

Article

Classification of Daily Irradiance Profiles and the Behaviour of Photovoltaic Plant Elements: The Effects of Cloud Enhancement

Isabel Santiago * , Jorge Luis Esquivel-Martin, David Trillo-Montero, Rafael Jesús Real-Calvo and Víctor Pallarés-López

Department of Electronic and Computer Engineering, Campus de Rabanales, University of Córdoba, 14071 Córdoba, Spain; jorgel.esquivelm@gmail.com (J.L.E.-M.); ma2trmod@uco.es (D.T.-M.); el1recar@uco.es (R.J.R.-C.); el1palov@uco.es (V.P.-L.)

* Correspondence: el1sachi@uco.es; Tel.: +34-957-218699

Abstract: In this work, the automatic classification of daily irradiance profiles registered in a photovoltaic installation located in the south of Spain was carried out for a period of nine years, with a sampling frequency of 5 min, and the subsequent analysis of the operation of the elements of the installation on each type of day was also performed. The classification was based on the total daily irradiance values and the fluctuations of this parameter throughout the day. The irradiance profiles were grouped into nine different categories using unsupervised machine learning algorithms for clustering, implemented in Python. It was found that the behaviour of the modules and the inverter of the installation was influenced by the type of day obtained, such that the latter worked with a better average efficiency on days with higher irradiance and lower fluctuations. However, the modules worked with better average efficiency on days with irradiance fluctuations than on clear sky days. This behaviour of the modules may be due to the presence, on days with passing clouds, of the phenomenon known as cloud enhancement, in which, due to reflections of radiation on the edges of the clouds, irradiance values can be higher at certain moments than those that occur on clear sky days, without passing clouds. This is due to the higher energy generated during these irradiance peaks and to the lower temperatures that the module reaches due to the shaded areas created by the clouds, resulting in a reduction in its temperature losses.

Keywords: irradiance daily profiles; unsupervised machine learning; clustering algorithm; performance ratio; photovoltaic module efficiency; inverter efficiency; cloud enhancement; irradiance enhancement



Citation: Santiago, I.; Esquivel-Martin, J.L.; Trillo-Montero, D.; Real-Calvo, R.J.; Pallarés-López, V. Classification of Daily Irradiance Profiles and the Behaviour of Photovoltaic Plant Elements: The Effects of Cloud Enhancement. *Appl. Sci.* **2021**, *11*, 5230. <https://doi.org/10.3390/app11115230>

Academic Editor: Manuela Sechilariu

Received: 28 April 2021

Accepted: 1 June 2021

Published: 4 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Renewable energy sources, which are clean and inexhaustible resources that are used for electricity generation through different technologies, are currently playing a crucial role in the race to achieve some of the Sustainable Development Goals of the United Nations [1], such as Goal 7, which aims to ensure access to affordable, reliable, sustainable and modern energy for all. The target set by the International Renewable Energy Agency (IRENA) is to produce 57% of the world's energy from renewable sources by 2030, up from 26% today, more than doubling the existing share, so this decade is considered to be the one in which renewable energies will drive the global energy transformation [2]. Even more ambitious are the objectives proposed by European countries through the European Green Deal, whose objective, in the fight against the threat of climate change and environmental degradation, is that there will be no net emissions of greenhouse gases by 2050 [3].

Within renewable energies, a wide range of typologies can be found, and the one that currently has the highest growth rate is photovoltaic (PV) energy [4]. In 2019, this was the most installed energy source among both renewables and non-renewable energies, accounting for 40% of new global capacity [4–6]. Globally, new PV capacity in 2019 was 115 GW, an increase of 12% compared to 2018, giving a cumulative total of 627 GW. These

figures were obtained due to the growth that took place on all continents, highlighting the increase in some countries such as China (30 GW), USA (13 GW), India (10 GW) and Japan (7 GW). In the European Union, PV energy saw very significant growth, with 16.7 GW installed, an increase of 104 % compared to 2018 (8.2 GW), the strongest growth since 2010 [4,7,8]. This increase in the installed capacity of renewable energies, and in particular the increase in PV power, has allowed around a third of the world's electricity generation to take place without fossil fuels and therefore with zero emissions [4].

With this rate of expansion, PV technology, which has come to be considered the world's fastest growing power generation source and a driving force behind energy transition in some countries, is expected to further increase in installed capacity in the near future [4,7,9,10]. In this regard, renewable energies in general, and PV in particular, have in recent years considerably increased their role in the generation mix of different countries, and their presence has led to changes in grid operation strategies [6]. These policies must continue to evolve in the future so that renewables energies form the basis of electricity generation and of the transition towards greater decarbonisation of sectors such as the demand for building power, air conditioning and transport [3,11–13].

Nowadays, the main disadvantage of renewable energy sources is their manageability or dispatchability, as the energy they produce is conditioned by the randomness of the weather, giving rise to rapid variations in their production [14]. If electricity generated from renewable sources cannot be stored at a reasonable cost, its penetration in the grid is limited, and the share of renewable energy in the generation mix will be even more difficult to increase.

For a better manageability of solar PV installations, which enables an increase in their penetration in the power grid, it is important to know in advance the electrical energy that is to be produced, in order to be able to predict more accurately how the product can be delivered in the time intervals considered in the daily and intraday energy markets. The need for better forecasting has for years increased interest in various solutions, and different authors have used different techniques [15–21], such as time series prediction with statistical and learning methods [22], or atmospheric modelling with the analysis of sky satellite images and the study of cloud propagation [23]. All these models are based on meteorological and/or historical production data. Among all of them, the categorisation of solar irradiance data is one of the earliest applied techniques [24], and in the literature there are many authors who have analysed and predicted the behaviour of irradiance profiles, using various procedures [24–27].

In order to characterise daily irradiance profiles, several authors define a series of indices, such as the clearness index, which represents the proportion of measured irradiance and theoretical irradiance under clear sky conditions. There is also the variability or intermittency index, or the fractal dimension [28–30], which are parameters that take into account the deviation of the actual irradiance values from clear sky conditions. For their characterisation, the irradiance data recorded over time can be used directly or, alternatively, a transformation of these temporal data to the frequency domain can be performed beforehand [24]. All these types of data can be used to make classifications, in order to extract different types of days, and group those types of daily profiles with similar behaviour. Different authors use different methods to make these classifications, considering a greater or a lesser number of types of days [24,25,31–37].

Although obviously the production of a PV plant at each moment in time will depend on the amount of irradiance received by the solar panels and the temperature at which they operate, the meteorological conditions that occur throughout the day, such as the variation in ambient temperature and wind speed, as well as cloud cover and the speed at which clouds are moving, generally determine the operation conditions and the behaviour of the plant's components, so that the efficiency and yield values of both the modules and the inverters are modified throughout the day, and from one day to the next [38,39]. This means that the ability of the components of the PV system to convert the received irradiance into electricity is not constant, but is influenced by the type of day that occurs.

In this context, one of the main objectives of this work was to apply a classification algorithm to a nine-year history of records of the global irradiance received by the PV modules of a solar installation, in order to automatically find patterns or types of days. Once the days were classified according to the daily irradiance profiles, the dependence between the patterns of days found and the operation and performance of the PV plant components were also analysed, which is one of the main contributions of this paper. To obtain the different types of days, a clustering method was used, within the unsupervised automatic machine learning techniques, making use of the existing tools in the Python programming language. The method used in this work stands out for its simplicity and ease of application, and requires almost no changes to the irradiance data measured at the PV plant, so no data filtering or transformation are required [24]. If the weather forecast is known in advance, it will be possible to know beforehand the type of day that is going to take place, and from the results found, it will be possible to know what the plant's behavioural interval will be, which will obviously influence its production. Once the behaviour of the installation components were obtained on each type of day, another of the aims of this work was to show the reasons for the variation in this behaviour on different types of days. To this end, an analysis of the operation of its elements on clear sky days was carried out, comparing it with that obtained on days with a greater presence of fluctuations.

On these non-stable days there may be peaks of irradiance higher than the values registered on clear sky days, due to the phenomenon known as cloud enhancement (CE) or irradiance enhancement (IE) [40], and the presence of these events was also analysed in this paper. According to Jarvela et al. [41], although the CE phenomenon is well known, its effects on actual PV systems has not been thoroughly studied.

The rest of the paper is organised as follows. Section 2 presents the methodology employed and the characteristics of the data used in the study, Section 3 shows and discusses the results obtained, and Section 4 shows the conclusions extracted from them.

2. Materials and Methods

2.1. Types of Algorithms Used for the Classification

In this work, the classification of daily global irradiance profiles was carried out using machine learning algorithms, which is the science of programming computers so that they can learn or extract information from data, without being explicitly programmed. Within machine learning, we are going to use a technique belonging to the so-called unsupervised learning, which in general refers to methods in which the model is adjusted to observations. Contrary to what happens in the so-called supervised learning techniques, in this case there are no training data made up of input and output data, which enable us to provide a priori the solution to the problem. The unsupervised learning techniques do not need an "expert" intervention that provides known output data to the algorithm, offering a prior knowledge of the result to be obtained, but only has input data, and the model is able to find hidden structure in its inputs without knowledge of outputs [15,35]. Within these unsupervised learning techniques, clustering algorithms were applied, whose objective is to divide the data set into different groups of elements with similar characteristics, in such a way that the objects grouped in the same group or cluster have similar characteristics and that these characteristics are also different from those of the objects in the other groups.

The classification algorithm chosen, from among the many existing ones, was the so-called k-means. This one, developed by MacQueen in 1967, is a well-known and commonly used partitional clustering algorithm [15,32,42]. Although it is not very well suited to irregularly shaped groups, it was chosen because it is well suited to a large number of samples, which is the case for the data used in this paper. This clustering method has the advantage of creating averaged groups, while other clustering algorithms (such as hierarchical clustering) tend to perform better in identifying the outliers [43]. For its application it is required that the number of clusters, k , must be specified in advance.

This k-means method groups the data by trying to separate the samples into k groups of equal variance, minimising within each group the criterion known as inertia or sum of

squares. The algorithm divides a number n of samples x , into a number of distinct groups (k), and each of these groups is described by the mean μ_j of the samples in the group. This average is known as the centre or centroid of the cluster, and generally they do not have to be points of x , but must lie in the same space. The algorithm aims to choose centres that minimise the inertia or sum of squares within the cluster, which is the Euclidean distance, and is governed by the following expression,

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2), \quad (1)$$

In general terms, the algorithm has three fundamental steps. The first step is based on choosing k samples from the data set n , which would be the cluster centres at the beginning. After initialisation, the algorithm consists of a loop between the next two steps. The second step assigns each data item x to its nearest centre, while the third step creates new centres by taking the average of the values of all samples assigned to the previous centre. The algorithm will repeat steps two and three until the centres do not change significantly.

For the initialisation of the centres, the k-means++ method was used in this work, which distributes the initial centres so that they are generally distant from each other, in order to try to prevent the algorithm from converging to a local, and therefore not optimal, minimum. In this way, better results are obtained than by random initialisation.

When performing point-to-point classifications, the Euclidean distance is used as the base metric (Equation (1)), but for grouping time series into similar groups, this type of metric is not the most appropriate. In this work, a distance metric dedicated to time series, the so-called Dynamic Time Warping (DTW), was used, which is an algorithm for measuring the similarity between two temporal linear sequences. DTW is an algorithm for finding an optimal alignment between two sequences and a useful distance metric, which possesses the capacity to cope with changes in data over time. The DTW Barycentre Averaging (DBA) algorithm minimises the sum of the squared DTW distance between the barycentre and the series in the group. As a result, the centroids have an average shape that mimics the shape of the group members, regardless of where temporal changes occur among the samples [44–47].

Given two time series $x = (x_0, \dots, x_n)$ and $y = (y_0, \dots, y_m)$, with lengths n and m respectively, the distance DTW from x to y is thus stated as an optimisation problem, such that,

$$DTW(x, y) = \min_{\pi} \sqrt{\sum_{(i,j) \in \pi} d(x_i, y_j)^2}, \quad (2)$$

where d is the Euclidean distance between two points in the series and $\pi = [\pi_0, \dots, \pi_k]$ is a path that satisfies a number of properties. On the one hand, $\pi_k = (i_k, j_k)$ represents pairs of indices such that $0 \leq i_k < n$ and $0 \leq j_k < m$. As boundary conditions $\pi_0 = (0, 0)$ and $\pi_k = (n - 1, m - 1)$ are considered. As a continuity condition, it is stated that for all $k > 0$, if $\pi_k = (i_k, j_k)$ and $\pi_{k-1} = (i_{k-1}, j_{k-1})$, it has to be fulfilled that $i_{k-1} \leq i_k \leq i_{k-1} + 1$ and $j_{k-1} \leq j_k \leq j_{k-1} + 1$. Therefore, the path is a temporal alignment of time series such that Euclidean distance between aligned or resampled time series is minimal [48].

In this way, the k-means clustering method applied to time series, in this case applied to daily irradiance profiles, using the DTW metric, created groups of irradiance daily profiles with similar shapes, and the centres or centroids of the clusters were calculated as barycentres or average sequences of a group of irradiance time series considered within the same group.

2.2. Software Used to Apply the Classification

To implement the above algorithm, the Python programming language was used. It is a general-purpose, freely distributed, interpreted, multi-paradigm, multi-platform and dynamically typed language. It owes its popularity to the large number of libraries and functions available, with a great technical support thanks to a large community of users.

Anaconda, which is a Python distribution, was used for its management, and it was used in the Windows environment. The main Python library used was Scikit-learn, which contains state-of-the-art machine learning algorithms, as well as exhaustive documentation on each one [49,50]. This library contains the clustering algorithms used in this work, which are available in its module `sklearn.clusters`. `ts-learn` was also used, which is a general-purpose machine learning Python library for time series, as is the case of the data analysed in this work, and from which the modules `tslearn.utils` y `tslearn.clustering` [48] were mainly used. The latter module offers the option of using DTW as the central metric in the *k-means* algorithm, leading to better clusters and centroids in the case of this study. Scikit-learn also relies on other Python packages that were used, such as NumPy, which is a library for scientific computing, necessary for processing data in the form of arrays [51]; SciPy, which has a collection of functions for scientific computing and on which Scikit-learn is based to implement its algorithms [52]. Pandas, the library needed to manage, modify and operate the files and data processed, was also used [53], and for the graphical representation of the results of the algorithm, the library Matplotlib was used [54].

