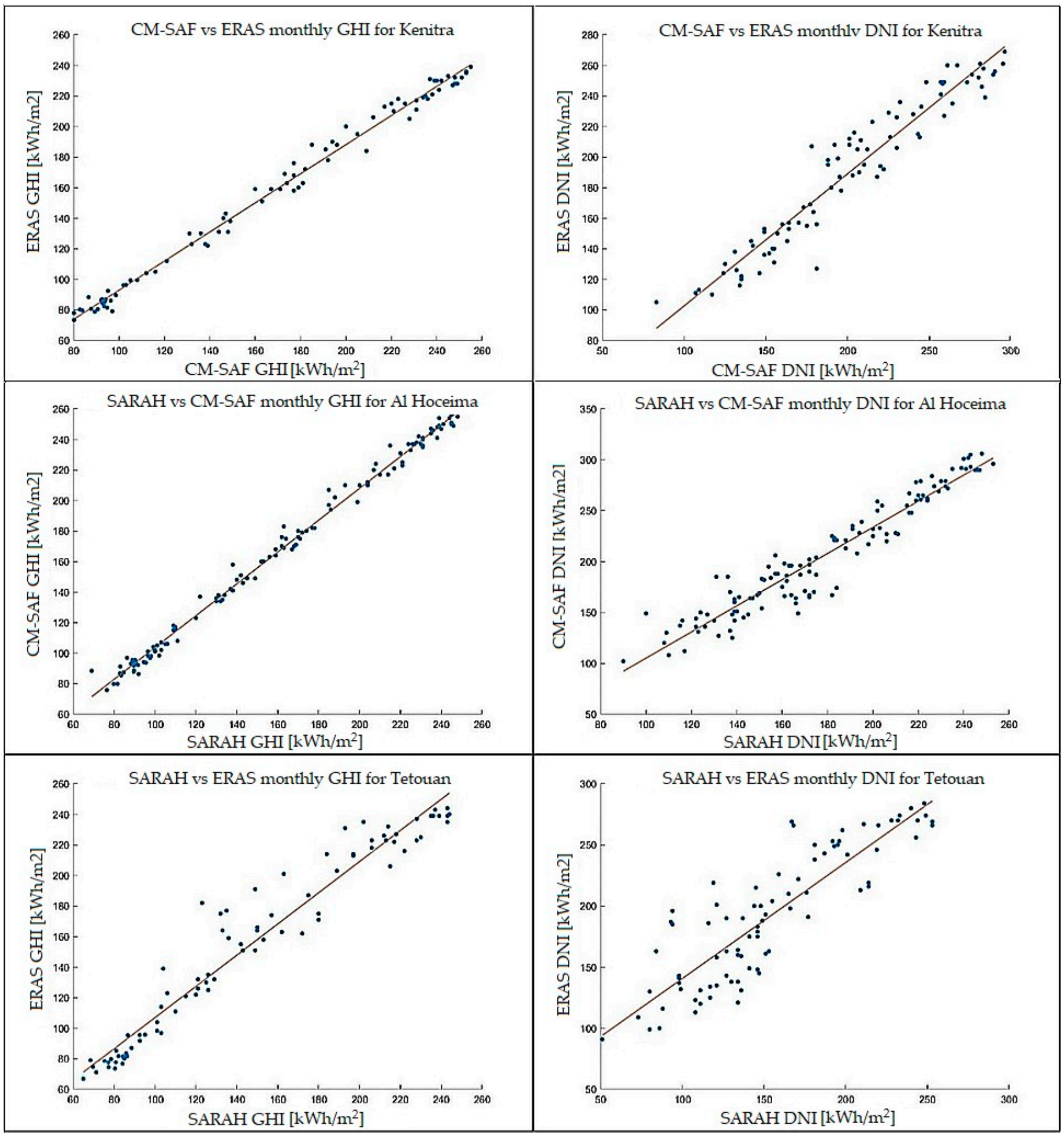

Figure 5. Monthly irradiation values within the analyzed time intervals comparing pairs of databases and linear regression for the validation of data. Source: own elaboration.

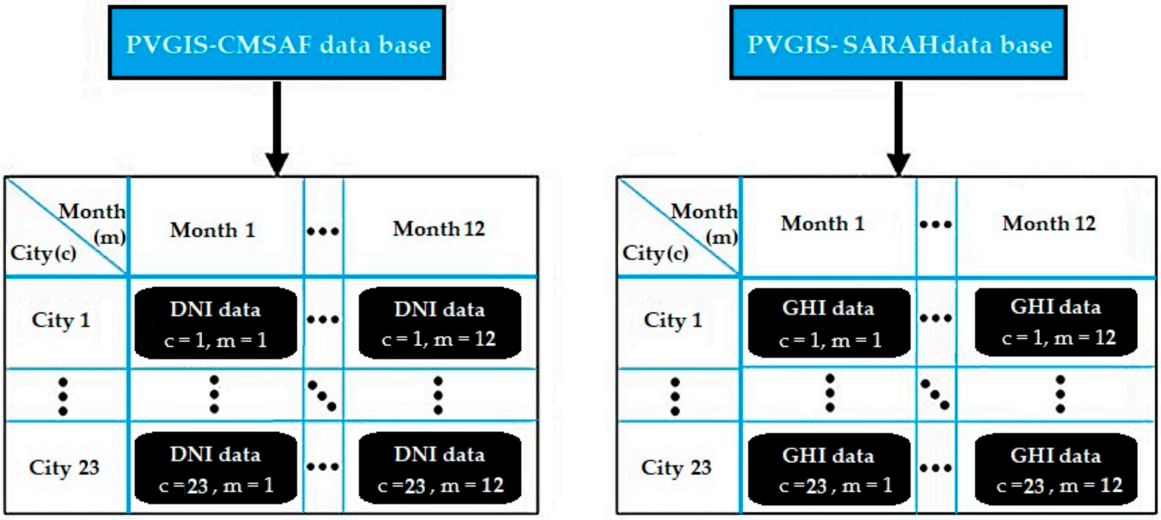

Figure 6. Categorization of the DNI and GHI data according to the month and considered city. Source: own elaboration.

2.3. Probability Distribution Function Fitting

Probability distribution fitting is the next step in the process. To this end, the overall procedure is illustrated in Figure 7.

For a given city and month, the input set of data shown in Figure 7 is collected either from the PVGIS-CMSAF in the case of the DNI or from the PVGIS-SARAH in the case of GHI. Then a prior test is conducted to determine whether the set of data follows a unimodal or a bimodal distribution. Other modalities such as tri-modal distributions will not be considered in this work. The particular selected test is Hartigan's dip test, because of the reliability of its results [65]. The test will use a uniform distribution that minimizes the maximum difference between the distribution of the sample that needs to be analyzed and the aforementioned uniform distribution. This is referred to as the "dip", and its value will be used to assess the unimodality of the distribution of the sample, as mentioned before. The detailed calculation procedure follows the one explained in [66]. From this, Hartigan's dip statistic can be obtained, for which values can range from 0 to 1. The classification criteria used is the following: if the statistic is less than 0.1, it means that the sample has a significant bimodal behaviour. Conversely, the sample can be considered as unimodal.

In the case of bimodality, it is frequently difficult or even unfeasible to fit a parametric PDF [67]. In this event, the so-called kernel density estimation is one of the most common non-parametric approaches for estimating the PDF [68], and it is the method of choice in this study.

On the other hand, if from Hartigan's dip statistic the unimodal distribution of the input sample is concluded, a parametric distribution fitting test will be employed for characterizing the PDF of the satellite radiation data. There are many tools that are useful to assign a parametric distribution to a set of data. Among these tests, we can find the Kolmogorov–Smirnov test, the chi-square test and the Anderson–Darling test, the latter being more powerful among the ones mentioned as it gives more weight to the tails than other tests [69]. This will allow to better factor in extreme solar irradiation values that were registered in the database.

