

Article

Design and Development of a Relational Database Management System (RDBMS) with Open Source Tools for the Processing of Data Monitored in a Set of Photovoltaic (PV) Plants

David Trillo-Montero, Samuel Cosano-Lucena, Miguel Gonzalez-Redondo, Juan Jesus Luna-Rodriguez and Isabel Santiago *

Department of Electronic and Computer Engineering, Campus de Rabanales, Universidad de Córdoba, 14071 Córdoba, Spain

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

Abstract: The objective of this work has been to implement an orderly, accessible, fast and space-saving storage system that allows the transfer to a Relational Database Management System (RDBMS) of all the data corresponding to the monitoring of a set of photovoltaic (PV) systems whose behaviour is to be analysed. The RDBMS consists of a series of linked databases, enabling all PV system information to be stored, and it is scalable so it can be expanded depending on the number of installations to be studied. The data recorded in the plants are found in a large number of very disaggregated files, and with different measured parameters, different formats, nomenclatures, or units of measurement, so the developed system is responsible for homogenising all the information for storage. For this purpose, a procedure has been developed to carry out the automatic transfer of all the data recorded in their corresponding databases. In addition, in this work, a web application called *S-lar 2* has been developed to facilitate selective access to all the data once stored in the corresponding tables. This application, which is connected to the designed databases, allows the storage and management of the information coming from the PV plants, in order to determine, among other things, the operation mode of each of the components of these facilities. Using the data already organised, it has also been possible to establish a system for comparing the production of inverters within the same plant in order to have a tool that allows the quick and visual detection of possible deviations between them and thus detect malfunctions in any of the components. The whole procedure has been carried out using free software, such as Maria DB and Python.

Keywords: Photovoltaic systems; software application; monitoring systems; databases; Maria DB; Python

Citation: Trillo-Montero, D.; Cosano-Lucena, S.; Gonzalez-Redondo, M.; Luna-Rodriguez, J.J.; Santiago, I. Design and Development of a Relational Database System (RDBMS) with Open Source Tools for the Processing of Data Monitored in a Set of Photovoltaic (PV) Plants. *Appl. Sci.* **2023**, *13*, 1357. <https://doi.org/10.3390/app13031357>

Academic Editor: Luis Hernández-Callejo

Received: 1 January 2023

Revised: 17 January 2023

Accepted: 17 January 2023

Published: 19 January 2023



Copyright: © 2023 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

In recent decades there has been a renaissance in renewable energies due to the proximity of the exhaustion of fossil energy sources, the progressive increase in their cost, the environmental problems derived from their exploitation, transport and consumption, and, of course, due to the great external dependence of many countries on this type of energy sources, which also implies that energy supply and the economy are directly linked to the geopolitical instabilities that may exist between different countries.

Renewable energies, through their different technologies, are playing a fundamental role in achieving sustainable development from the energy point of view and in meeting objectives against climate change and environmental degradation, objectives expressed by different international organisations and institutions in recent years, such as the United

Nations and its Sustainable Development Goals [1], specifically Goal 7, which aims to ensure access to affordable, reliable, sustainable and modern energy for all. The International Renewable Energy Agency (IRENA) has set a target of 57% of the world's energy to be produced from renewable energy sources by 2030, compared to 26% at present [2]. This increase in the share of renewable generation, which the European Union, through the European Green Deal and Green Recovery, is taking to an even greater extreme with the proposal for zero greenhouse gas emissions by 2050, means that this decade will be seen as the one in which a major global energy transformation will take place [3]. A pathway to net zero emissions (PNZ) is therefore proposed, which requires the electrification of systems that are currently based on fossil fuels, such as air conditioning, the transport sector, and the mechanisation and activities of the secondary and primary sectors.

Thus, the challenge at the energy level is not only to generate the electricity currently required in a cleaner way, eliminating polluting energy sources from the generation mix, but also to further increase electricity generation capacity in order to accommodate within the electricity system sectors that require a high demand for energy and which are currently based on other polluting energy sources. Renewable energies will therefore have to provide the energy supply required for all current activity.