2.3. Monitored Data and Specifications of the Analysed PV System

A set of 36 PV modules connected to an inverter were the elements used to carry out the analysis shown in this paper. They belong to a PV installation located on the rooftop of a building in the city of Cordoba ($37^{\circ}53'00''$ N, $4^{\circ}46'00''$ O), in the south of Spain. The modules, from BP, are made of polycrystalline silicon and have a peak power of 165 W. Their characteristics are listed in Table 1. The modules are configured as three strings in parallel, with 12 modules connected in series in each of them, and they are mounted on an open frame with 30° of inclination and 18° of azimuth. The inverter to which the modules are connected is from SMA, model SMC-100, and its specifications are shown in Table 2.

Table 1. General characteristics of the installation and its PV modules.

	Specifications
Number of modules per inverter	$36 = 12 \times 3$
Module model	BP-3165
PV module peak power	165 W
Number of cells per module	72
Surface area of each module	$1,573,790 \text{ mm}^2$
PV cell surface	$125 \text{ mm} \cdot 125 \text{ mm}$
Useful module area	1125 m^2
Power of modules per inverter	5940 W
Power temperature coefficient	$-(0.5 \pm 0.005) \% / ^{\circ}\text{C}$
Temperature coefficient of I_{sc}	$(0.065 \pm 0.015) \% / ^{\circ}\text{C}$
Temperature coefficient of V_{oc}	$-(160 \pm 20) \text{ mV} / ^{\circ}\text{C}$
NOCT	$47^{\circ}\text{C} \pm 2^{\circ}\text{C}$
Open-circuit voltage (V_{oc})	44.2 V
Voltage at maximum power (V_{max})	35.2 V
Short-circuit current (I_{sc})	5.2 A
Current maximum power (I_{max})	4.7 A

Table 2. Inverter characteristics.

	Specifications
Inverter model	SMA SMC-5000
Maximum performance	96%
European performance	>95.1%
Maximum input current	26 A
Maximum dc voltage	600 V
Maximum dc power	5750 W
PV voltage range, MPP	246–480 V
Number of MPPT	1
Maximum output current	26 A
Maximum nominal ac power	5000 W
Maximum ac power	5500 W

The data acquisition system of the PV installation is integrated in the inverter. The measured parameters were stored with a frequency of 5 min, and they correspond to an average value of each parameter in this time interval.

The irradiance values analysed in this work, which correspond to the global tilted irradiance (GTI), were collected by an equipment called Sunny SensorBox, from the company SMA, to which the sensor of this parameter was connected, which was a calibrated PV ASI amorphous cell, with a measurement range of 0 to 1500 W/m², an accuracy of $\pm 8\%$ and a resolution of 1 W/m². This irradiance sensor was installed next to the PV modules, with the same orientation and inclination, so that the recorded irradiance values would represent those received by the PV panels. During the nine years of operation, a degradation of the cell was detected, resulting in a decrease of 1.8% in the recorded irradiance data, which is a percentage lower than the accuracy given by the sensor manufacturer. In addition, this equipment also collected measurements from two temperature sensors. One of them, a PT-100M sensor, with a range of -20 to $+110$ °C, was used to measure the temperature of the modules. Its accuracy and its resolution specified by the manufacturer were ± 0.5 °C and 0.1 °C, respectively. The other, a PT-100M-NR sensor, was used for measuring the ambient temperature, which had the same characteristics as the previous one, but with an accuracy of ± 0.7 °C given by the manufacturer. Assuming effects of decoupling and stability for these sensors, the error in temperature measurements can be considered ± 1 °C.

The SensorBox and the inverter were connected via RS485 to a device called Sunny-Webbox from SMA. It is a device with a central communications unit that continuously collects all the data from the inverters and the SensorBox and is a multifunctional, energy-saving data logger. The measurement data were transmitted via a GSM modem from remote locations where there was no telephone or ADSL connection, so it was read directly from a computer, and connected to the network, consuming between 4 and 12 W of power.

The data were recorded at the PV facility over 9 years, during the period from 1 January 2011 to 4 February 2020, which means a total of 3322 days. For each full day the first record took place at 00.00 h and the last at 23.55 h, resulting in 288 samples daily, which were the time series to be classified in this work.

2.4. Characterisation of the Behaviour of the PV System Components

In order to know the performance of the PV installation for each type of irradiance profile, the values of the efficiency of the panels, the efficiency of the inverters, and the performance ratio (PR) of the set formed by modules and inverter were determined for each record. These parameters were calculated by using the data recorded by the inverter, corresponding to the current and voltage at the inverter input, and the power at the inverter

output, which were also recorded during the indicated period with the same recording frequency of 5 min.

The value of the efficiency of the PV modules is given by the expression:

$$\eta_G = (E_{dc}/E_r) \cdot 100, \quad (3)$$

where E_{dc} is the power generated by the PV modules associated with each inverter, which is determined by the expression:

$$E_{dc} = I_{dc} \cdot V_{dc} \cdot t, \quad (4)$$

where I_{dc} is the direct current from the modules, measured in A, V_{dc} is the voltage at the input of the inverter from the modules, measured in V, and t is the time corresponding to the monitoring interval expressed in hours. E_r is the reference energy given by:

$$E_r = E_G \cdot A_G, \quad (5)$$

where E_G is the in-plane irradiation of the PV panels expressed in Wh/m² and given by the product of the irradiance (G_i) multiplied by the monitoring time (t), and A_G the area of PV panels associated with each inverter.

The efficiency of the inverter is determined by the expression:

$$\eta_{inv} = (E_{ac}/E_{dc}) \cdot 100, \quad (6)$$

where E_{ac} is the energy obtained at the output of the inverter, given by the expression:

$$E_{ac} = P_{ac} \cdot t, \quad (7)$$

in which P_{ac} is the output power of the inverter and t is the time, which corresponds to the monitoring interval expressed in hours.

Also determined was the parameter known as Performance Ratio (PR), by means of the expression:

$$PR = Y_f/Y_r, \quad (8)$$

where Y_f , named final yield, are given by the expression:

$$Y_f = E_{ac}/P_0, \quad (9)$$

where E_{ac} is the ac power at the output of the inverter and P_0 is the nominal power at standard conditions STC (AM1.5, 1 kW/m², 25 °C) of the PV modules associated to the inverter.

Y_r is the reference yield, given by the expression:

$$Y_r = E_G/G^*, \quad (10)$$

where E_G is the in-plane total irradiation and $G^* = 1000$ W/m² is the irradiance at standard conditions STC.

PR is a dimensionless quantity that corresponds to the ratio of the ac energy generated compared to the value of the energy that would be generated with the same irradiation in a system without losses. It would be calculated for the set formed by the inverter and the modules associated with it and gives a value that corresponds to the operation of this set.

The average daily values of these parameters were determined for the entire period analysed and grouped for each of the types of days found through the classification process, in order to determine the relationship between the daily irradiance profiles of each type and these three parameters that represent the efficiency of the plant's components.

For the analysis of some specific days given as an example, the losses in the inverter were studied, which are given by the following expression:

$$L_s = Y_A - Y_f, \quad (11)$$

where, as already seen, Y_f is the final yield and Y_A is the array yield,

$$Y_A = E_{dc}/P_0, \quad (12)$$

which represents the time, measured in h/d, that PV generator must be operating with nominal power P_0 to generate E_{dc} energy, reflecting the real operation of the PV generator compared to its nominal capacity.

The temperature losses of the modules, which are considered as the system losses due to PV modules are operating above 25 °C, are given by

$$L_{ct} = Y_r - Y_T, \quad (13)$$

where Y_r is the name reference yield and Y_T is the temperature corrected reference yield, given by the expression

$$Y_T = Y_r \cdot [1 - c_T(T_c - T_0)], \quad (14)$$

where c_T is the module temperature coefficient (% °C⁻¹), T_0 is the reference temperature 25 °C, and T_c is the PV cell temperature (°C), given by:

$$T_c = T_{amb} + (NOCT - 20) \cdot (G_i/800), \quad (15)$$

where T_{amb} is the ambient temperature, G_i is the in-plane irradiance (W/m²) and $NOCT$ is the nominal operating cell temperature. All these parameters are described in more detail in a previous paper [55].

2.5. Procedure Used to Apply the Clustering Algorithms

The pre-processing carried out on the data was to eliminate from the analysis those days in which the records are null, due to some type of failure in the monitoring system or due to disconnection for maintenance work. In this regard, data from 3237 of the total of 3322 days comprising the monitoring interval were available for the study.

The classification methodology used consisted of two types of groupings. Firstly, the set of 3237 days was divided, using the *k-means* algorithm, into three categories, taking into account the value of the total daily irradiance as a classification index. Based on this value, the days were classified into three types. Type H, M and L, corresponding respectively to high, medium and low total daily irradiance. As indicated above, this classification was carried out using the *sklearn.clusters* module of the Scikit-learn library. As previously indicated, the initialisation of the centroids was carried out using the *k-means++* method, and the Euclidean metric was used.

Once this classification was carried out, a second clustering process was made within each of the three categories obtained. For this second classification, the variations in the instantaneous irradiance values recorded throughout the day were taken into account. To do so, the value of the irradiance at each moment in time was subtracted from its value immediately before, and with the absolute value of these differences, a set of time series was obtained for each day, with 287 records. It is an index which represents the variability of the irradiance that takes place throughout the day. These variations in irradiance values are due to the passage of clouds, and give rise to random variations in the PV plant output. Then, the series corresponding to each of the three groups of days of categories H, M and L made previously, were classified again into three other categories. This second classification corresponds to days of low, medium and high irradiance variability, denoted by 1, 2 and 3, respectively. Thus, the total of the 3237 days finally analysed will be automatically grouped into a total of nine categories denoted as H1, H2, H3 (days with high total daily irradiance, and with low, medium and high variability of irradiance, respectively); M1, M2, M3 (days with medium values of total daily irradiance, and with low, medium and high variability of irradiance, respectively); and L1, L2, L3 (days with low total daily irradiance, and with low, medium and high variability of irradiance respectively). This set of nine types of days allows the variability of existing daily irradiance profiles to be gathered, but in a general

way, so that, based on the time of year and weather forecasts, it can be easy to predict in advance the type of day that is going to occur.

As this second classification is not a point-to-point classification, but a time series classification, the TimeSeriesKMeans algorithm, which is a variation of the original *k-means*, adapted to time series clustering, was now employed. This classification was performed with the DTW metric described above. The `tslearn.clustering` module belonging to the `ts-learn` library was used for this second clustering process.

3. Results and Discussion

Figure 1 shows the total irradiance values received on the plane of the PV modules, which were recorded every 5 min during the entire period analysed in this work, from January 2011 to February 2020. There are two periods, in 2012 and 2016, where, due to failures in the monitoring system, no data records were available, and which, as indicated above, were excluded from the classification process. In the colour map represented in Figure 1, it can be seen that the maximum daily irradiance values logically occur in the central hours of the day, between 12 p.m. and 4 p.m., and these values are higher in the summer months. It can also be seen how the irradiance values decrease from the maximum values in the central hours of the day, towards the hours of sunrise and sunset, with obviously zero irradiance in the night hours. The greater number of hours of sunshine in the summer months in the geographical location where the measurements were taken is evident. It can also be observed how, on days with passing clouds, there are variations or fluctuations in the irradiance values throughout part or all the day. The daily irradiance profiles with non-zero values shown in this figure are therefore those that were used for the classification carried out in this work.

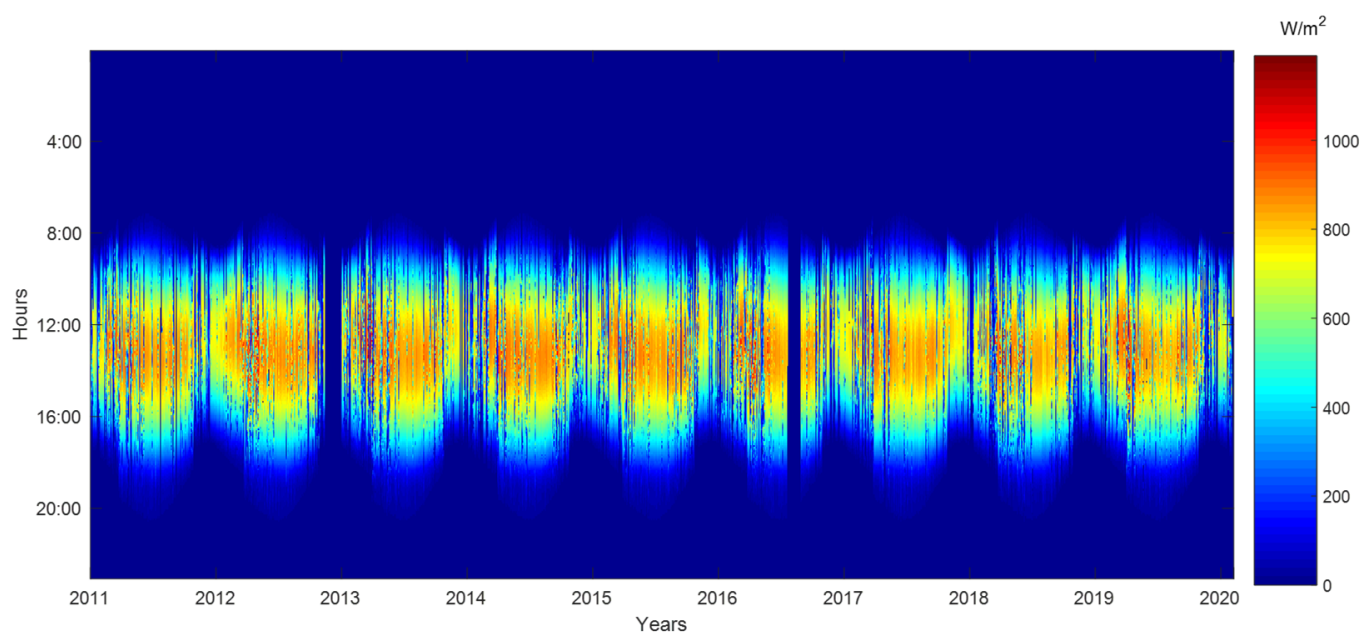


Figure 1. Instantaneous in-plane irradiance values registered from 2011 to 2020 with a frequency sample of 5 min.