Figure 8 illustrates the step-by-step calculation of the Anderson–Darling test statistic and corresponds to the red block labelled "Apply the AD test" in Figure 7. The input data to this procedure is the same as shown in Figure 7 in case that the unimodality is confirmed by Hartigan's dip test ($p > 0.1$).

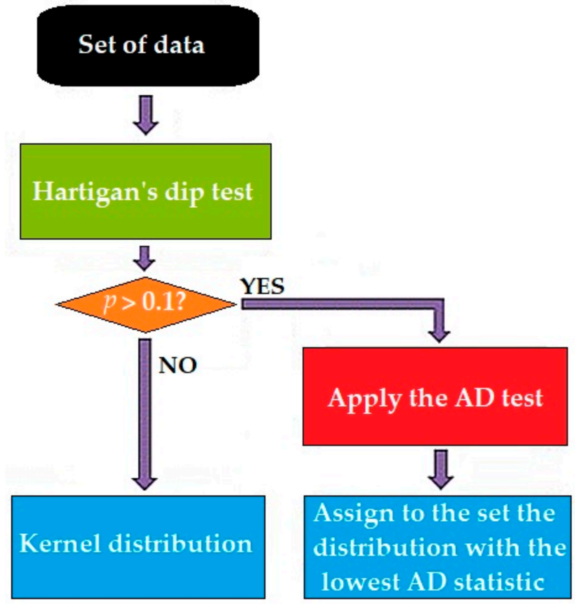

Figure 7. Step-by-step calculation for distribution fitting. Source: own elaboration.

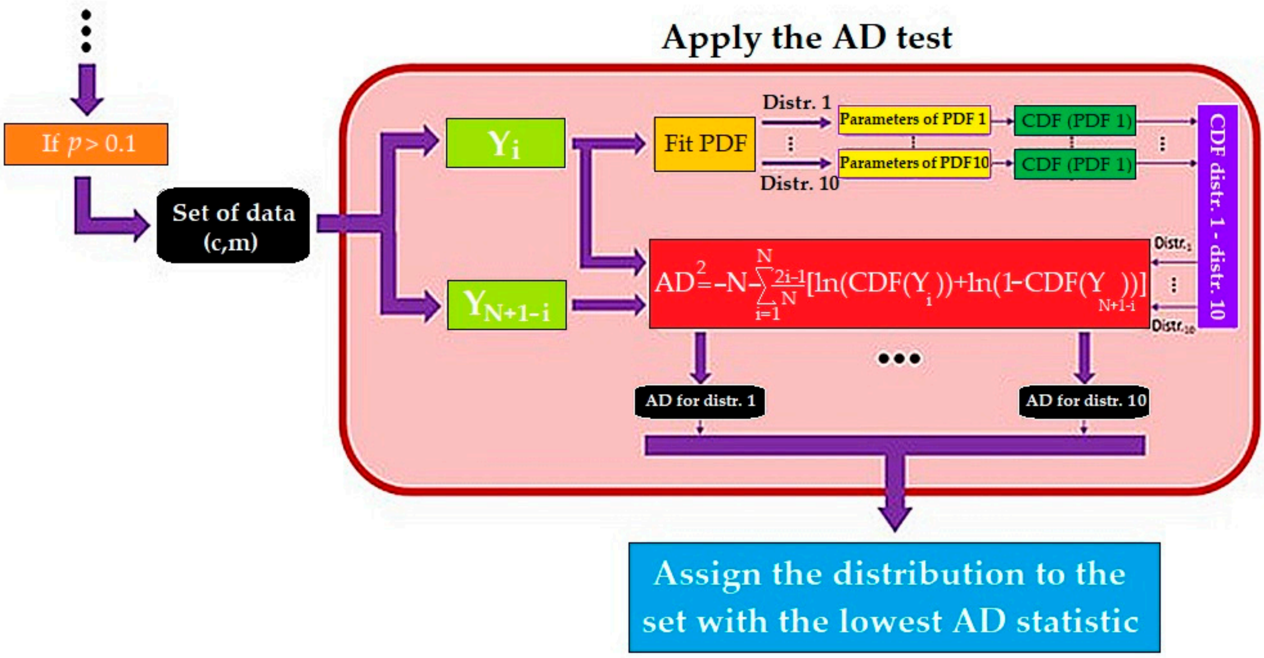

Figure 8. Step-by-step calculation of the Anderson–Darling (AD) test. Source: own elaboration.

The Anderson–Darling (AD) test statistic is defined as

$$AD^2 = -N - \sum_{i=1}^{N} \frac{2i-1}{N}[\ln(CDF(Y_i)) + \ln(1 - CDF(Y_{N+1-i}))], \quad (4)$$

where

N: number of elements in the input data for every city and month through the analyzed period.

i: index taking values in the range $1 \le i \le N$.

Y_i: element of the input data sorted in an ascending order.

Y_{N+1-i}: element of the input data sorted in a descending order.

$CDF(\ )$: cumulative distribution function (CDF) of a specified PDF, calculated for an element Y_i or Y_{N+1-i} of the input data.

To determine which probability distribution fits for each sample, the AD statistic is calculated assuming it follows a specified distribution. In this study, the 10 most frequent distributions found in the literature were chosen to have the data samples fitted into. Table 5 displays these PDFs and their CDFs for each of the selected distributions:

Table 5. Distributions considered with their respective PDF and CDF. Source: own elaboration based on [69–71].

Distribution Name	PDF [f(x)]	CDF [F(x)]
Beta	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(p,q)}$	$\frac{\int_0^z \left(t^{p-1}(1-t)^{q-1}dt\right)}{B(p,q)}$
Chi-square	$\frac{\exp(-x/2)\ x^{\frac{v}{2}-1}}{2^{v/2}\ \Gamma(v/2)}$	$\frac{\gamma\left(\frac{v}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}$
Exponential	$\frac{1}{\beta}\exp\left(-\frac{x-\mu}{\beta}\right)$	$1-\exp(-x/\beta)$
Extreme value	$\frac{1}{\beta}\exp\left(\frac{x-\mu}{\beta}\right)\exp\left(-\exp\left(\frac{x-\mu}{\beta}\right)\right)$	$1-e^{-e^{x}}$
Gamma	$\frac{x^{(\gamma-1)}\exp(-x)}{\Gamma(\gamma)}$	$\frac{\int_0^x \left(t^{\alpha-1}\exp(-t)dt\right)}{\Gamma(\gamma)}$
Lognormal	$\frac{\exp\left(-\left(\frac{\left(ln\left(\frac{x-\theta}{m}\right)\right)^2}{2\sigma^2}\right)\right)}{(x-\theta)\sigma\sqrt{2\pi}}$	$\Phi\left(\frac{\ln(x)}{\sigma}\right)$
Normal	$\frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$	$\int_{-\infty}^{x}\frac{\exp(-x^2/2)}{\sqrt{2\pi}}$
Rayleigh	$\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$	$1-\exp\left(-\frac{x^2}{2\sigma^2}\right)$
T-distribution	$\frac{\left(1+\left(\frac{x^2}{v}\right)\right)^{-\frac{v+1}{2}}}{B(0.5,0.5v)\sqrt{v}}$	$\frac{1}{2}+\frac{1}{2}\left[I\left(1;\frac{1}{2}r,\frac{1}{2}\right)-I\left(\frac{r}{r+t^2},\frac{1}{2}r,\frac{1}{2}\right)\right]sgn(t)$
Weibull	$\frac{\gamma}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{\gamma-1}\exp\left(-\left(\frac{x-\mu}{\alpha}\right)^{\gamma}\right)$	$1-\exp(-(x^{\gamma}))$

As can be seen in Equation (4), the CDFs corresponding to each of the tested PDFs are needed. The parameters of the several CDFs are obtained by adjusting their PDFs to the input data sorted in ascending order (Y_i), as can be seen in the upper part of Figure 8. Then, the CDF values of the different elements in the input data sorted both in ascending (Y_i) and descending (Y_{N+1-i}) orders are computed and the rest of calculations represented in Equation (4) are conducted, in order to obtain the corresponding AD statistic for each of the tested PDFs in Table 5 (see Figure 8). If the AD statistic for a certain distribution has the lowest value amongst the rest, the input data are best fitted into the distribution in question [72].