These renewable energies experienced an increase of 9.1% in installed capacity in 2021, which grew by 257 GW according to IRENA, a growth trend that was also experienced in 2020, which has shown its resilience in an adverse context such as the pandemic experienced due to Covid-19 [4,5]. Solar energy is, within the framework of renewable energies, one of those that can stand out for having greater technological development and greater possibilities of competing in the market. Specifically, in 2021, solar PV energy was the leading renewable energy technology and the most installed worldwide. New PV capacity in 2021 was 175 GW, an increase of 21% compared to 2020, exceeding 100 GW for the fourth consecutive year and reaching a cumulative 942 GW [4–6]. This technology is driving the largest increase in renewable energy, overtaking wind in absolute terms [2,4]. It is an inexhaustible source of energy, it is distributed all over the world and it is one of the main technologies on which users are focused in order to meet the objectives of reducing polluting emissions [7,8]. In recent years, the technological development of this type of installation, together with the reduction in prices that has taken place (although the current crisis has reversed this trend in recent months), has meant that its use has become increasingly widespread [4,7,9–13]. The growth of this technology is explained by the significant increase in the market on all continents, led by countries such as China, which held 31% of the global market, the United States, which installed 26.9 GW in 2021, India, with an estimated annual installed capacity of 13 GW in 2021, Japan with 6.5 GW, Brazil with 5.5 GW, Germany with 5.3 GW, Spain with 4.9 GW, Australia with 4.6 GW, South Korea with 4.2 GW and France with 3.4 GW [5].

The opportunities for improvement in this sector are based on development at the level of installed capacity. However, there are also a series of challenges that must be addressed, among which the achievement of technological improvements must be highlighted, which undoubtedly seek to increase generation capacity, cost reduction and better integration into the grid. Improving the competitiveness of PV installations compared to other electricity generation technologies must be a priority for this sector, and to this end, it is essential to improve the efficiency and production of its elements and to have the least possible impact on the electricity grid.

In terms of increasing efficiency, research is focused on improving both the production processes and the long-term stability of cells and modules.

On the one hand, the aim is to increase the theoretical performance of the PV plant's generation components by improving both the semiconductor materials currently used to manufacture PV modules, such as silicon or thin-film panels, and by making use of emerging panel technologies based on organic cells [13–15]. It is also possible to improve the configuration and internal structure of the panels, as well as to make progress in the en-

capsulating or coating materials of the modules [16–19]. Work is also being done to optimise the mode of operation of the inverters in order to extract as much power as possible from the generator panels [20–22]. It is also important, especially considering the increasing growth of this sector, to optimise the lifetime of the panels by analysing the causes of their degradation [23–25], and, once this lifetime is over, to ensure that their impact is reduced as much as possible [26–29].

Regarding the operation of PV plants, once a certain technology has been installed, two important aspects to consider are the optimisation of their operation and the optimal integration of their production into the grid, if they are connected to it.

PV systems may have significant fluctuations in operation due to cloud cover, unplanned outages, dust, soiling, or temporary events such as failures and damages, or operation and maintenance (O&M) issues [30]. Therefore, the implementation of means to assist in the diagnosis, monitoring and detection of these variations in production will allow easier performance of maintenance of the plant. The characterisation of the real margins in which a given PV installation operates allows preventive and predictive maintenance to be carried out, enabling a more active action instead of reactive fixes [31]. This will result in a higher number of operating hours and, therefore, an increase in profitability. In this sense, an efficient O&M service is essential to achieve the best technical and economic performance of any PV system during its entire life cycle [27,32–34]. To improve this service, which can be extremely complex and time-consuming [27], further innovations and developments are needed to provide a range of technological solutions or advanced services such as tools for automated fault or deviation detection [35–38], including even the use of digital twins [27,39], intelligent monitoring of modules by drones [40,41], proposed anti-dirt solutions [41,42] or systems to provide cooling [43–46] to optimise production [47,48].

Numerous authors have conducted various studies related to the analysis of the performance of components of grid-connected PV plants in different geographical locations, with different climatic conditions, studying the different factors that can influence it [49]. To name a few, Roumpakias et al. analysed the performance of a plant in Greece for six years, studying the effects of degradation [50], while Murat-Ates et al. analysed the performance of a rooftop plant in Turkey [51]. Gopi et al. carried out a performance analysis at a utility-scale located in a rainy tropical monsoon climate [52], Bansal et al. in a hot and dry climate [53], Ihaddadene et al. in a dry arid climate in Algeria [54] and Daher et al. in a maritime desert climate [55]. Kumar et al. analysed the performance of PV installations located in water channels in India [56]. Srivastava et al. analysed the performance of plants located in various parts of the world [57] and found that their performance has improved over the years due to improved technologies. Adar et al. also compared the performance of panels of different technologies, such as monocrystalline, polycrystalline and amorphous Si, analysing their dependence on weather conditions [58]. The performance of technologies was also compared in the work of Rasheedi et al., in this case at a plant in Kuwait [59]. This type of performance analysis is usually based on the IEC 61724 PV system performance—Part 1: Monitoring standard, which regulates the parameters that must be measured and calculated to carry out this performance analysis depending on the type of PV installation [60].