Figure 2 shows a histogram in which the total daily irradiance values for each of the days shown in Figure 1 can be observed. It can be seen that, influenced by the greater or lesser number of daily hours of sunshine throughout the year, and by the meteorological conditions, the total daily irradiance received by the panels in the geographical location analysed in this work can vary from values of less than 10 to values greater than 80 kW/m² per day. The red dashed lines show the values of total daily irradiance received by the panels that were obtained by the *k-means* algorithm to classify the analysed days in the three types named low (L), medium (M) and high (H).

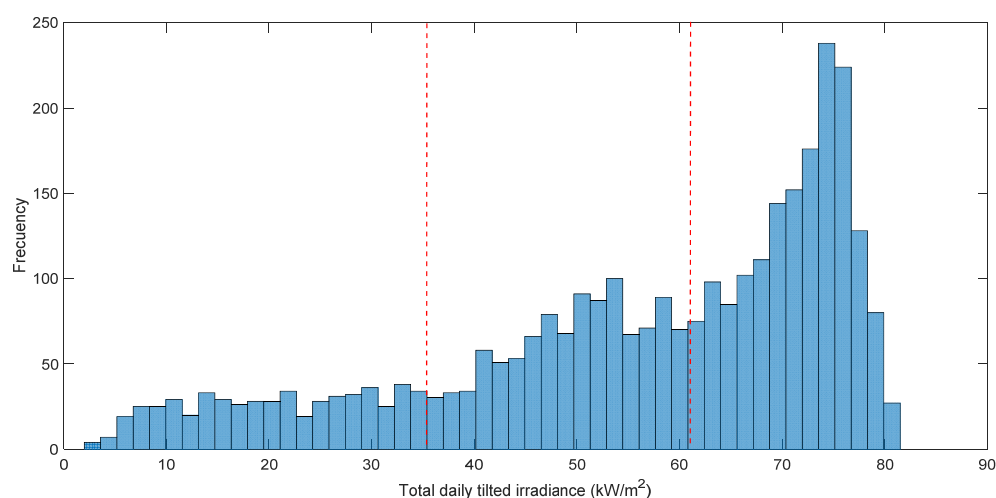


Figure 2. Histogram corresponding to the daily total values of irradiance recorded on the plane of the PV modules during the period 2011–2020.

Figure 3 shows, for each of the days of the period analysed, the average daily efficiency values of the set of 36 PV modules studied, connected to the same inverter, and organised, as mentioned above, into three parallel strings, each with 12 modules connected in series. The efficiency values of the modules represented in this figure were obtained by averaging for each day the efficiency values (η_G) determined by Equation (3), which were calculated every 5 min. It can be seen that, for the days with the irradiance profiles shown in Figure 1, the set of panels worked on average with efficiencies ranging from around 8% to values of over 15%, so that this daily operating parameter of the modules can present variations of up to 8%.

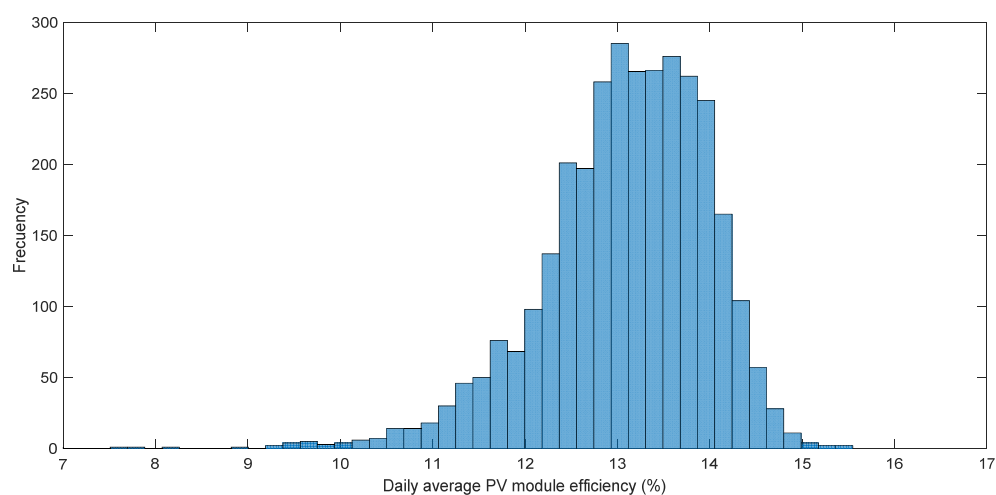


Figure 3. Histogram corresponding to the daily average efficiency values of PV modules during the period 2011–2020.

With regard to the operation of the inverter, the daily average efficiency values for the period of days analysed are shown in the histogram in Figure 4. The efficiency values of the inverter represented in this figure were obtained by averaging for each day the efficiency values (η_{inv}) determined by Equation (6), which were calculated every 5 min. It can be seen that, in this case, the inverter under study can operate with an efficiency ranging from values of around 75% to values close to 95%, although the values with the highest incidence were above 90%. The European weighted performance of the inverter, given by its manufacturers, is 95.1%, and 96% is the maximum efficiency, as reflected in Table 2. These values were not reached in Figure 4, since the daily average values of this parameter

are shown. However, efficiency values below 90% indicate that the inverter is working below its proper operating state.

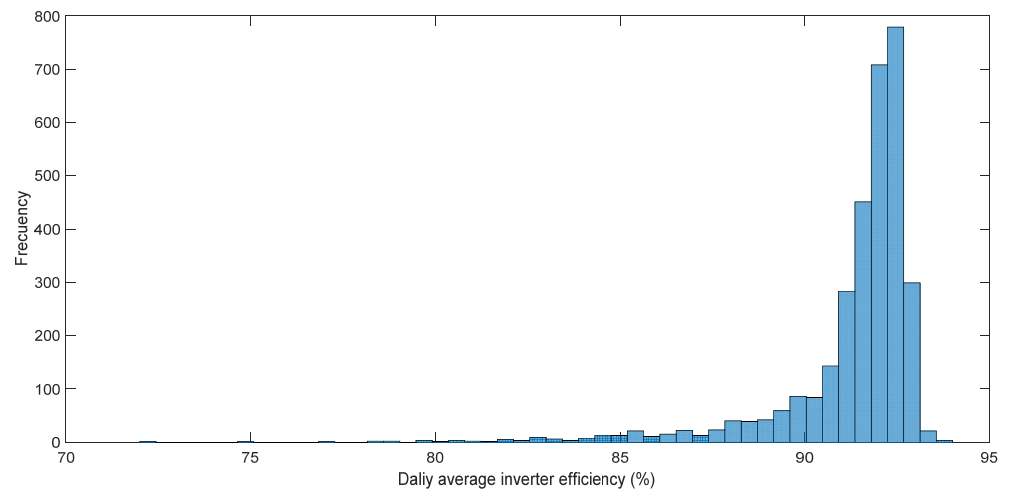


Figure 4. Histogram of daily average values of inverter efficiency over the period 2011–2020.

And in the case of the so-called performance ratio (PR) (Equation (8)), the daily average values for the period analysed are shown in the histogram in Figure 5. It can be seen that it varies between values above 0.3 and values very close to 1, although for the data set represented, the daily average PR values that determine the interquartile range ($Q1$ – $Q3$) are 0.79 and 0.87, and 0.83 is the median of those data.

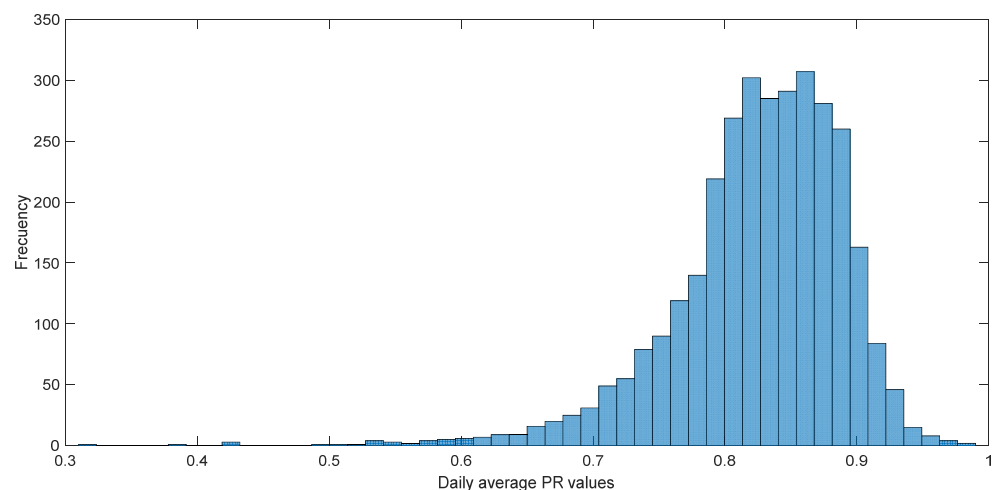


Figure 5. Histogram corresponding to the daily average values of the performance ratio (PR) for the inverter together with the associated modules over the period 2011–2020.

It can be seen, then, that once the adopted configuration of the modules and the inverter associated with them is fixed by the design and dimensioning of the set, their operation presents margins of variation, which shows how these components respond to changes in the meteorological conditions. In this way, the energy production will not be exclusively proportional to the irradiance received, but the variation in the meteorological conditions affects the operation performance of the PV installation components, which can lead to a variation in the expected production.

With respect to the first classification carried out in this work, obtained by applying the k -means algorithm to all the total daily irradiance values of the 3237 days studied, the results obtained are shown in Figure 6, where the number of days obtained in each of

the three types is represented, as well as the values of the total daily in-plane irradiance received in the modules for each of these days. Of the total number of days analysed, 550 days were obtained belonging to the low type (which represents 17.0% of the total number of days analysed), 1043 corresponded to the medium category (which represents 32.2%), and 1644 corresponded to the high category (which represents 50.8%). The total daily irradiance received on the modules on L days was between 2.5 and 35.8 kW/m², with an average value of 21.3 kW/m². For M days it was between 35.8 and 61.2 kW/m², with an average value of 50.4 kW/m², and for H days, the range of the total daily irradiance values varied between 61.2 and 81.9 kW/m², with an average value of 72.1 kW/m². This implies that on L days, the total in-plane irradiation received by the PV panels was between 0.2 y 3.0 kWh/m² (an average value of 1.8 kWh/m²); between 3.0 and 5.1 kWh/m² (an average value of 4.2 kWh/m²) on M days; and between 5.1 and 6.8 kWh/m² (an average value of 6.0 kWh/m²) on H days. Of the total irradiance or irradiation received by the panels during the nine years of monitoring considered in this work, 6.4% would have been received on L days, 29.2% would have been received on M days, and 64.4% would have been received on H days.

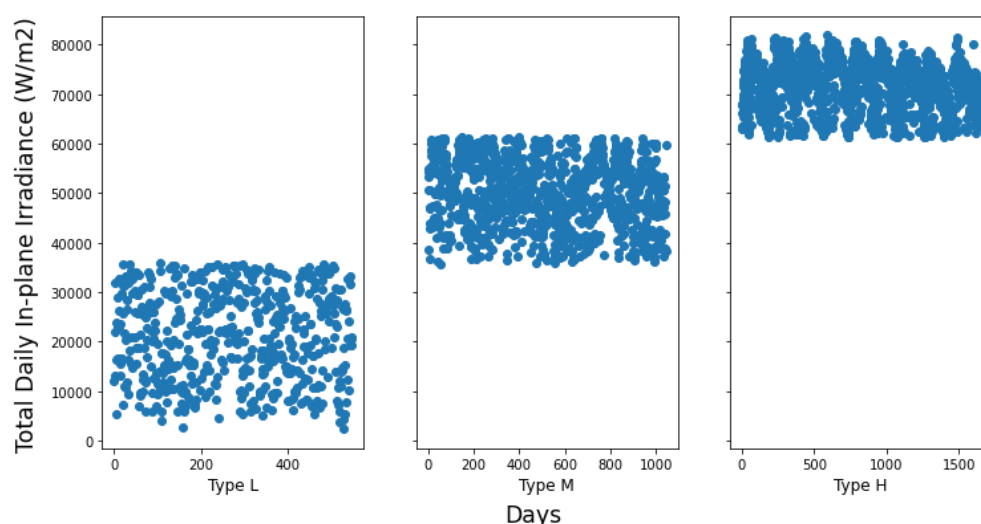


Figure 6. Results of the first classification of the days according to the total daily irradiance value using the k-means algorithm.

Once this first classification was carried out, the second classification was applied, from which a total of nine types of days or clusters were obtained, with different characteristics from one to another, in terms of the amount of irradiance and its variation or fluctuation throughout the day.