3. Results and Discussion

3.1. Obtained Results

Table 6 shows the DNI PDFs for each month and for each city. The cities have been sorted according to their own climatic zone shown in Table 2.

In this case, the predominant PDF is the extreme value for all zones except for the cities in zone 5, where the logarithmic PDF prevails. The gamma PDF is the least predominant among all the distributions, with an occurrence that is not significant at zone 1. In zone 4, there are no occasions where the PDF is bimodal, while for the rest of the zones, there are few instances when the bimodality is present and a kernel PDF is assigned.

Table 6. PDFs for the DNI case for each city and month (normal = Nor, logarithmic = Log, extreme value = ExV, gamma = Gmm, Weibull = Web, kernel = Krn). Source: own elaboration.

Zone	City	Month											
		Jan.	**Feb.**	**Mar.**	**Apr.**	**May**	**Jun.**	**Jul.**	**Aug.**	**Sept.**	**Oct.**	**Nov.**	**Dec.**
1	Agadir	Log	ExV	ExV	ExV	Nor	Nor	ExV	Log	ExV	Nor	Log	Web
	Casablanca	ExV	Web	ExV	ExV	Log	Nor	ExV	Web	ExV	Nor	Krn	ExV
	Essaouira	Krn	ExV	Nor	Nor	ExV	ExV	Nor	ExV	ExV	ExV	ExV	Log
	Kenitra	ExV	Web	ExV	Log	Nor	Nor	ExV	Nor	ExV	Web	Gmm	Nor
	Laayoune	ExV	Nor	Log	ExV	ExV	Log	Nor	Web	ExV	Web	Log	Log
	Rabat	ExV	Web	Krn	Nor	Nor	Nor	Nor	Nor	Nor	ExV	Log	Nor
	Safi	Log	Web	ExV	Nor	ExV	Nor	Nor	ExV	ExV	Log	Nor	Log
	Sidi Ifni	Log	Log	ExV	Nor	ExV	Nor	Nor	ExV	ExV	ExV	ExV	Log
2	Al Hoceima	ExV	Nor	Nor	Nor	ExV	Log	Log	Log	Log	Web	Log	ExV
	Larache	Log	Log	ExV	ExV	Log	Nor	Nor	Nor	Log	Nor	ExV	ExV
	Nador	ExV	ExV	Nor	Krn	ExV	ExV	Log	ExV	Nor	ExV	Nor	Nor
	Tanger	ExV	ExV	ExV	Log	Nor	Nor	Nor	Krn	Web	ExV	Log	ExV
	Tetouan	ExV	Nor	Krn	Nor	ExV	Web	Log	Log	Nor	Nor	Log	Nor
3	Beni Mellal	Krn	ExV	ExV	Nor	ExV	Log	Log	ExV	Nor	Log	Nor	ExV
	Fes	Nor	ExV	ExV	Web	Log	Log	Log	ExV	Nor	ExV	Log	ExV
	Meknes	Nor	Web	ExV	ExV	Log	Krn	Nor	ExV	Web	Nor	Log	ExV
	Oujda	ExV	ExV	Nor	ExV	ExV	Web	Web	ExV	ExV	Nor	Nor	Log
	Taza	Nor	Log	ExV	ExV	ExV	Log	Web	Web	ExV	Web	Log	Log
4	Ifrane	ExV	ExV	ExV	Nor	Nor	Log	Web	Log	ExV	Nor	Log	Nor
	Midelt	Log	ExV	Log	ExV	ExV	Log	Nor	ExV	Log	Log	ExV	Log
5	Marrakech	Nor	Krn	ExV	Nor	ExV	ExV	Log	Log	Log	Log	Krn	Log
6	Er-Rachidia	Log	ExV	Log	ExV	ExV	ExV	Log	ExV	Log	Log	ExV	Log
	Ouarzazate	Nor	ExV	Log	Web	ExV	ExV	Web	Nor	Krn	Nor	Web	Log

The same information presented in Table 6 is depicted at Figure 9, in order to graphically show the percentage that a certain PDF appears in each of the zones. From Figure 9, it can be seen that certain PDFs appear more often in some zones, while in other zones, they do not appear, stating the great location-dependent radiation characterization.

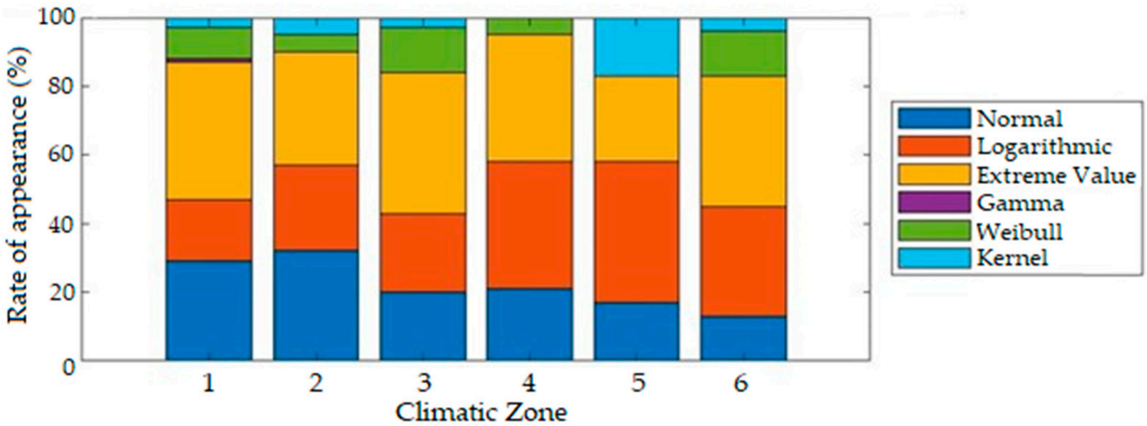

Figure 9. Bar chart with the rate of appearance of the PDFs in each month per climatic zone (DNI case). Source: own elaboration.

Figure 10 displays examples of the PDF fitting test results. The histograms of the data samples for the month of October in Nador, the month of December in Kenitra and the month of March in Tetouan are shown along with the characterization of the adjusted extreme value, normal and kernel PDFs, respectively. The results show a good performance of the test for PDF fitting.