A sample of the daily irradiance profiles obtained in each of the nine different types of days or clusters is shown in Figure 7. The centroids corresponding to each of the nine clusters obtained are shown in Figure 8. In addition, the values of total in-plane daily irradiance, for each of the days obtained in each of the nine subcategories, are shown in Figure 9, where the number of days obtained in each one is also shown on x-axes. The mean values of total daily irradiance for days H1, H2 and H3 were, respectively, 72.4, 72.0 and 67.8 kW/m². The corresponding values for days M1, M2 and M3 were 51.5, 48.5 and 47.8 kW/m², and those of days L1, L2 and L3 were of 12.9, 25.6 and 27.9 kW/m², respectively. In the case of L days, there was a larger variation in the average value of the total daily irradiance value from L1 to L2 and L3 days. However, in the case of M days the average total daily irradiance values presented by the three groups were similar to each other, slightly higher on days with smaller fluctuations, and lower when there are more fluctuations; the same behaviour was found for H days. It was observed that the days within the three L-type categories, which are those with the lowest irradiance values, corre-

sponded to days with passing clouds. There were hardly any days with daily irradiance values in the range belonging to L days that corresponded to clear sky days, so that in all three types, including L1, the passage of clouds and, therefore, fluctuations of irradiance values, were the most frequent situation, resulting in these low daily irradiance values.

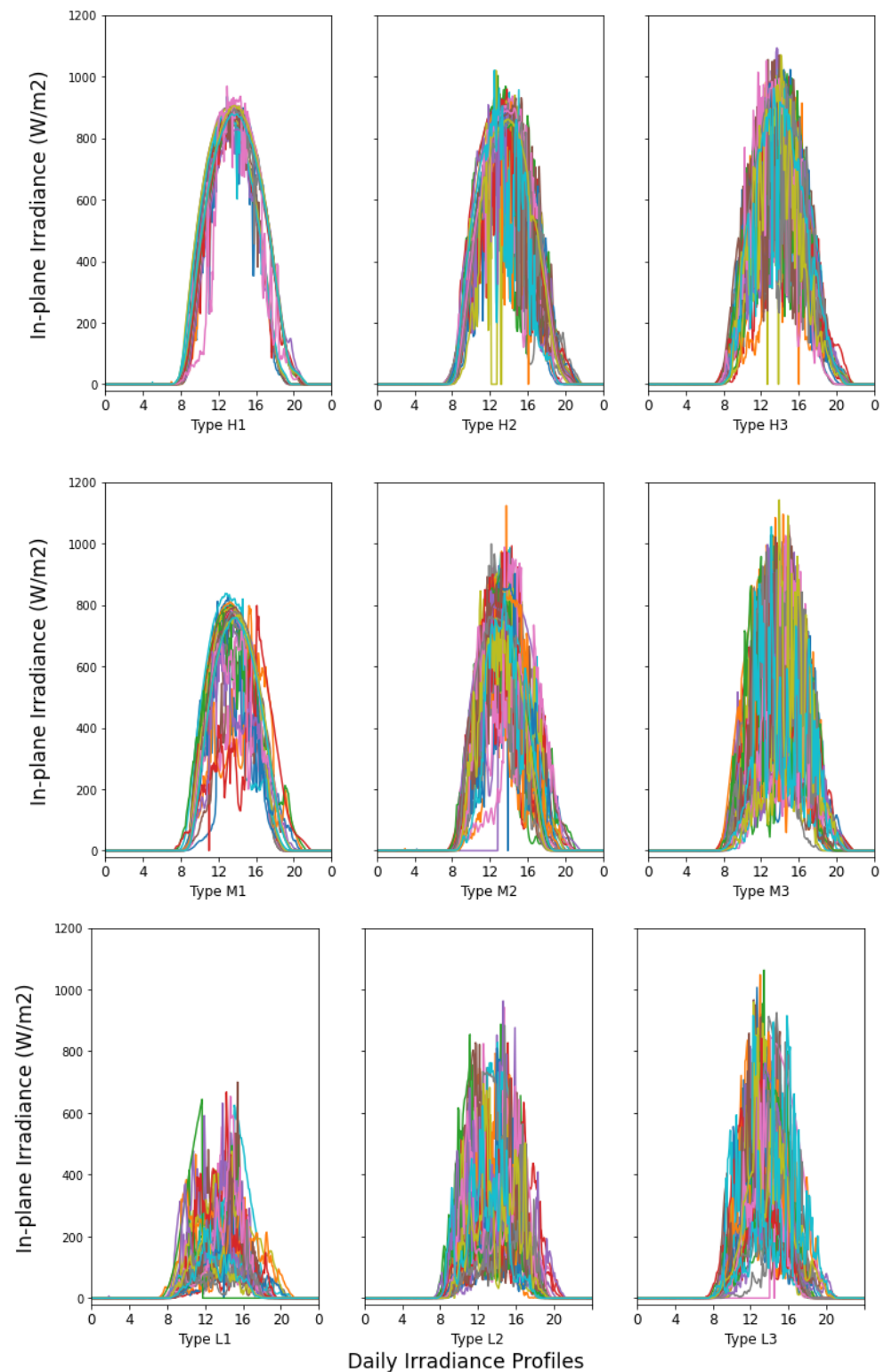


Figure 7. Sample of daily irradiance profiles obtained in the clusters L1, L2, L3, M1, M2, M3, H1, H2 and H3.

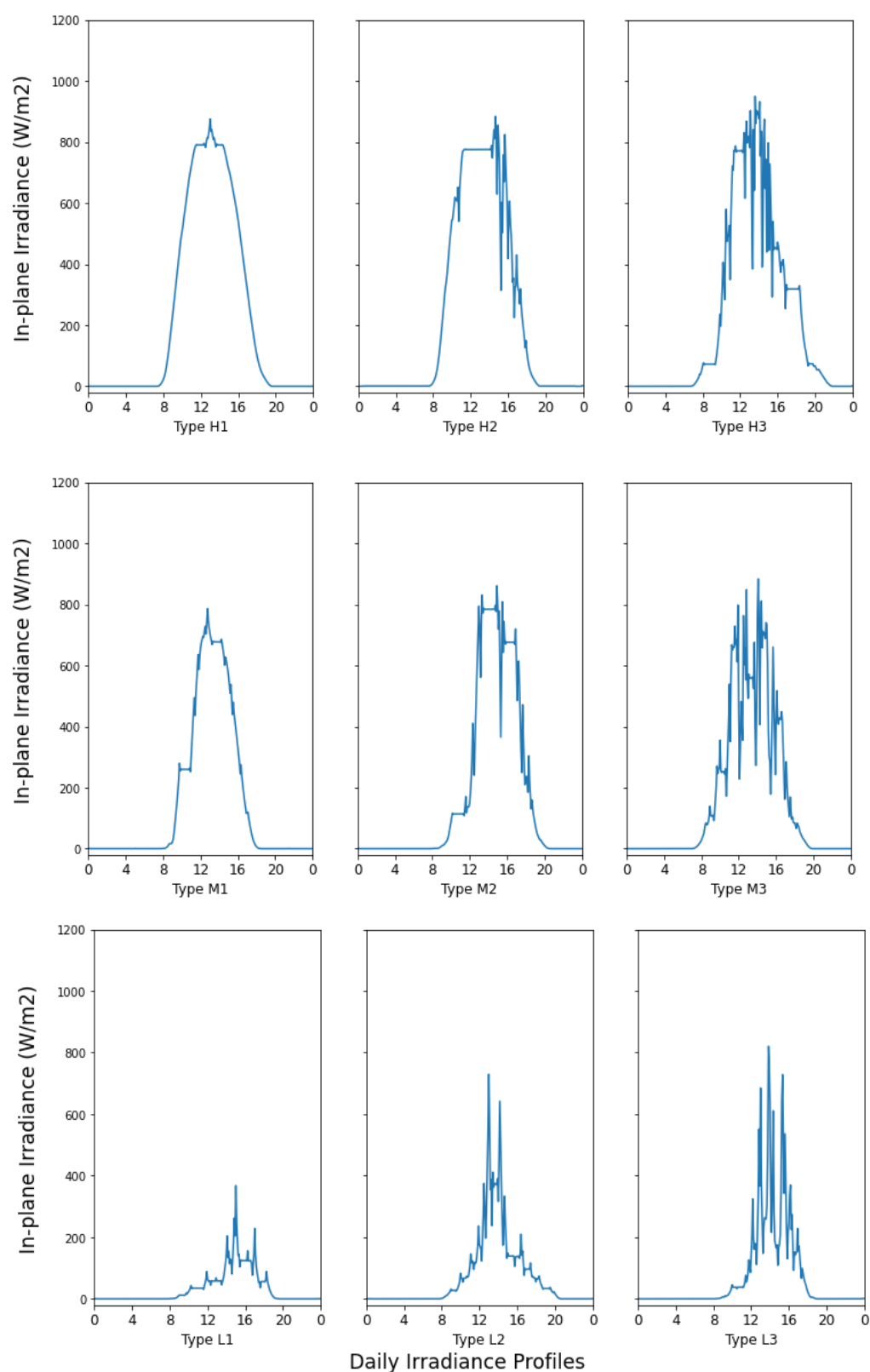


Figure 8. Centroids of the daily irradiance profiles obtained in the clusters L1, L2, L3, M1, M2, M3, H1, H2 and H3.

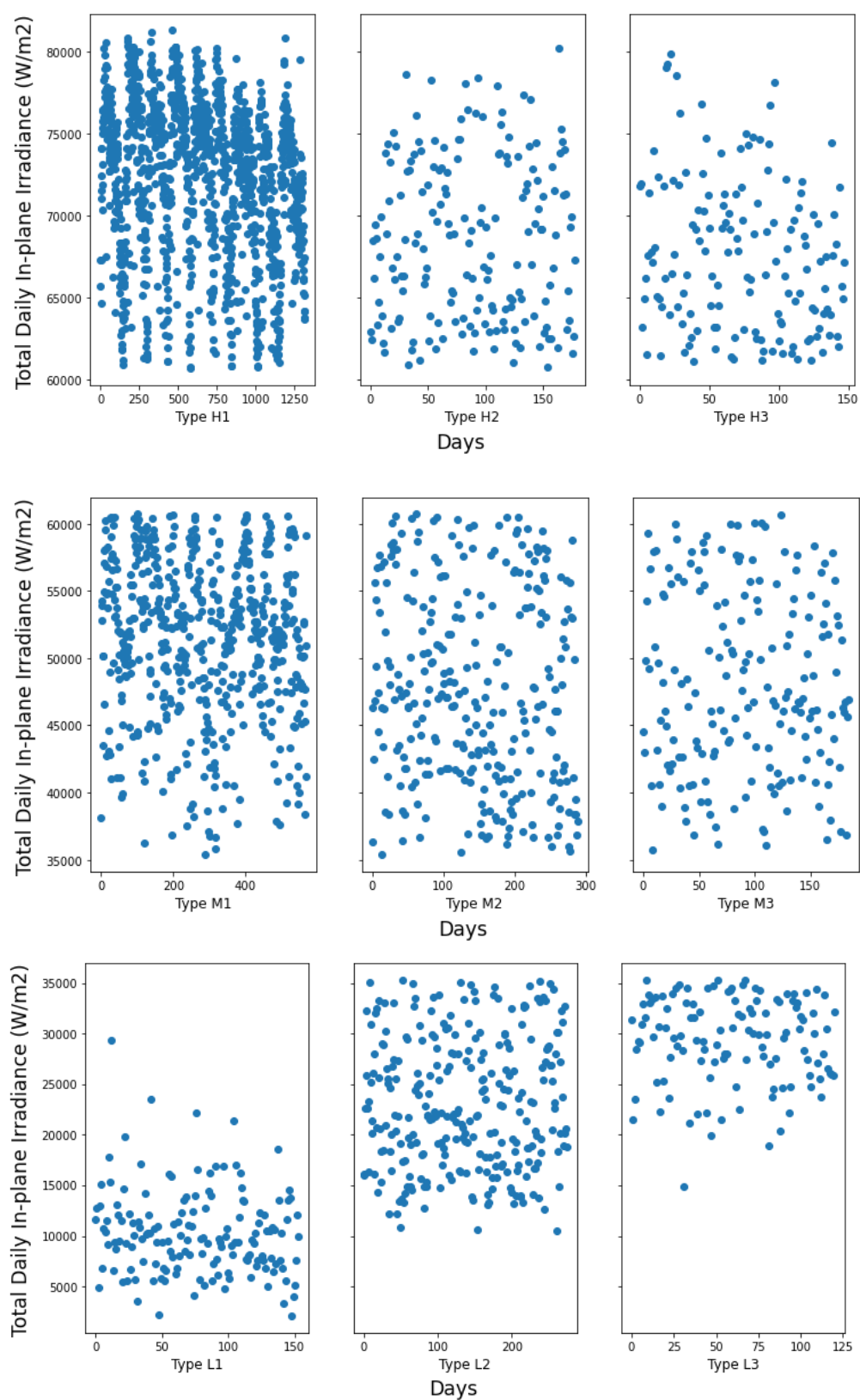


Figure 9. Total in-plane daily irradiance and number of days obtained in the clusters L1, L2, L3, M1, M2, M3, H1, H2 and H3.

With respect to the type M or days with medium values of the total daily irradiance, the days with smaller fluctuations due to the lower cloud passage are within the M1 subgroup. In addition, in this M1 group, some clear sky days were obtained, which corresponded to the irradiance received by the panels in January and December, winter days in this geographical location, with a lower number of hours of sunshine than during the rest of the year, which, despite being clear sky, give rise to values of total daily irradiance in the range of the named medium days. In the M2 and M3 categories, as a result of the classification, the medium days which presented greater fluctuations in the irradiance values throughout the day, due to the passage of clouds, logically obtained the highest fluctuations in the M3 cluster.

Within the H clusters, with higher values of total daily irradiance, the days that the algorithm considered to be within the H1 cluster corresponded to clear sky days or days with less cloud passage, with fluctuations increasing on H2 type days, and to a greater extent on H3 type days. The type of day with the highest incidence in this geographical location corresponded to H1 days, which account for 40.2% of the total of 3237 days analysed.