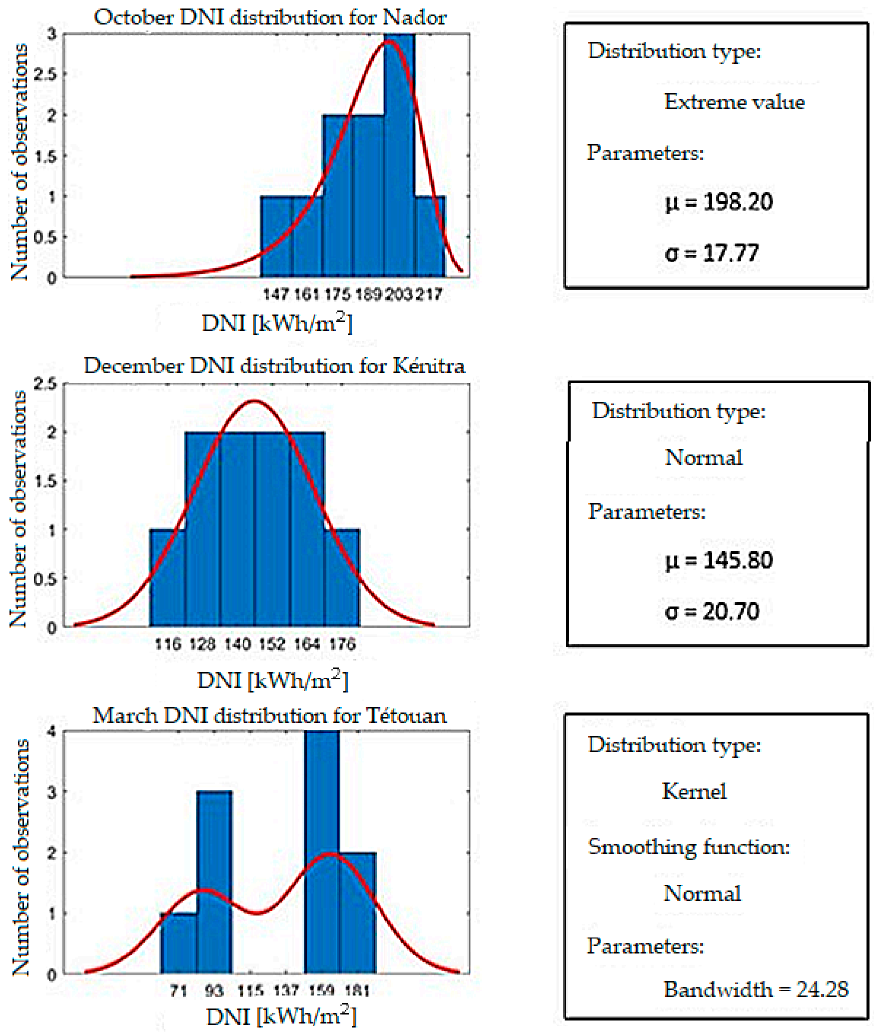

Figure 10. Example of the extreme value, normal and kernel PDFs' fitting test results for the month of October at Nador, December at Kenitra and March at Tetouan (DNI case). Source: own elaboration.

For the GHI, the results are displayed in Table 7. As shown in Figure 11, the most dominant PDFs are the extreme value, logarithmic and the normal ones. Specifically, the extreme value PDF clearly prevails at zone 4, while it is less present at zone 2. In contrast to the DNI case, now the gamma PDF has a higher rate of appearance, particularly at zone 5. Again, a great location dependence of the obtained results is exhibited.

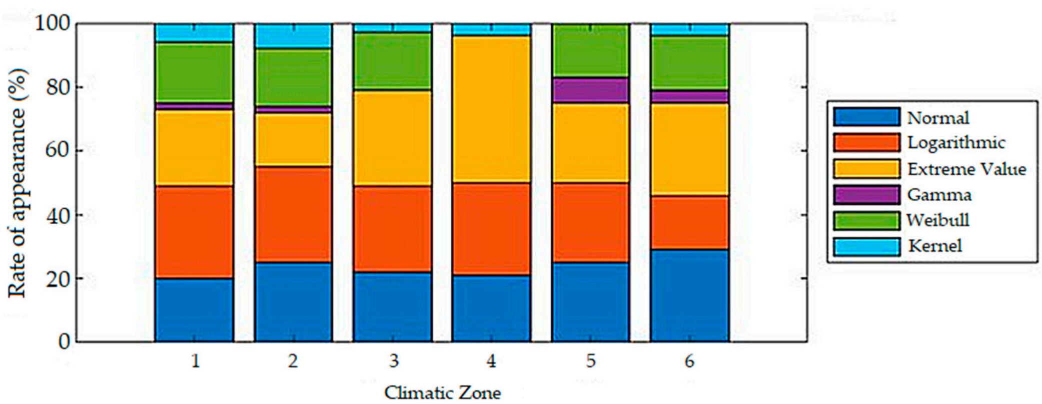

Figure 11. Bar chart with the rate of appearance of the PDFs in each month per climatic zone (GHI case). Source: own elaboration.

Table 7. PDFs for the GHI case for each city and month (normal = Nor, Logarithmic = log, extreme value = ExV, gamma = Gmm, Weibull = Web, kernel = Krn). Source: own elaboration.

Zone	City	Month											
		Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
1	Agadir	ExV	Web	Web	Nor	Web	Web	Krn	Log	ExV	Web	Web	Log
	Casablanca	Nor	Gmm	Web	Nor	Nor	Web	ExV	Nor	Web	ExV	Nor	Web
	Essaouira	Log	Gmm	ExV	Log	ExV	Log	ExV	ExV	Web	ExV	Log	Log
	Kenitra	Nor	Log	Web	Log	Log	Web	ExV	ExV	Nor	Nor	Nor	Krn
	Laayoune	ExV	Web	Nor	ExV	ExV	Log	ExV	Web	Web	Log	Log	Krn
	Rabat	Log	Log	Web	Log	Nor	Log	ExV	Krn	Log	Nor	Log	Krn
	Safi	Log	Log	ExV	Krn	ExV	Log	ExV	ExV	Nor	Nor	Log	Log
	Sidi Ifni	Nor	Log	Log	Log	Nor	Nor	ExV	ExV	Nor	ExV	Web	Log
2	Al Hoceima	Log	Log	Krn	Nor	ExV	Nor	Nor	Log	Log	Web	Log	Web
	Larache	Log	Log	Web	Log	Nor	Nor	Nor	Web	Nor	Nor	Log	ExV
	Nador	Nor	Krn	Krn	Nor	ExV	Nor	Log	Nor	Log	Web	Web	Nor
	Tanger	Log	Gmm	Web	Log	Web	Krn	Log	ExV	Nor	Web	Log	ExV
	Tetouan	Log	Web	Web	ExV	ExV	ExV	Krn	Log	Log	ExV	Nor	ExV
3	Beni Mellal	Nor	ExV	Log	Web	ExV	Web	Web	ExV	Web	Log	Nor	ExV
	Fes	Log	Nor	ExV	Log	Log	Web	Nor	Nor	ExV	Nor	Nor	ExV
	Meknes	Log	Nor	ExV	Log	Log	ExV	Nor	ExV	ExV	Krn	Nor	ExV
	Oujda	Nor	ExV	Log	ExV	ExV	Log	Web	Web	Nor	Web	Web	Log
	Taza	Log	Log	Log	ExV	Krn	Log	Web	Nor	ExV	Web	Log	ExV
4	Ifrane	ExV	ExV	ExV	ExV	Log	Log	Nor	Log	ExV	Nor	Nor	ExV
	Midelt	ExV	Nor	Log	Krn	ExV	Log	ExV	Nor	ExV	Log	Log	ExV
5	Marrakech	Nor	Nor	Log	ExV	ExV	ExV	Log	Web	Web	Gmm	Log	Nor
6	Er-Rachidia	Gmm	Log	Log	Nor	Nor	Web	Nor	ExV	Log	Krn	Web	ExV
	Ouarzazate	Nor	Log	Nor	Web	ExV	ExV	ExV	Nor	Nor	Web	ExV	ExV

Figure 12 shows the histograms of the month of February at Ifrane, the month of October at Nador and the month of March at Tetouan, along with the adjusted extreme value, normal and kernel PDFs, respectively. It can be seen that the results of the distribution fitting test displayed good performance.

3.2. Analysis and Discussion of Results

The analysis of the results presented in Tables 6 and 7, and Figures 9 and 11 shows the great existing variety of PDFs for the DNI and GHI data at the different months and cities. This demonstrates the need for undertaking a proper characterization of the irradiation when the assessment of solar energy projects is involved.

In this regard, the deterministic approach for evaluating the profitability of solar systems is still prevalent in the literature since the probabilistic methodologies are frequently perceived by practitioners as a deterrent. Nevertheless, the presented methodology might allow developers to conduct more accurate assessments as it takes into account the inherent variability of the solar resource without adding excessive mathematical burden. This methodology is composed of simple and clear steps making use of widely known statistical tools, and therefore, it is an easy application when compared with the use of ANN.

The objective of foreseeing the radiation using ANN is to determine with a high rank of precision its specific value for a particular day within a precise time frame (usually an hour). It is used primarily when dealing with the management of energy systems. Although the methodology presented here could be used for foreseeing the specific value of radiation for a particular day and hour, its results would not be as accurate as those obtained using ANN. The ANN techniques for radiation forecasting provide a value according to the observed weather conditions of the previous years, requiring the knowledge of a complex data set of variables such as air temperature, humidity, wind speed and direction, atmospheric pressure, as well as solar radiation data. In this regard, the ANN techniques are not only far more complex, but are also more demanding in terms of the needed input meteorological data set. Taking into account the characteristics of both ANN and the presented methodology, it could be concluded that each method could be best suited for different types of applications. When the accuracy of radiation forecasting is the ultimate result, the performance of ANN is far superior. Nevertheless, when it is simply needed to embed the radiation uncertainty in the analysis, as is the case of the techno-economic valuation of solar energy assets, the presented methodology can reach this objective with much less complexity and less demanding input data.

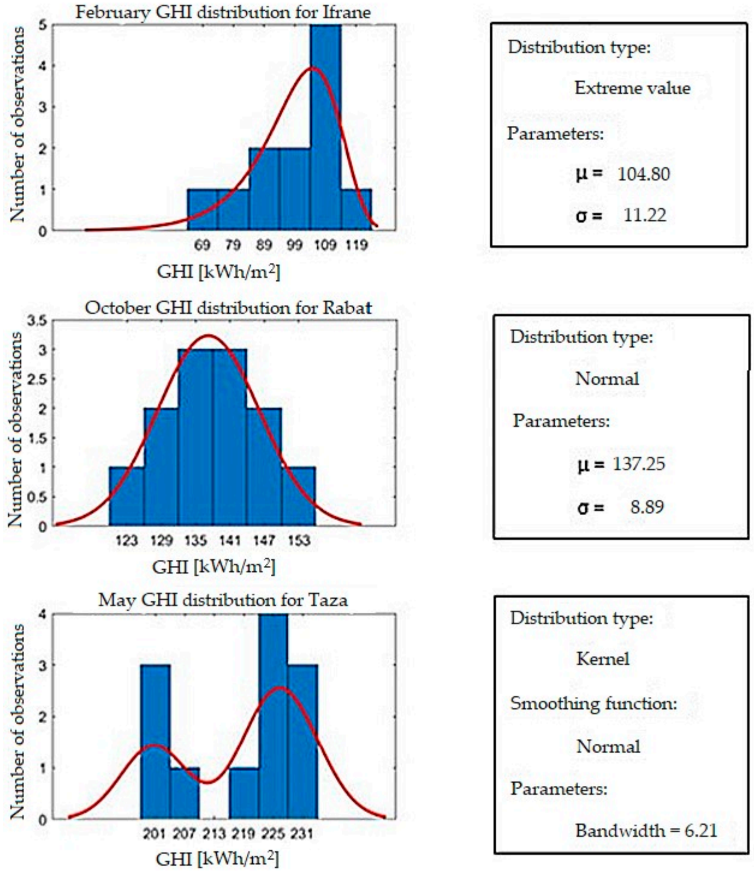

Figure 12. Examples of the extreme value, normal and kernel PDFs' fitting test results for the month of February in Ifrane, the month of October in Nador and the month of March in Tetouan (GHI case). Source: own elaboration.

The obtained PDFs can be embedded into probabilistic mathematical models for obtaining a big number of possible outcomes of a variable of interest, whether it is solar irradiation, or a profitability metric depending on it. Particularly, in the case of the well-known Monte Carlo simulation method, repeated sampling from the solar irradiation PDF is performed, in order to obtain the statistical properties of the uncertain output variable and provide the likelihood of occurrence of a determined outcome.

In order to illustrate the use and to assess the performance of the obtained PDFs, the GHI PDF for the month of October in the city of Rabat has been embedded within a probabilistic mathematical model, specifically a Monte Carlo simulation.

The obtained results were tested against the most recent satellite hourly data available within the PVGIS platform for Rabat, corresponding to the year 2020 SARAH2 database. The accumulated monthly GHI for the October 2020 benchmark was then obtained by adding all the hourly data for this month, resulting in 139.12 kWh/m^2. Next, from the PDF obtained for the city of Rabat in the month of October depicted at Figure 12, a Monte Carlo simulation was run.

Figure 13 shows the basic structure of the algorithm. Initially, for the selected city and month, the appropriate PDF is fed into the probabilistic model. An iterative random sampling process is then started for obtaining irradiation values. This process ends when the predefined number of samples (*Nsamples*) is achieved. In this case, 10,000 samples have been obtained.

As can be seen in Figure 14, the mean value of the simulated accumulated monthly GHI for October in the city of Rabat was 137.01 kWh/m^2. This figure also shows the evolution of the average of all the simulated values as the number of samples grows. The obtained mean of 137.01 kWh/m^2 represents a relative difference of 1.5% regarding the 2020 benchmark data of 139.12 kWh/m^2. The interpretation of this result is that the PDF in Figure 12 can provide a set of *Nsamples* feasible values of accumulated monthly GHI for the benchmark city and month, whose behaviour properly reflects the implied stochastic pattern of the irradiation.

In this sense, the determined PDFs can be considered as a new solar irradiation database for the corresponding locations and time, since random values can be generated from them, all of which reflect the variability of the solar resource of the site.

It must be clear that this methodology is not intended for predicting the real irradiation value for a specific location and month, but for providing feasible irradiation data able to characterize the irradiation variability. This methodology may be more than sufficient when assessing the economic profitability of solar assets, and far better than the deterministic approaches usually employed.

The same method can be applied for studying other phenomena, such as the variability of wind speeds, the ambient temperature of a location, or the seasonal humidity. This method is not limited to the energy sector but can be applied in other areas of science as well.