Comparing M days with H days, it can be seen in Figure 7 how, in the irradiance profiles of H days, the daily sunshine hours were somewhat higher (with sunset taking place later and sunrise taking place before) than those of M days. Furthermore, it can also be seen that, in the irradiance profiles of M days, the passage of clouds resulted in lower irradiance values than in the fluctuations due to the passage of clouds on H days, with these values being even lower in the profiles of L days. Thus, it can be considered that the *k-means* algorithm correctly carried out the classification of the daily irradiance profiles according to the total daily irradiance values and the fluctuations of the daily irradiance throughout the day. The applied technique can be considered to be robust and easy to apply [24,25].

Once the classifications were made, the performance of the PV system elements was determined for each of the days obtained in the nine clusters. The results obtained for the average daily *PR* value, corresponding to the operation of all modules together with the inverter, are shown in Figure 10.

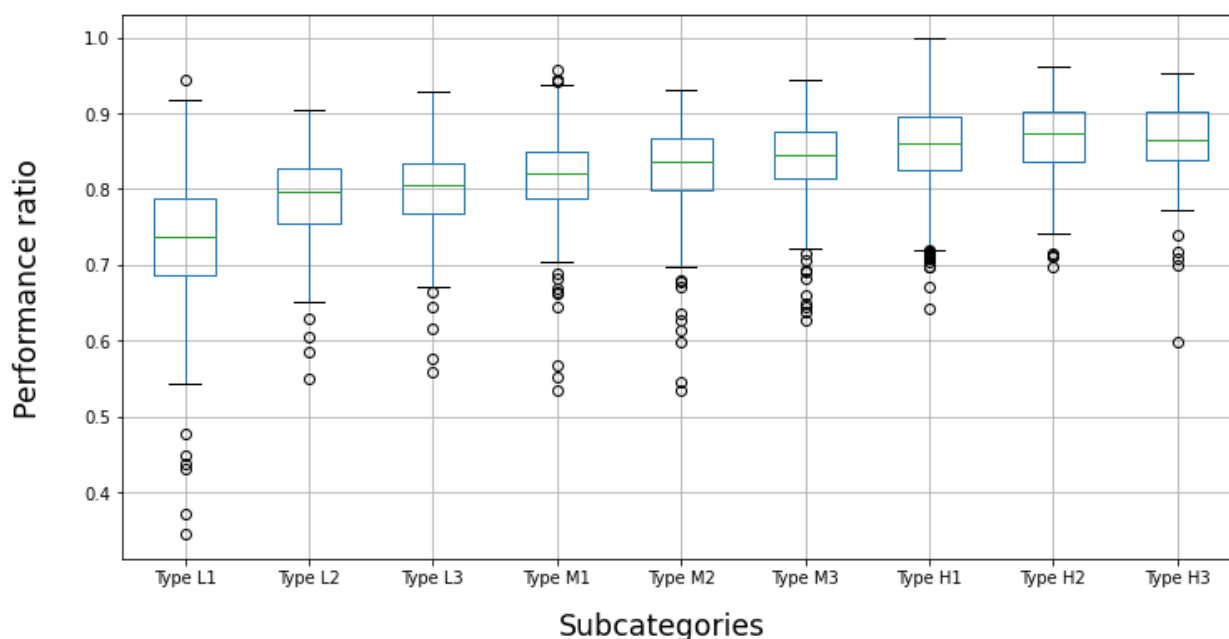


Figure 10. Average daily performance ratio (*PR*) of the set of 26 PV modules together with the inverter in the clusters L1, L2, L3, M1, M2, M3, H1, H2 and H3.

It can be seen that there is a relationship between the type of day and the performance with which the set of panels and the inverter works, ranging from mean values slightly below 0.75 for L1 days, which are those with the lowest performance, to mean values close to 0.9 for H2 days, which are the types of days on which there is a higher average *PR* value. The greatest dispersion of *PR* values within the different groups occurs on L1 days, indicating that a greater heterogeneity of the days selected by the algorithm was obtained within this group.

With regard to the behaviour of the set of modules and inverter, it can be observed that on H days, the performance of their operation was similar on the three types of days, regardless of the irradiance fluctuations that occurred. Within the M days, the performance of M3 and M2 days were similar to each other, despite the differences in the fluctuations of the daily irradiance profiles in these two groups, and both were slightly higher than the performance of M1 days. The same was true for L days, in which the *PR* values of the L2 and L3 days were similar to each other, and higher than those of the L1 days, which corresponded to lower irradiance values.

It is noteworthy that H1 and M1 are not the days with the best average *PR* within the categories of H and M days respectively, but it was found that days with passing clouds, and therefore with fluctuations in production, had better average performance ratio than clear sky days or days with fewer transitions. To see the reason for this behaviour, Figure 11 shows the average daily efficiency of the PV modules obtained for the days on each group.

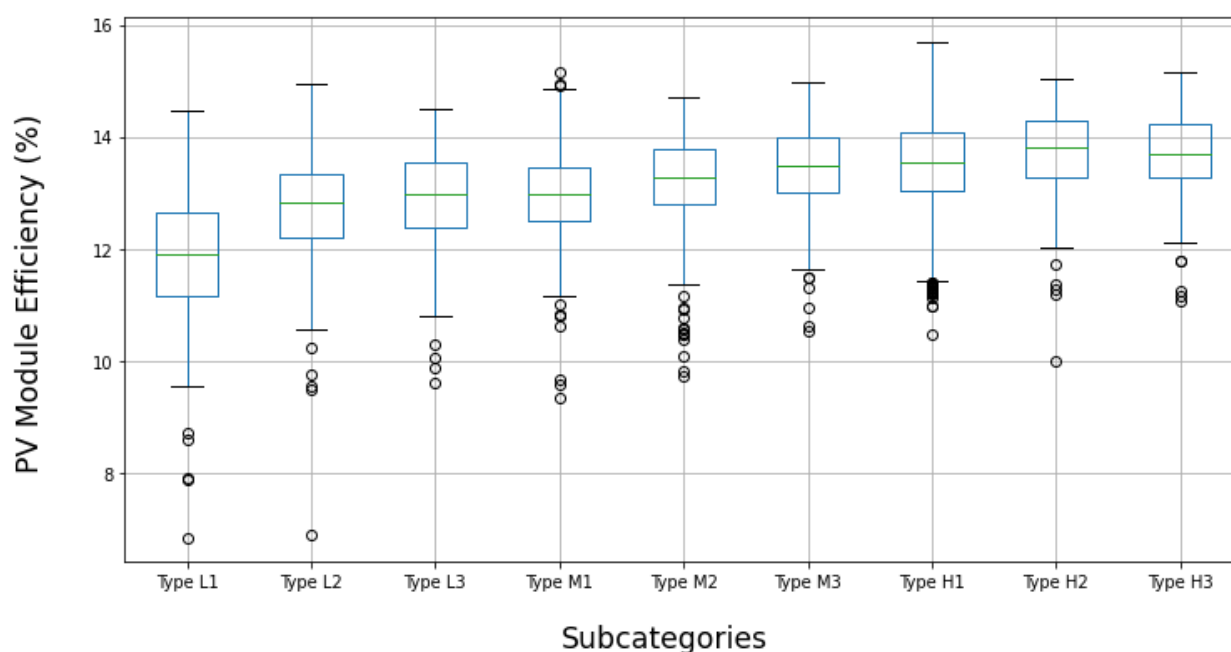


Figure 11. Average daily efficiency of PV modules in the clusters L1, L2, L3, M1, M2, M3, H1, H2 and H3.

Figure 11 shows that, although the production of energy in the modules is proportional to the irradiance received by them, the efficiency with which the modules transform solar radiation into electricity varied depending on the type of day, such that, on L1 days, modules worked with an average efficiency of 12%, and on H2 days, which are those with the best average efficiency of the panels, had average values close to 14%. It can be observed that in this H2 group, there was less dispersion of the average efficiency of the module from one day to another within the same group. It can be seen that, in general, the days with the lowest module efficiency corresponded to L days. Within the subcategories, the efficiency improved from L1 to L2 and L3, as the total daily irradiance increased. Within these types of days, the variability or dispersion of the results within the same group was

greater, which indicates, as already mentioned, the heterogeneity of patterns that may exist within these subcategories. The behaviour of the efficiency of the PV modules improved for days with medium irradiance. However, it is worth noting that within the M1, M2 and M3 groups, there was no better performance on the days with lower fluctuations (M1), but, contrary to what might be expected at first, on M3 days, which had higher fluctuations, the average efficiency was higher, even taking into account that, for the three types of days, M1, M2 and M3, the total daily irradiance is within the same range of values (Figure 9). With respect to the behaviour of the modules on H days, the average efficiency value was slightly higher than on M days, but a similar behaviour appeared with respect to the presence of fluctuations. Although the best efficiency results, above 15%, were found on some H1 days, in general, the average module efficiency was higher on H2 days, which had more fluctuations in irradiance, than on H1 days. The best average efficiency value was found on H2 days, as on H3 days there was a small decrease, but the average efficiency was still slightly higher on H3 than on H1 days.

It is difficult to make a comparison of the results obtained with those found in previous research, since irradiance profiles are highly dependent on geographical location. Moreover, the number of clusters considered varies from one author to another [24,36]. Witkoff et al. [39] performed a study dividing the days also into nine categories—although clustering techniques were not used—and obtained the best *PR* values of the installation on the cloudiest or overcast days. However, the meteorological conditions throughout the year in the city of Singapore, where the study was carried out, are different from those of south Spain, where the installation analysed in this work is located. Maafi and Harrouni [30] performed a fractal modelling of daily solar irradiances and used it to classify daily irradiances registered in the desert of Algeria. For this geographical location they considered only three types of days, and they also found that the *PR* values were slightly higher on partly and completely cloudy days, than clear sky days. However, the authors only showed *PR* data and did not comment on this behaviour.

The better average module efficiency of days with higher irradiance fluctuations compared to clear sky days or days with lower fluctuations can be related to the phenomenon called CE, also named over-irradiance or IE, which has been analysed by numerous authors [40,41,56–60]. As it can be seen in Figures 7 and 8, on partly cloudy days, at some moments in time, the irradiance values can exceed the expected values of the clear sky irradiance or the theoretical global irradiance values for that time of the year on a day with a cloudless atmosphere. This increase in the irradiance values is due to reflection phenomena caused by cloud edges surrounding the unobstructed solar disk. Diffuse irradiance may be enhanced due to the reflection of solar irradiance from the base of the clouds and from the scattering of direct irradiance due to cloud particles [61]. On the other hand, Yordanov et al. [60] stated that the CE phenomenon is mainly due to strong forward Mie scattering inside the clouds, and that the strongest CE events occur when a narrow gap is surrounded by thin clouds within 5° around the solar disk [41,56]. The CE phenomenon can be observed all around the world, from the equator to regions to the north and south, and from sea level to high altitudes, and such periods of enhanced solar irradiance may last from several seconds to some minutes, depending on the cloud motion [58].

The presence of these higher irradiance values in the PV installation analysed in this work is shown in more detail in Figure 12, where the daily irradiance profiles during 2011 are represented. It can be seen that, on days with passing clouds, there are fluctuations in the irradiance values and this parameter reaches values at some moments in time higher than on clear sky days at the same time of the year. It is shown that this phenomenon is observed throughout the year. For total irradiance values on a tilted plane (30°) and records every 5 min, it can be seen that the irradiance values recorded can be up to 1.25 times higher than the values corresponding to clear sky days. Although only the values for one year are shown in this figure, in order to show the image in greater detail, this type of effect is observed during all the years corresponding to the monitoring period of the data processed in this work.

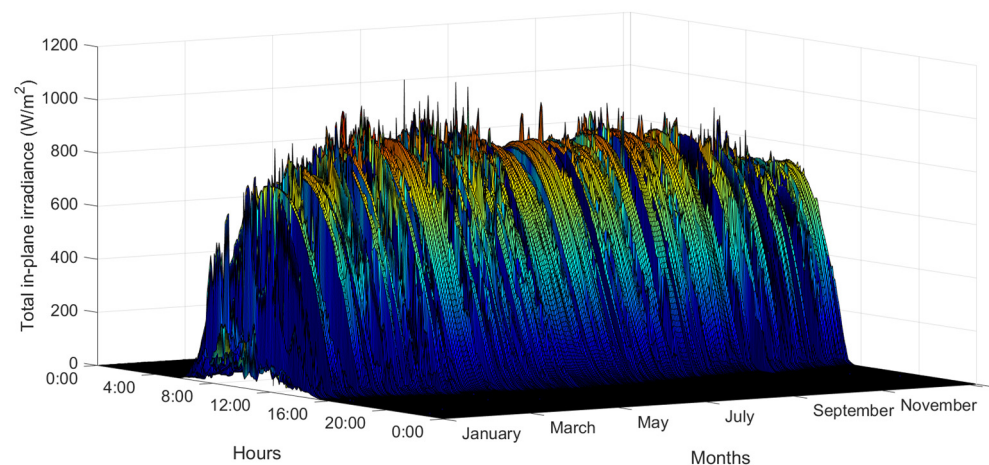


Figure 12. Daily irradiance profiles registered, every 5 min, from January to December 2011.

These higher irradiance values resulted in higher production in the PV panels at those specific moments in time when the irradiance is higher. As an example, Figure 13 shows the I_{dc} current produced in the group of 36 PV panels on 12, 14 and 17 May 2011, while Figure 14 shows the irradiance received on the modules on those days, both parameters recorded every 5 min. Day 14 corresponds to a clear day with high irradiance (type H1), while days 12 and 17 show high irradiance values and fluctuations (type H2). The presence of the CE phenomenon is evident in Figure 14. Therefore, it can be observed that, as the irradiance increases at these times, so does the current produced by the PV panels.

In addition to the high global irradiance that may occur at these moments in time, it can be observed that lower ambient and module temperatures are recorded on these days, as shown in Figures 15 and 16, due to cloud cover between the peaks of increased irradiance. Due to the thermal mass of the PV panel, the change in operating temperature is slow in comparison to the changes in irradiance, so if there has previously been shading by clouds, there may be a situation where there is a peak in irradiance but at that moment, there is no temperature increase equivalent to the situation where there is no cloud cover. This shading and temperature reduction results in a decrease in temperature losses in the PV panel output (Figure 17).

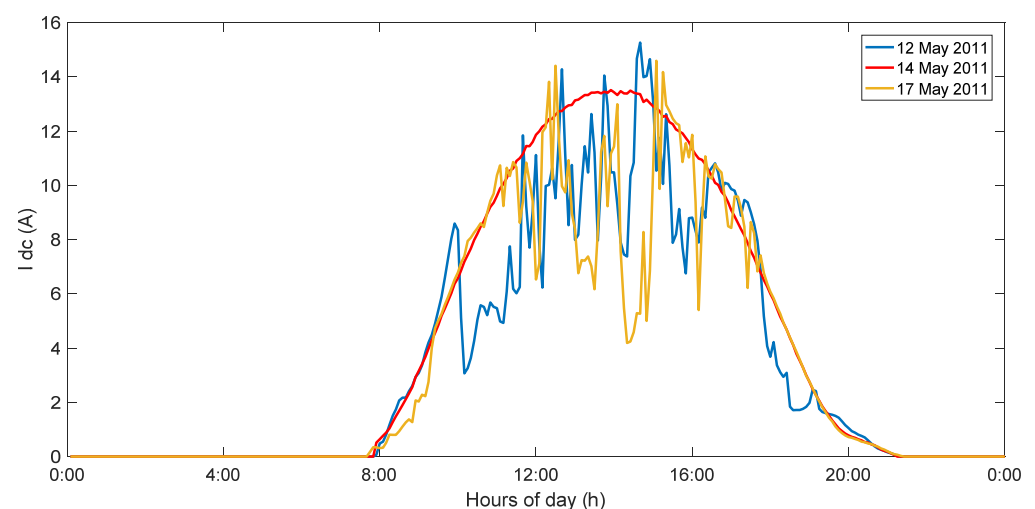


Figure 13. Daily profiles of the direct current, I_{dc} , produced in the set of 36 PV modules during 12, 14 and 17 May 2011.

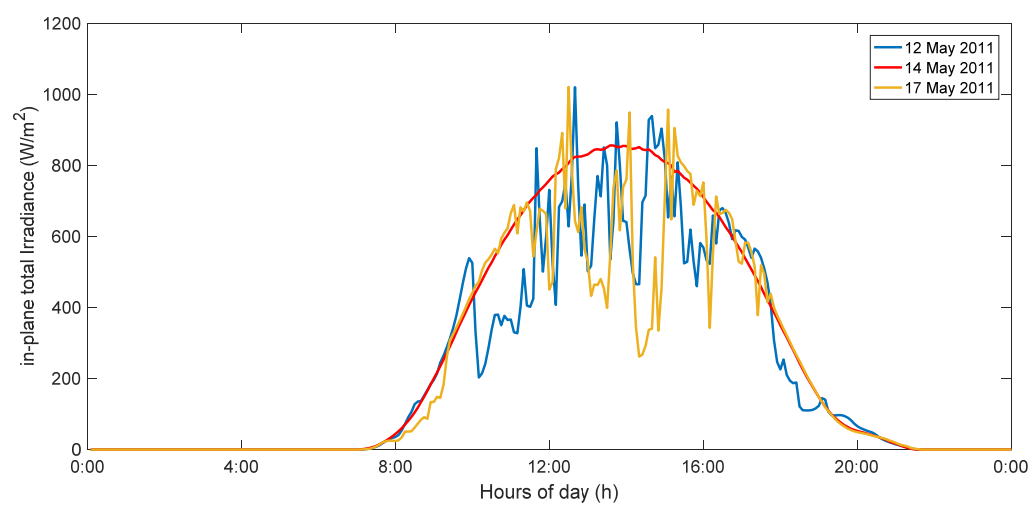


Figure 14. Daily profiles of total in-plane irradiance during 12, 14 and 17 May 2011.

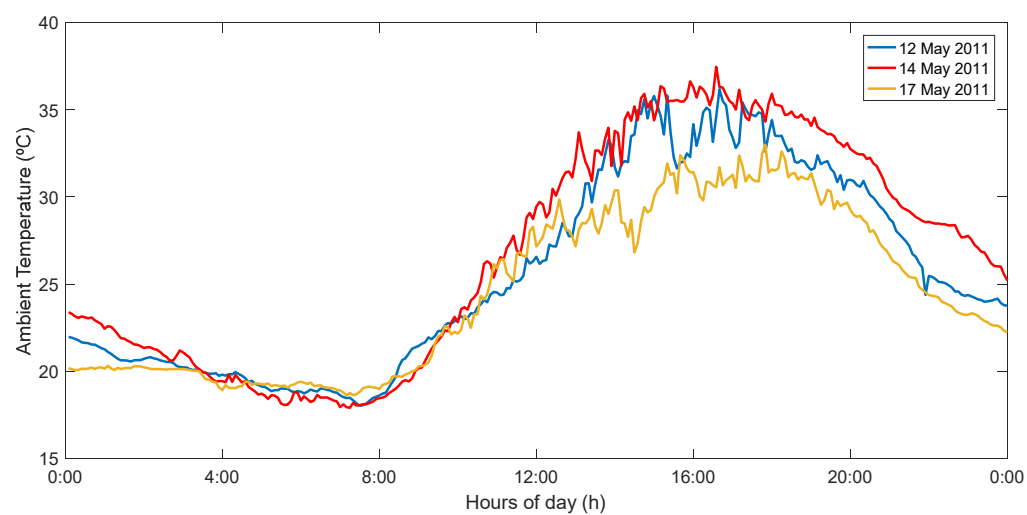


Figure 15. Daily profiles of ambient temperature registered during 12, 14 and 17 May 2011.

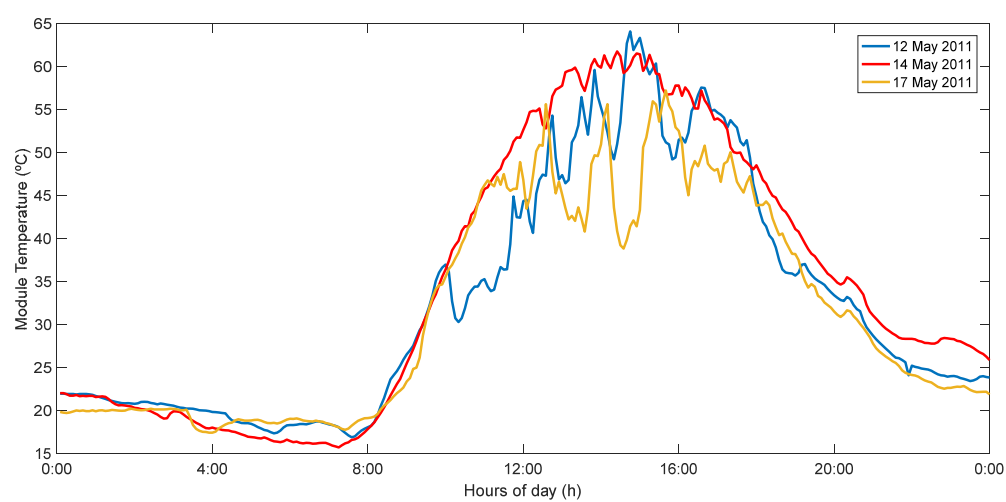


Figure 16. Daily profiles of module temperature registered during 12, 14 and 17 May 2011.

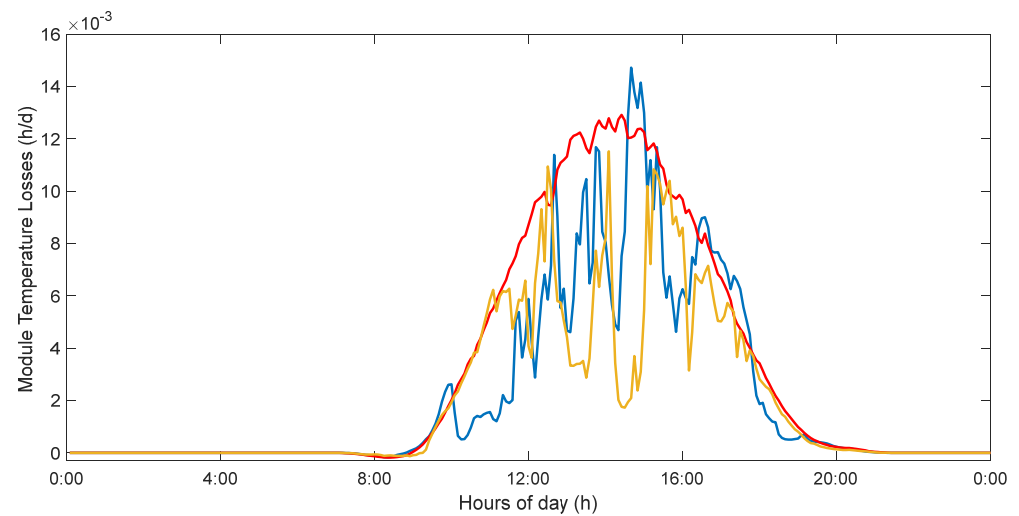


Figure 17. Daily profiles of temperature losses in PV module obtained during 12, 14 and 17 May 2011.

As the temperature decreases at these time instants, due to the presence of cloud shading, the response in the I-V curve of the module results in an increase in the voltage of the panels, with respect to their values without passing clouds, as can be seen in Figure 18. As a result of the increase in voltage and current, the power generated in the panels at these times with higher irradiance values can be up to 20% higher than in the case where the irradiance values correspond to clear days (Figure 19), and this leads to an occasional improvement in the working efficiency of the PV panels during these times, when clouds are passing and the CE effect is present (Figure 20). Zehner et al. had previously analysed the operation of a PV panel during CE events, showing that the power could be 30% higher than the nominal power [59].

With respect to the inverter performance for each of the types of days obtained in the classifications made in this work, Figure 21 shows that the type of day also influences the efficiency with which the inverter works.

The average daily efficiency of the inverter was better on days with higher daily irradiance values (H days) and lower on M days, with the lowest values obtained on L days. In this case, as opposed to what happened with the modules, the inverter worked with a higher average efficiency on days with lower fluctuations in irradiance values.

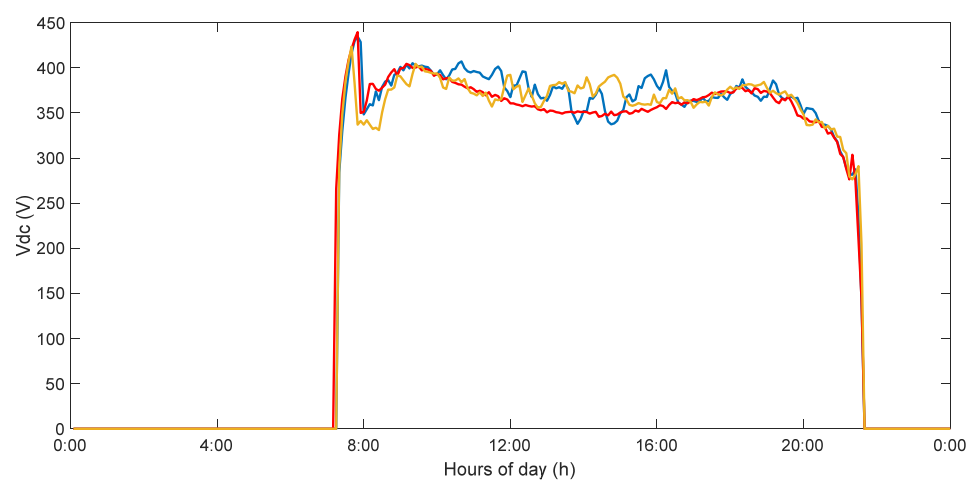


Figure 18. Daily profiles of the working voltage of the PV modules during 12, 14 and 17 May 2011.

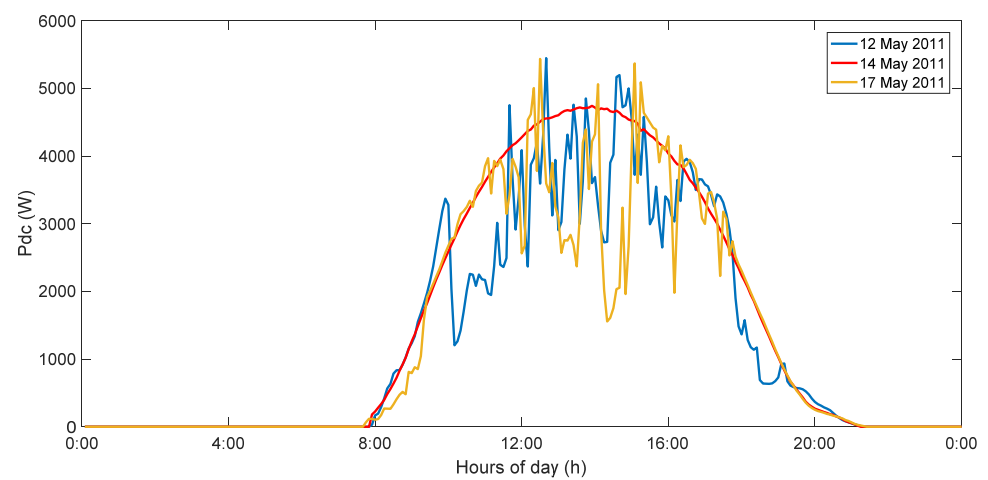


Figure 19. Daily profiles of the dc power generated by the PV panels during 12, 14 and 17 May 2011.