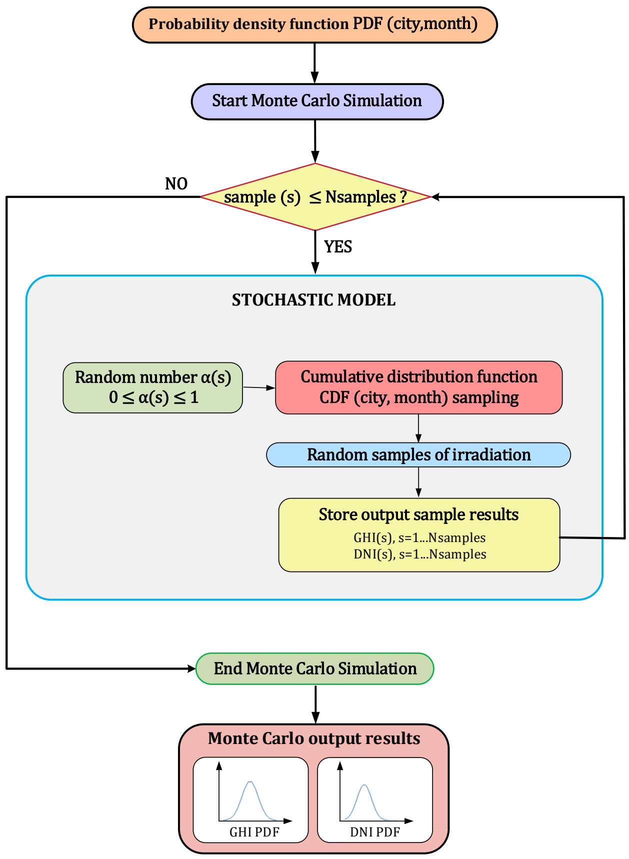

Figure 13. Basic structure of the Monte Carlo simulation algorithm for testing the PDFs. Source: own elaboration.

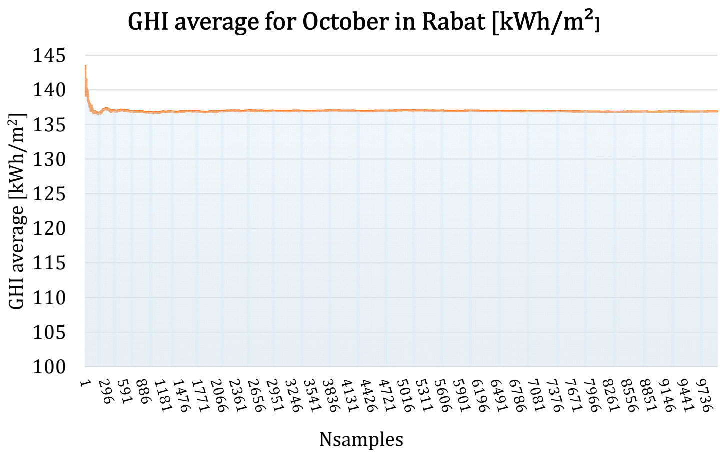

Figure 14. Evolution of the GHI average for October in Rabat according to the number of samples obtained in the Monte Carlo simulation. Source: own elaboration.

4. Conclusions

The main aim of this paper was to present a straightforward and comprehensive method to determine the PDF of the solar irradiation, addressed to handle the inherent variability of the solar resource for the techno-economic assessment of the energy assets. The final product of this methodology is a set of PDFs that characterizes the solar irradiation of the desired site and for a selected time resolution, which can be used as a solar irradiation database when assessing CSP and PV projects. In this regard, a second objective of the paper was to illustrate the selection of the best satellite database available for a determined location, which is exemplified on the country of Morocco, applying it to 23 representative cities belonging to the different climatic zones. Morocco being one of the most outstanding countries in the active promotion of solar energies in the MENA region, the presented method can be a useful tool for the accurate assessment of these projects.

First, monthly solar irradiation data were retrieved from various satellite databases (PVGIS-SARAH, PVGIS-ERA5 and PVGIS-CMSAF). The DNI and the GHI were the types of solar irradiation data used when applying the method. A quality control was performed on the retrieved data in order to determine which of the considered databases presented the lowest statistical errors. From the quality control results, the databases PVGIS-CMSAF and PVGIS-SARAH were considered the most convenient for characterizing the DNI and the GHI, respectively.

Next, a probability distribution fitting test was presented, able to identify bimodal behaviours in the irradiation data. When the bimodality was confirmed, a kernel distribution function was applied. Conversely, the set of data was regarded as unimodal and the PDF that fitted best was chosen.

It can be concluded that for both the DNI and the GHI cases, the extreme value, the logarithmic and the normal PDFs are predominant for the extensive set of Moroccan cities examined. On the other hand, other distributions are not that often. Likewise, the bimodal behaviour appears in some instances but does not follow a recognizable pattern. These varied results remark the need for incorporating the characterization of solar irradiation at the different climatic zones when undertaking the economic feasibility analysis of solar energy assets.

The main merit of the presented methodology is that it is composed of simple and clear steps making use of widely known statistical tools, and therefore it is an easy application when compared with the use of other state-of-the-art methods, as ANN. When it is simply needed to embed the radiation uncertainty in the analysis, as is the case of the techno-economic valuation of solar energy assets, the presented methodology can reach this objective with much less complexity and less demanding input data.

In order to assess the validity of the proposed methodology, the obtained PDF for the GHI of the month of October in the city of Rabat was employed in a Monte Carlo simulation, resulting in an average monthly GHI of 137.01 kWh/m^2. This value was then confronted to the most recent satellite data corresponding to the same city and month, i.e., the year 2020. A relative difference of 1.5% was obtained, thus evincing the goodness of the method for representing the stochastic pattern of the irradiation.

The results obtained in this paper can be used for future solar related projects. A monthly interval set of data can be useful for long-term studies, whereas a smaller interval is favourable when it comes for a more technical point of view. The reliability of the data is also important, as a low trustworthiness can lead to complications in a project assessment. Additionally, the method presented in this paper for determining the distribution that fits the best to a set of data is not limited to solar resource assessment, but can be used in other fields of study.

Author Contributions: Conceptualization, N.B. and H.M.; methodology, N.B., H.M. and J.d.l.H.; software, N.B.; validation, N.B., H.M. and J.d.l.H.; formal analysis, N.B. and H.M.; investigation, N.B., H.M. and S.C.; resources, N.B., H.M. and S.C.; data curation, N.B.; writing—original draft preparation, N.B.; writing—review and editing, H.M. and J.d.l.H.; visualization, N.B.; supervision, H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. PV Magazine. Solar Is Gaining Traction in MENA Region—But Plenty of Obstacles Remain. 2020. Available online: https://www.pv-magazine.com/2020/01/17/mesia-outlines-past-progress-future-promise-in-sweeping-look-at-solar-across-middle-east-and-north-africa/ (accessed on 29 September 2022).
2. EcoMENA. Countries Driving Solar Surge in MENA. 2019. Available online: https://www.ecomena.org/solar-surge-in-mena/ (accessed on 29 September 2022).
3. NREL Transforming Energy. Concentrating Solar Power Projects. Available online: https://solarpaces.nrel.gov/concentrating-solar-power-projects-country (accessed on 29 September 2022).
4. Masen | L'Agence Marocaine Pour l'Energie Durable. Available online: http://www.masen.ma (accessed on 29 September 2022).
5. Royaume du Maroc—Ministère de l'Energie, des Mines et de l'Environnement. Available online: https://www.mem.gov.ma/Pages/secteur.aspx?e=2 (accessed on 29 September 2022).
6. REN21—Building the Sustainable Energy Future with Renewable Energy. Available online: https://www.ren21.net (accessed on 29 September 2022).
7. Blaga, R.; Sabadus, A.; Stefu, N.; Dughir, C.; Paulescu, M.; Badescu, V. A current perspective on the accuracy of incoming solar energy forecasting. *Prog. Energy Combust. Sci.* **2019**, *70*, 119–144. [CrossRef]
8. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.-L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [CrossRef]
9. Kumler, A.; Xie, Y.; Zhang, Y. A Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique. *Sol. Energy* **2019**, *177*, 494–500. [CrossRef]
10. Wang, P.; van Westrhenen, R.; Meirink, J.F.; van der Veen, S.; Knap, W. Surface solar radiation forecasts by advecting cloud physical properties derived from Meteosat Second Generation observations. *Sol. Energy* **2019**, *177*, 47–58. [CrossRef]
11. Kallio-Myers, V.; Riihelä, A.; Lahtinen, P.; Lindfors, A. Global horizontal irradiance forecast for Finland based on geostationary weather satellite data. *Sol. Energy* **2020**, *198*, 68–80. [CrossRef]
12. Ohtake, H.; da Silva Fonseca, J.G., Jr.; Takashima, T.; Oozeki, T.; Shimose, K.-i.; Yamada, Y. Regional and seasonal characteristics of global horizontal irradiance forecasts obtained from the Japan Meteorological Agency mesoscale model. *Sol. Energy* **2015**, *116*, 83–99. [CrossRef]
13. Ruiz-Arias, J.A.; Gueymard, C.A. A multi-model benchmarking of direct and global clear-sky solar irradiance predictions at arid sites using a reference physical radiative transfer model. *Sol. Energy* **2018**, *171*, 447–465. [CrossRef]
14. Zempila, M.-M.; Giannaros, T.M.; Bais, A.; Melas, D.; Kazantzidis, A. Evaluation of WRF shortwave radiation parameterizations in predicting Global Horizontal Irradiance in Greece. *Renew. Energy* **2016**, *86*, 831–840. [CrossRef]