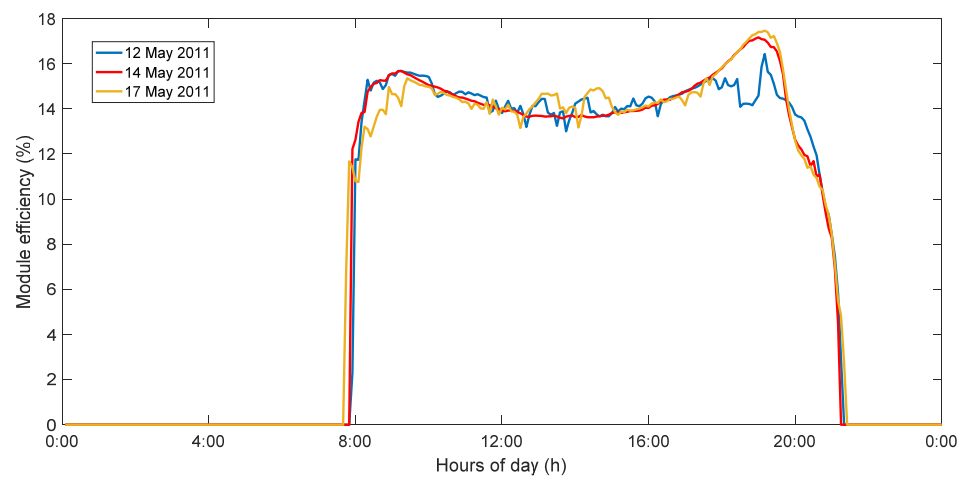


Figure 20. Daily profiles of module efficiency obtained during 12, 14 and 17 May 2011.

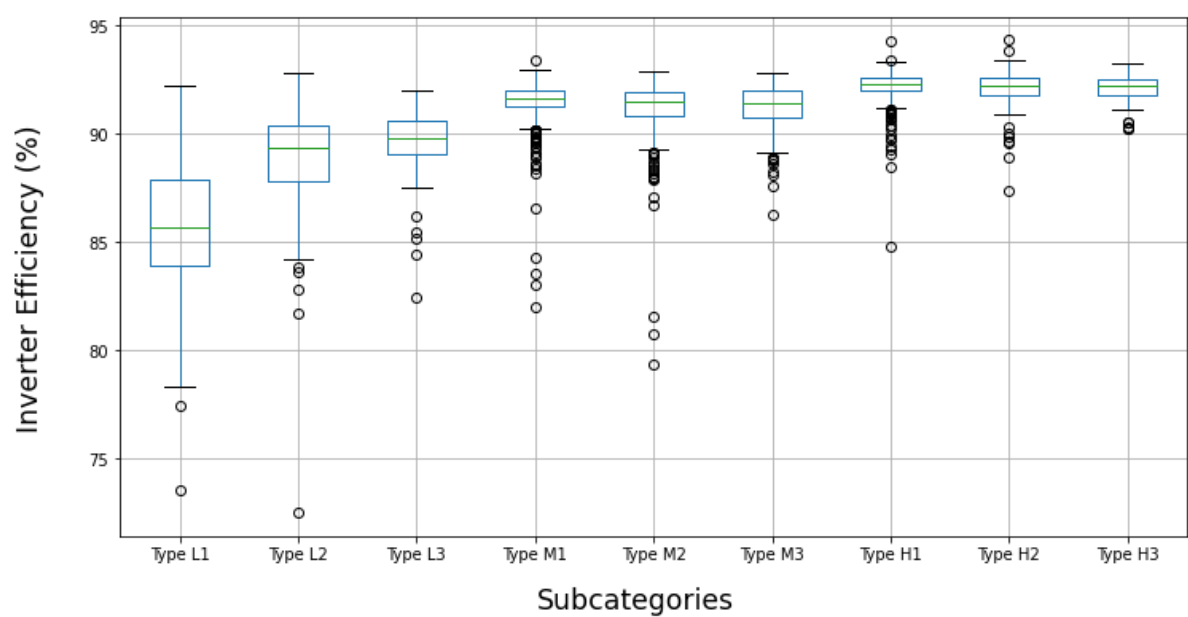


Figure 21. Inverter efficiency in the clusters L1, L2, L3, M1, M2, M3, H1, H2 and H3.

Figure 22 shows the working status of the inverter on the days shown as an example, which is coded by the numbers 0, 3, 7 and 10.

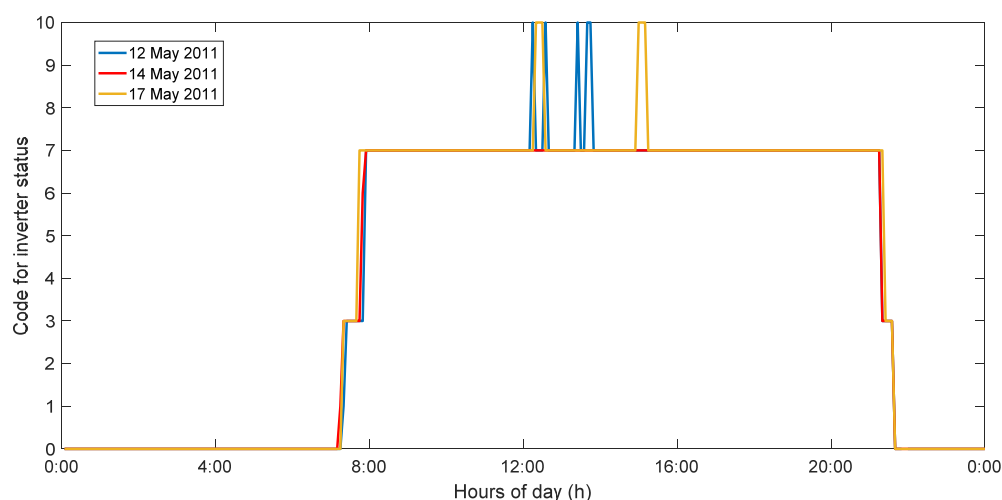


Figure 22. Daily profiles of inverter state obtained during 12, 14 and 17 May 2011.

The inverter state goes from having a value of 0, when it is not operating because there is no energy production in the PV modules, to a value or state of 3, in which the modules are starting to produce energy early in the morning, and in this state the inverter is synchronising with the grid to start injecting energy into it. State 7 corresponds to the operation of the inverter in maximum power point tracking (MPPT) mode. In it, the inverter tries to maximise the energy yield by operating the PV generator in its MPP. It can be seen in Figure 22 that on a clear sky day, with no cloud cover, the inverter operates all day in this MPPT state, until the production is so low at the end of the day that it switches back to state 3 and finally to the non-operational state 0. However, on days with passing clouds and the presence of CE, there are intervals of time when the inverter operates above the MPPT mode, a state represented by the code 10. If this mode of operation occurs frequently, as indicated by Jarvela et al. [41,56], it can lead not only to a reduction in inverter efficiency, but also to an increase in inverter operating temperature and long-term deterioration. In the design carried out in some PV installations, the MPPT range is narrower than the DC voltage range of the inverter. If the MPP power is higher than the allowed power, the inverter could operate in power-limiting mode. The PV generator power can be adjusted by controlling its operating voltage. Typically, power limiting is carried out by increasing the operating voltage, which reduces the current, and consequently, the power [56]. This is exactly what can be seen in Figure 18. Therefore, it is important to dimension the elements of the PV installations, taking into account the presence of these irradiance peaks. However, a balance has to be found, because increasing inverter size reduces saturation losses during high irradiance conditions, but decreases inverter conversion efficiency under low irradiance conditions, so, authors such as Luoma et al. [57] indicate in their work the range of values that optimise the performance of these elements of the PV installation taking into account the presence of CE events.

Figures 23 and 24 show, respectively, the losses that occur in the inverter on the three example days analysed, and the efficiency of the inverter throughout those days. It can be seen how, at times when there are irradiance peaks, the inverter may increase its losses in the process of converting the dc energy generated in the PV modules to ac energy. The presence of passing clouds leads to fluctuations in the irradiance and temperature values, which means that the inverter cannot work with a constant efficiency value throughout the day, as happens on 14 May, which is a clear sky day, but instead, this efficiency fluctuates to a greater extent. For the days 12 and 17 May, these changing operating conditions resulted in an average daily inverter efficiency value slightly lower than that of the clear sky day.

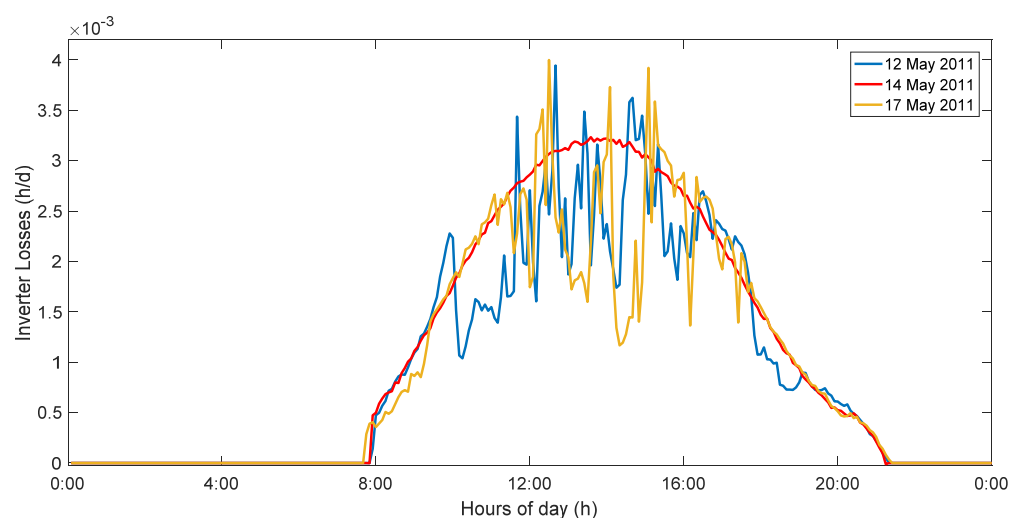


Figure 23. Daily profiles of inverter losses obtained during 12, 14 and 17 May 2011.

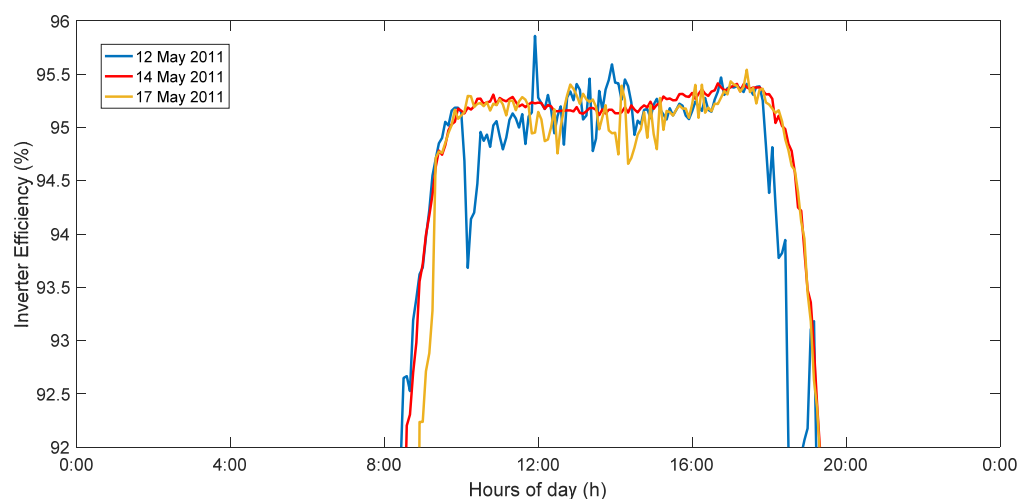


Figure 24. Daily profiles of inverter efficiency obtained during 12, 14 and 17 May 2011.

4. Conclusions

In this work, the daily irradiance profiles recorded in southern Spain for a nine-year monitoring period were categorised by applying the *k-means* clustering method, and were grouped into nine different day types. According to the total daily irradiance values, three categories were obtained, named low, medium and high, and the fluctuation of this parameter throughout the day, due to the presence of clouds, gave rise to three categories within each of the above. It is considered that the methodology was efficient and easy to apply, and it allowed the irradiance profiles recorded for the monitoring period analysed to be correctly classified. By determining the operation of the plant elements during this period, it was possible to observe that the behaviour of the modules and the inverter depended on the type of day, such that the weather forecasts would allow us to know the range of efficiency with which the plant elements are going to work on the following day. For large PV plants, with high nominal power values, a variation of several units in the efficiency of the elements could represent a significant amount of energy. The opposite behaviour of modules and inverter against days with fluctuations is what makes the *PR* values so similar in some of the clusters, regardless of the increase in fluctuations. It was found that inverters worked with a better average efficiency on days with higher irradiance values and lower fluctuations, i.e., days with no or little cloud cover, whereas PV modules improved their average efficiency on days with fluctuations and cloud passages.

This may be due to the presence, on those days, of higher irradiance values than on clear sky days, due to the so-called CE effect, and which, together with lower ambient and module temperatures due to the shading effect of the clouds, mean that at those instants of time there are not only peaks in production, but also that the module works with lower temperature losses and therefore with better yields. It was observed that it is important to take into account the presence of these events when sizing the inverter, in order to improve its operation and lifetime.

The tendency of the results obtained in this work could be applied to PV installations with similar configurations to the one analysed in this work, and with analogous meteorological conditions, such as Mediterranean countries. In any case, the methodology proposed in this work can also be easily applied to data monitored in PV installations located in geographical locations with very different meteorological conditions, in order to analyse the behaviour of their elements on different types of days. In future works, classification could be carried out using other algorithms, and performing classifications with different numbers of clusters. It would also be interesting to modify the size of the data set used to perform the classification in order to optimise the minimum amount necessary to obtain adequate results, as well as to apply the study to the operation of other PV installations with different configurations, in order to be able to generalise the results found. Another important aspect to analyse, which was also beyond the scope of this study, would be to determine the probabilities of obtaining a type of day depending on the type that is present on the preceding day or on the previous days.

With respect to the behaviour of the components of the installation in the presence of the CE events, the analysis was carried out in a small PV installation, but it would be very interesting to extend this study to a larger utility-scale plant, which requires a larger surface area for the PV panels, in order to study the limitation in size of the affected land areas by the CE effect, and to study how the cloud speed affects it. In larger installations, these peaks in production could affect their management in terms of their impact on the electricity grid. It would also be interesting to use data recorded with a higher sampling frequency, in order to be able to assess the duration of the CE events.