15. Fountoukis, C.; Martín-Pomares, L.; Perez-Astudillo, D.; Bachour, D.; Gladich, I. Simulating global horizontal irradiance in the Arabian Peninsula: Sensitivity to explicit treatment of aerosols. *Sol. Energy* **2018**, *163*, 347–355. [CrossRef]
16. Yang, D.; Sharma, V.; Ye, Z.; Lim, L.I.; Zhao, L.; Aryaputera, A.W. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy* **2015**, *81*, 111–119. [CrossRef]
17. Kim, C.K.; Kim, H.-G.; Kang, Y.-H.; Yun, C.-Y.; Kim, S.Y. Probabilistic prediction of direct normal irradiance derived from global horizontal irradiance over the Korean Peninsula by using Monte-Carlo simulation. *Sol. Energy* **2019**, *180*, 63–74. [CrossRef]
18. Zang, H.; Liu, L.; Sun, L.; Cheng, L.; Wei, Z.; Sun, G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew. Energy* **2020**, *160*, 26–41. [CrossRef]
19. Lopes, F.M.; Silva, H.G.; Salgado, R.; Cavaco, A.; Canhoto, P.; Collares-Pereira, M. Short-term forecasts of GHI and DNI for solar energy systems operation: Assessment of the ECMWF integrated forecasting system in southern Portugal. *Sol. Energy* **2018**, *170*, 14–30. [CrossRef]
20. Huang, J.; Rikus, L.J.; Qin, Y.; Katzfey, J. Assessing model performance of daily solar irradiance forecasts over Australia. *Sol. Energy* **2018**, *176*, 615–626. [CrossRef]
21. Pedro, H.T.; Coimbra, C.F.; David, M.; Lauret, P. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. *Renew. Energy* **2018**, *123*, 191–203. [CrossRef]
22. Huertas-Tato, J.; Aler, R.; Galván, I.M.; Rodríguez-Benítez, F.J.; Arbizu-Barrena, C.; Pozo-Vazquez, D. A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning. *Sol. Energy* **2020**, *195*, 685–696. [CrossRef]
23. Lauret, P.; Voyant, C.; Soubdhan, T.; David, M.; Poggi, P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Sol. Energy* **2015**, *112*, 446–457. [CrossRef]
24. Pereira, S.; Canhoto, P.; Salgado, R.; Costa, M.J. Development of an ANN based corrective algorithm of the operational ECMWF global horizontal irradiation forecasts. *Sol. Energy* **2019**, *185*, 387–405. [CrossRef]
25. Aguiar, L.M.; Pereira, B.; Lauret, P.; Díaz, F.; David, M. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renew. Energy* **2016**, *97*, 599–610. [CrossRef]
26. Mazorra Aguiar, L.; Pereira, B.; David, M.; Díaz, F.; Lauret, P. Use of satellite data to improve solar radiation forecasting with Bayesian Artificial Neural Networks. *Sol. Energy* **2015**, *122*, 1309–1324. [CrossRef]
27. Hejase, H.A.; Al-Shamisi, M.H.; Assi, A. Modeling of global horizontal irradiance in the United Arab Emirates with artificial neural networks. *Energy* **2014**, *77*, 542–552. [CrossRef]
28. Scolari, E.; Sossan, F.; Paolone, M. Irradiance prediction intervals for PV stochastic generation in microgrid applications. *Sol. Energy* **2016**, *139*, 116–129. [CrossRef]
29. Pedro, H.T.; Coimbra, C.F. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. *Renew. Energy* **2015**, *80*, 770–782. [CrossRef]
30. Tolba, H.; Dkhili, N.; Nou, J.; Eynard, J.; Thil, S.; Grieu, S. GHI forecasting using Gaussian process regression: Kernel study. *IFAC-PapersOnLine* **2019**, *52*, 455–460. [CrossRef]
31. Belmahdi, B.; Louzazni, M.; Akour, M.; Cotfas, D.T.; Cotfas, P.A.; El Bouardi, A. Long-Term Global Solar Radiation Prediction in 25 Cities in Morocco Using the FFNN-BP Method. *Front. Energy Res.* **2021**, *9*, 733842. [CrossRef]
32. Bounoua, Z.; Chahidi, L.O.; Mechaqrane, A. Estimation of daily global solar radiation using empirical and machine-learning methods: A case study of five Moroccan locations. *Sustain. Mater. Technol.* **2021**, *28*, e00261. [CrossRef]
33. Belmahdi, B.; Louzazni, M.; El Bouardi, A. A hybrid ARIMA–ANN method to forecast daily global solar radiation in three different cities in Morocco. *Eur. Phys. J. Plus* **2020**, *135*, 925. [CrossRef]
34. El Mghouchi, Y.; Chham, E.; Zemmouri, E.; El Bouardi, A. Assessment of different combinations of meteorological parameters for predicting daily global solar radiation using artificial neural networks. *Build. Environ.* **2018**, *149*, 607–622. [CrossRef]
35. Belmahdi, B.; Louzazni, M.; El Bouardi, A. One month-ahead forecasting of mean daily global solar radiation using time series models. *Optik* **2020**, *219*, 165207. [CrossRef]
36. Stoffel, T.; Renne, D.; Myers, D.; Wilcox, S.; Sengupta, M.; George, R.; Turchi, C. *Concentrating Solar Power: Best Practices Handbook for the Collection and Use of Solar Resource Data (CSP)*; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2010. [CrossRef]
37. HELIOSCSP. Steps for Solar Resource Assessment in Concentrated Solar Power Projects. 2018. Available online: http://helioscsp.com/steps-for-solar-resource-assessment-in-concentrated-solar-power-projects/ (accessed on 29 September 2022).
38. Renne, D.; George, R.; Wilcox, S.; Stoffel, T.; Myers, D.; Heimiller, D. *Solar Resource Assessment*; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2008. [CrossRef]
39. Mohandes, M.; Rehman, S.; Halawani, T. Estimation of global solar radiation using artificial neural networks. *Renew. Energy* **1998**, *14*, 179–184. [CrossRef]
40. Mohandes, M.; Balghonaim, A.; Kassas, M.; Rehman, S.; Halawani, T. Use of radial basis functions for estimating monthly mean daily solar radiation. *Sol. Energy* **2000**, *68*, 161–168. [CrossRef]
41. Mubiru, J.; Banda, E. Estimation of monthly average daily global solar irradiation using artificial neural networks. *Sol. Energy* **2008**, *82*, 181–187. [CrossRef]
42. Marchand, M.; Ghennioui, A.; Wey, E.; Wald, L. Comparison of several satellite-derived databases of surface solar radiation against ground measurement in Morocco. *Adv. Sci. Res.* **2018**, *15*, 21–29. [CrossRef]

43. Wahab, M.M.A.; El-Metwally, M.; Hassan, R.; Lefèvre, M.; Oumbe, A.; Wald, L. Assessing surface solar irradiance and its long-term variations in the northern Africa desert climate using Meteosat images. *Int. J. Remote Sens.* **2010**, *31*, 261–280. [CrossRef]
44. Aguiar, L.M.; Polo, J.; Vindel, J.; Oliver, A. Analysis of satellite derived solar irradiance in islands with site adaptation techniques for improving the uncertainty. *Renew. Energy* **2019**, *135*, 98–107. [CrossRef]
45. Urraca, R.; Gracia-Amillo, A.M.; Huld, T.; Martinez-De-Pison, F.J.; Trentmann, J.; Lindfors, A.V.; Riihelä, A.; Sanz-Garcia, A. Quality control of global solar radiation data with satellite-based products. *Sol. Energy* **2017**, *158*, 49–62. [CrossRef]
46. Schumann, K.; Beyer, H.G.; Meyer, R.; Chhatbar, K. Improving Satellite-Derived Solar Resource Analysis with Parallel Ground-Based Measurements. In Proceedings of the ISES Solar World Congress, Kassel, Germany, 28 August–2 September 2011.
47. Meyer, R.; Butron, J.T.; Marquard, G.; Schwandt, M.; Geuder, N.; Hoyer-Klick, C.; Lorenz, E.; Hammer, A.; Beyer, H.G. Combining Solar Irradiance Measurements and Various Satellite-Derived Products to a Site-Specific Best Estimage. In Proceedings of the SolarPACES Symposium 2008, Las Vegas, NV, USA, 4–7 March 2008. Available online: https://elib.dlr.de/53849/ (accessed on 29 September 2022).
48. Ineichen, P. Long Term Satellite Global, Beam and Diffuse Irradiance Validation. *Energy Procedia* **2014**, *48*, 1586–1596. [CrossRef]
49. Kariuki, B.W.; Sato, T. Interannual and spatial variability of solar radiation energy potential in Kenya using Meteosat satellite. *Renew. Energy* **2018**, *116*, 88–96. [CrossRef]
50. Plain, N.; Hingray, B.; Mathy, S. Accounting for low solar resource days to size 100% solar microgrids power systems in Africa. *Renew. Energy* **2019**, *131*, 448–458. [CrossRef]
51. Müller, R.; Pfeifroth, U.; Träger-Chatterjee, C.; Trentmann, J.; Cremer, R. Digging the METEOSAT Treasure—3 Decades of Solar Surface Radiation. *Remote Sens.* **2015**, *7*, 8067–8101. [CrossRef]
52. Huld, T.; Müller, R.; Gambardella, A. A new solar radiation database for estimating PV performance in Europe and Africa. *Sol. Energy* **2012**, *86*, 1803–1815. [CrossRef]
53. Palmer, D.; Koubli, E.; Cole, I.; Betts, T.; Gottschalg, R. Satellite or ground-based measurements for production of site specific hourly irradiance data: Which is most accurate and where? *Sol. Energy* **2018**, *165*, 240–255. [CrossRef]
54. Urraca, R.; Huld, T.; Gracia-Amillo, A.; Martinez-De-Pison, F.J.; Kaspar, F.; Sanz-Garcia, A. Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data. *Sol. Energy* **2018**, *164*, 339–354. [CrossRef]
55. El Mghouchi, Y.; El Bouardi, A.; Sadouk, A.; Fellak, I.; Ajzoul, T. Comparison of three solar radiation models and their validation under all sky conditions—Case study: Tetuan city in northern of Morocco. *Renew. Sustain. Energy Rev.* **2016**, *58*, 1432–1444. [CrossRef]
56. El Mghouchi, Y.; Ajzoul, T.; El Bouardi, A. Prediction of daily solar radiation intensity by day of the year in twenty-four cities of Morocco. *Renew. Sustain. Energy Rev.* **2016**, *53*, 823–831. [CrossRef]
57. AMEE | Adhérons à un Avenir Énergétique Durable. Available online: https://www.amee.ma/index.php?lang=en (accessed on 29 September 2022).
58. Direction de la Météorologie Nationale | Ministère de l'Equipement, du Transport, de la Logistique et de l'Eau Département de l'Eau. Available online: http://www.marocmeteo.ma (accessed on 29 September 2022).
59. Agence Nationale Pour le Développement des Energies Renouvelables et de l'Éfficacité Energétique. Règlement Thermique de Construction au Maroc (RTCM). Available online: https://www.amee.ma/sites/default/files/inline-files/Guide%20RTCM%20.pdf (accessed on 29 September 2022).
60. Google Maps. Available online: https://www.google.com/maps/@29.3231834,-7.7723258,5.5z (accessed on 29 September 2022).
61. PANGEA. BSRN Stations. Available online: https://www.pangaea.de/ddi?request=bsrn/BSRNEvent&format=html&title=BSRN+Stations (accessed on 29 September 2022).
62. European Commission. Photovoltaic Geographical Information System (PVGIS). Available online: http://re.jrc.ec.europa.eu/pvg_tools/en/tools.html (accessed on 29 November 2018).
63. Romero, S.M. CSP & Solar Resource Assessment. In: Bank TW. Washington DC, USA: Energy Sector Management Assistant Program. Available online: https://www.esmap.org/sites/default/files/esmap-files/ESMAP_IFC_RE_CSP_Training_World_Bank_Romero.pdf (accessed on 29 September 2022).
64. Gueymard, C.A. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renew. Sustain. Energy Rev.* **2014**, *39*, 1024–1034. [CrossRef]
65. Freeman, J.B.; Dale, R. Assessing bimodality to detect the presence of a dual cognitive process. *Behav. Res. Methods* **2012**, *45*, 83–97. [CrossRef]
66. Hartigan, J.A.; Hartigan, P.M. The Dip Test of Unimodality. *Ann. Stat.* **1985**, *13*, 70–84. Available online: http://links.jstor.org/sici?sici=0090-5364%28198503%2913%3A1%3C70%3ATDTOU%3E2.0.CO%3B2-A (accessed on 29 September 2022). [CrossRef]
67. Machine Learning Mastery. A Gentle Introduction to Probability Density Estimation. 2020. Available online: https://machinelearningmastery.com/probability-density-estimation/ (accessed on 29 September 2022).
68. Kim, T.Y.; Park, B.U.; Moon, M.S.; Kim, C. Using bimodal kernel for inference in nonparametric regression with correlated errors. *J. Multivar. Anal.* **2009**, *100*, 1487–1497. [CrossRef]
69. *Engineering Statistics Handbook. 1.3.5.14. Anderson-Darling Test*; National Institute of Standards and Testing (NIST): Gaithersburg, MD, USA, 2013. [CrossRef]

70. Stephanie, G. Rayleigh Distribution: Definition, Uses, Mean, Variance. Statistics How to. 2014. Available online: https://www.statisticshowto.com/rayleigh-distribution/ (accessed on 29 September 2022).
71. Weisstein, E.W. "Student's t-Distribution". From MathWorld—A Wolfram Web Resource. Available online: https://mathworld.wolfram.com/Studentst-Distribution.html (accessed on 29 September 2022).
72. Hernández, J.A.; Trujillo, C.L.; Vallejo, W.A. Modelamiento de la irradiancia y la temperatura ambiente utilizando funciones de probabilidad. *Rev. Tecnura* **2014**, *18*, 128–137. [CrossRef]