Author Contributions: I.S. proposed the topic of study for the paper, carried out the literature review, and proposed the methodology. She participated in the data processing, in the application of the algorithms, in the calculation of the photovoltaic installation components' yields and, in the writing and editing the manuscript. She carried out the supervision of the work for the implementation of the paper. J.L.E.-M. carried out the application of the classification algorithms for the types of days, and also participated in the writing of the paper. D.T.-M. participated in the preparation of the data, from the inverter monitoring files, and converted them into matrices in order to facilitate their use for the application of the classification algorithms. He also participated in the writing of the paper. R.J.R.-C. participated in carrying out the literature review and in writing and editing the paper. He also participated in the funding acquisition. V.P.-L. participated in the data processing and in the writing and editing of the paper. He also participated in the funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Project Monitoring And Integration of energy data with Seamless Temporal Accuracy for photovoltaic plants (MISTA), PID2019-108953RA-C22, funded by the Spanish Ministry of Economy and Competitiveness.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank the company Solar del Valle SL, which, via their collaboration agreements with our Research Group, provided data monitored in photovoltaic installations; and to the Down Syndrome Association in Córdoba, Spain, who are the owners of the PV installation under study, and with which the University of Cordoba granted the UCO Social Innova project called SENSEable Nursing homeS (SENS).

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

References

1. Departement of Economic and Social Services. Sustainable Development. United Nations: The Sustainable Goals. Available online: <https://sdgs.un.org/es/goals> (accessed on 15 March 2021).
2. IRENA (International Renewable Energy Agency). Double the Share of Renewables in the ‘Decade of Action’ to Achieve Energy Transition Objectives. Available online: <https://www.irena.org/newsroom/pressreleases/2020/Jan/Double-the-Share-of-Renewables-in-the-Decade-of-Action-to-Achieve-Energy-Transition-Objectives> (accessed on 15 March 2021).
3. European Commission. A European Green Deal. Available online: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en (accessed on 15 March 2021).
4. Unión Española Fotovoltaica (UNEF). *The Photovoltaic Sector towards a New Era, Annual Report UNEF 2020*; UNEF: Madrid, Spain, 2020; Available online: <https://unef.es/informacion-sectorial/informe-anual-unef/> (accessed on 15 March 2021).
5. International Renewable Energy Agency (IRENA). *The Post-COVID Recovery. An Agenda for Resilience, Development and Equality*; IRENA: Abu Dhabi, United Arab Emirates, 2020; Available online: <https://www.irena.org/publications/2020/Jun/Post-COVID-Recovery> (accessed on 17 March 2021).
6. Santiago, I.; Moreno-Munoz, A.; Quintero-Jiménez, P.; Garcia-Torres, F.; Gonzalez-Redondo, M.J. Electricity demand during pandemic times: The case of the COVID-19 in Spain. *Energy Policy* **2021**, *148*, 111954–111964. [CrossRef]
7. Solar Power Europe. Available online: <https://www.solarpowereurope.org/> (accessed on 18 March 2021).
8. International Energy Agency. Photovoltaic Power System Programme. Available online: <https://iea-pvps.org/> (accessed on 30 March 2021).
9. Rauegi, M.; Leccisi, E. A comprehensive assessment of the energy performance of the full range of electricity generation technologies deployed in the United Kingdom. *Energy Policy* **2016**, *90*, 46–59. [CrossRef]
10. Schaefer, J.L.; Siluk, J.C.M. An Algorithm-based Approach to Map the Players’ Network for Photovoltaic Energy Businesses. *Int. J. Sustain. Energy Plan. Manag.* **2021**, *30*, 43–60. [CrossRef]
11. Devetaković, M.; Djordjević, D.; Radojević, M.; Krstić-Furundžić, A.; Burduhos, B.-G.; Martinopoulos, G.; Neagoe, M.; Lobaccaro, G. Photovoltaics on Landmark Buildings with Distinctive Geometries. *Appl. Sci.* **2020**, *10*, 6696. [CrossRef]
12. Martinopoulos, G. Are rooftop photovoltaic systems a sustainable solution for Europe? A life cycle impact assessment and cost analysis. *Appl. Energy* **2020**, *257*, 114035. [CrossRef]
13. Strzalka, A.; Alam, N.; Duminil, E.; Coors, V.; Eicker, U. Large scale integration of photovoltaics in cities. *Appl. Energy* **2012**, *93*, 413–421. [CrossRef]
14. Santiago, I.; Trillo-Montero, D.; Luna-Rodríguez, J.J.; Real-Calvo, R.J.; Moreno-Muñoz, A. Analysis of variations in PV production, focussing on storage and dispatchability decisions. In Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Seville, Spain, 17–19 March 2015; pp. 2884–2889.
15. 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]
16. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. [CrossRef]
17. Inman, R.H.; Pedro, H.T.C.; Coimbra, C.F.M. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* **2013**, *39*, 535–576. [CrossRef]
18. Wan, C.; Zhao, J.; Song, Y.; Xu, Z.; Lin, J.; Hu, Z. Photovoltaic and solar power forecasting for smart grid energy management. *CSEE J. Power Energy Syst.* **2016**, *1*, 38–47. [CrossRef]
19. Sobri, S.; Koohi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [CrossRef]
20. Ren, Y.; Suganthan, P.N.; Srikanth, N. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renew. Sustain. Energy Rev.* **2015**, *50*, 82–91. [CrossRef]
21. Zhang, Y.; Beaudin, M.; Taheri, R.; Zareipour, H.; Wood, D. Day-Ahead Power Output Forecasting for Small-Scale Solar Photovoltaic Electricity Generators. *IEEE Trans. Smart Grid* **2015**, *6*, 2253–2262. [CrossRef]
22. Reikard, G.; Haupt, S.E.; Jensen, T. Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models. *Renew. Energy* **2017**, *112*, 474–485. [CrossRef]
23. Du, J.; Min, Q.; Zhang, P.; Guo, J.; Yang, J.; Yin, B. Short-term solar irradiance forecasts using sky images and radiative transfer model. *Energies* **2018**, *11*, 1107. [CrossRef]
24. Hartmann, B. Comparing various solar irradiance categorization methods—A critique on robustness. *Renew. Energy* **2020**, *154*, 661–671. [CrossRef]
25. Munshi, A.A.; Mohamed, Y.A.R.I. Photovoltaic power pattern clustering based on conventional and swarm clustering methods. *Sol. Energy* **2016**, *124*, 39–56. [CrossRef]
26. Torregrossa, D.; Le Boudec, J.Y.; Paolone, M. Model-free computation of ultra-short-term prediction intervals of solar irradiance. *Sol. Energy* **2016**, *124*, 57–67. [CrossRef]
27. Moreno-Tejera, S.; Silva-Pérez, M.A.; Ramírez-Santigosa, L.; Lillo-Bravo, I. Classification of days according to DNI profiles using clustering techniques. *Sol. Energy* **2017**, *146*, 319–333. [CrossRef]
28. Harrouni, S.; Maafi, A. Classification des Eclairissements Solaires à l’Aide de l’Analyse Fractale. *Revue Internationale des Énergies Renouvelables* **2002**, *5*, 107–122.

29. Harrouni, S. Fractal classification of typical meteorological days from global solar irradiance: Application to five sites of different climates. In *Modeling Solar Radiation at the Earth's Surface: Recent Advances*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 29–54.
30. Maafi, A.; Harrouni, S. Preliminary results of the fractal classification of daily solar irradiances. *Sol. Energy* **2003**. [\[CrossRef\]](#)
31. Chee Keong, C.; Yi Hong, L. Prediction of Solar Radiation Using Data Clustering and Time-Delay Neural Network. *J. Comput. Commun.* **2018**, *6*, 61–97. [\[CrossRef\]](#)
32. Govender, P.; Brooks, M.J.; Matthews, A.P. Cluster analysis for classification and forecasting of solar irradiance in Durban, South Africa. *J. Energy S. Afr.* **2018**, *29*, 51–62. [\[CrossRef\]](#)
33. Govender, P.; Sivakumar, V. Investigating diffuse irradiance variation under different cloud conditions in Durban, using k-means clustering. *J. Energy S. Afr.* **2019**, *30*, 22–32. [\[CrossRef\]](#)
34. Chicco, G.; Cocina, V.; Spertino, F. Characterization of solar irradiance profiles for photovoltaic system studies through data rescaling in time and amplitude. In Proceedings of the Universities Power Engineering Conference, Perth, Australia, 28 September–1 October 2014.
35. Badescu, V. *Modeling Solar Radiation at the Earth's Surface: Recent Advances*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 29–54.
36. Gastón-Romeo, M.; Leon, T.; Mallor, F.; Ramírez-Santigosa, L. A Morphological Clustering Method for daily solar radiation curves. *Sol. Energy* **2011**, *85*, 1824–1836. [\[CrossRef\]](#)
37. Jeanty, P.; Delsaut, M.; Trovalet, L.; Ralambondrainy, H.; Lan-Sun-luk, J.D.; Bessafi, M.; Charton, P.; Chabriot, J.P. Clustering daily solar radiation from Reunion island using data analysis methods. *Renew. Energy Power Qual. J.* **2013**, *1*, 461–466. [\[CrossRef\]](#)
38. Santiago, I.; Trillo Montero, D.; Luna Rodriguez, J.J.; Moreno Garcia, I.M.; Palacios Garcia, E.J. Graphical diagnosis of performances in photovoltaic systems: A case study in southern Spain. *Energies* **2017**, *10*, 1964. [\[CrossRef\]](#)
39. Wittkopf, S.; Valliappan, S.; Liu, L.; Ang, K.S.; Cheng, S.C.J. Analytical performance monitoring of a 142.5kW p grid-connected rooftop BIPV system in Singapore. *Renew. Energy* **2012**, *47*, 9–20. [\[CrossRef\]](#)
40. Lappalainen, K.; Kleissl, J. Analysis of the cloud enhancement phenomenon and its effects on photovoltaic generators based on cloud speed sensor measurements. *J. Renew. Sustain. Energy* **2020**, *12*, 043502–043515. [\[CrossRef\]](#)
41. Järvelä, M.; Lappalainen, K.; Valkealahti, S. Characteristics of the cloud enhancement phenomenon and PV power plants. *Sol. Energy* **2020**, *196*, 137–145. [\[CrossRef\]](#)
42. Bouveyron, C.; Brunet-Saumard, C. Model-based clustering of high-dimensional data: A review. *Comput. Stat. Data Anal.* **2014**, *71*, 52–78. [\[CrossRef\]](#)
43. Chicco, G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* **2012**, *1*, 68–80. [\[CrossRef\]](#)
44. Müller, M. *Information Retrieval for Music and Motion*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 69–84.
45. Izakian, H.; Pedrycz, W.; Jamal, I. Fuzzy clustering of time series data using dynamic time warping distance. *Eng. Appl. Artif. Intell.* **2015**, *39*, 235–244. [\[CrossRef\]](#)
46. Petitjean, F.; Ketterlin, A.; Gançarski, P. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognit.* **2011**, *44*, 678–693. [\[CrossRef\]](#)
47. Cuturi, M.; Blondel, M. Soft-DTW: A differentiable loss function for time-series. In Proceedings of the 34th International Conference on Machine Learning, ICML, Sydney, Australia, 6–11 August 2017.
48. Tavenard, R.; Faouzi, J.; Vandewiele, G.; Divo, F.; Androz, G.; Holtz, C.; Payne, M.; Yurchak, R.; Rußwurm, M.; Kolar, K.; et al. Tslearn, a Machine Learning Toolkit for Time Series Data. *J. Mach. Learn. Res.* **2020**, *21*, 1–6.
49. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
50. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API design for machine learning software: Experiences from the scikit-learn project. In Proceedings of the ECML PKDD Workshop: Languages for Data Mining and Machine Learning, Prague, Czech Republic, 23 September 2013; pp. 108–122.
51. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [\[CrossRef\]](#)
53. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 56–61.
54. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [\[CrossRef\]](#)
55. Trillo-Montero, D.; Santiago, I.; Luna-Rodriguez, J.J.; Real-Calvo, R. Development of a software application to evaluate the performance and energy losses of grid-connected photovoltaic systems. *Energy Convers. Manag.* **2014**, *81*, 144–159. [\[CrossRef\]](#)
56. Järvelä, M.; Valkealahti, S. Operation of a PV power plant during overpower events caused by the cloud enhancement phenomenon. *Energies* **2020**, *13*, 2185. [\[CrossRef\]](#)
57. Luoma, J.; Kleissl, J.; Murray, K. Optimal inverter sizing considering cloud enhancement. *Sol. Energy* **2012**, *86*, 421–429. [\[CrossRef\]](#)
58. Tapakis, R.; Charalambides, A.G. Enhanced values of global irradiance due to the presence of clouds in Eastern Mediterranean. *Renew. Energy* **2014**, *62*, 459–467. [\[CrossRef\]](#)

-
59. Zehner, M.; Weigl, T.; Hartmann, M.; Thaler, S.; Schrank, O.; Czakalla, M.; Mayer, B.; Betts, T.R.; Gottschalg, R.; Behrens, K.; et al. Energy Loss Due to Irradiance Enhancement. In Proceedings of the 26th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, Germany, 5–8 September 2011.
 60. Yordanov, G.H. A study of extreme overirradiance events for solar energy applications using NASA's I3RC Monte Carlo radiative transfer model. *Sol. Energy* **2015**, *122*, 954–965. [[CrossRef](#)]
 61. Espinosa-Gavira, M.J.; Agüera-Pérez, A.; de la Rosa, J.J.; Palomares-Salas, J.C.; Sierra-Fernández, J.M. An On-Line Low-Cost Irradiance Monitoring Network with Sub-Second Sampling Adapted to Small-Scale PV Systems. *Sensors* **2018**, *18*, 3405. [[CrossRef](#)]