With regard to the integration of this type of renewable generation system into the electricity grid, it is also important to have production forecasting systems that make it possible to anticipate the amount of energy that will be produced, and the fluctuations and ramps in this production, which will help in the sizing of storage systems to support the management and participation in the electricity market of the production of this type of energy sources [30,48,61–63]. Artificial intelligence techniques such as machine learning or deep learning are making it possible to address this type of question, and numerous tools are already available [64–67].

On the other hand, the aim is to minimise the impact of this production on the deviation of the parameters of the electricity network signal, so that these are within the supply

quality intervals imposed by the regulations (standard EN-50160 Voltage characteristics of electricity supplied by public electricity networks [68]) for the electricity signal that reaches consumers and, thus, prevent the operation of these plants from exceeding the limits that could cause damage to the elements of the network [69–72].

For all these issues, it is essential to first have monitoring systems in these plants that make it possible to know their operating parameters, as well as to obtain information on what is happening at their point of common coupling (PCC) and to have access to all the parameters that are recorded in them. Numerous studies have addressed the development of monitoring systems that make this information available [31,73–77]. However, the cost and complexity of existing PV monitoring systems restrict their use to large-scale PV plants [78], and, most of the time, private PV systems are not being checked or supervised [75]. A PV monitoring system should be low-cost to be broadly accepted [75]. Although the increasing number of smaller-scale plants currently in operation do not usually have such monitoring systems in place, many of them do have the data recorded by the plants' own inverters, which can be used to analyse the plant's performance without the need for additional measuring equipment with an extra cost [69,79,80].

In general, whether there is a monitoring system or not, the parameters and magnitudes that are measured must be processed and analysed in order to obtain useful information to extract knowledge to help carry out intelligent supervision of the different elements of the plants and to make decisions in the shortest possible time regarding their operation and integration in the electricity grid or information that helps to improve the design and installation of future plants.

Even in the absence of additional measurement sensors, PV plants generally produce a considerable amount of data [75], which creates a complex analysis and performance calculation process [27], so tools are needed to help manage this large volume of data, which, in addition to being acquired, must be processed, stored and analysed easily and effectively [30,48,81]. Some authors propose solutions and tools based on Big Data for the analysis of data from PV plants [27,82,83], highlighting among other aspects the importance of the correct organisation and storage of the data [78].

A system of this type is inherently complex due to the wide variety of systems and elements that must be integrated [83]. Different PV installations usually have different monitoring elements, with equipment of various models, which can record different types of parameters related to the production of the modules, the operation of the inverter itself, or some parameters of the grid signal, with different notations and/or units of measurement; and all this recorded information can be found in files with different types of formats. Even when monitoring is carried out by the plant's own inverters, all this information does not necessarily have to be homogeneous but usually may vary from one inverter model to another, even within the same company.

Some leading companies in system technologies for PV applications, along with their inverter models and monitoring systems, provide Internet portals for the visualisation of the data recorded by their own equipment. These Internet portals provide access to display the data of numerous plants that owners have registered, and even have the option to display them in a shared way so that any user registered in the system can have access to them. However, on such platforms, the parameters that can be displayed and their graphical representations are often pre-configured or the configuration options are often very limited. These applications do allow the downloading of data in formats such as CSV. However, the files that can be downloaded are often highly disaggregated and have an organisation that may vary from one equipment model to another. Therefore, even with the use of these platforms, if you want to carry out a deeper and more personalised analysis of the data contained in these files, it is necessary to be able to manage them in an organised way in order to process all the information they contain.

In this context, the aim of this work is the development and implementation of an application to store, in an orderly and automated way, all the data files from the monitoring of the inverters of a set of PV plants, and to have access to them in a flexible and

versatile way, in order to be able to process the data and analyse them in a much more detailed and in-depth way. The need arises from the agreement between the authors' research group and the company called Solar del Valle S.L., whose activity consists of the installation and O&M of PV plants, in order to be able to make use of the data recorded in these plants and to be able to take advantage of all the information hidden in them.

The work has resulted in a tool, called *S-lar 2*, which has allowed us to analyse the data recorded by the inverters and additional sensors in a group of PV plants. An earlier version of this tool is available in some previous publications [79,80].

The presented solution has several advantages, such as its capacity for the simultaneous processing of data from multiple plants, integrating all the plant's measurement sensors, adapting to a variety of typologies and structures of measured data, calculating element performance parameters and aggregating and averaging data at different time intervals. The system is scalable to increase the number of plants to be analysed and allows the data to be homogenised. This facilitates the joint processing of their data for their possible aggregation and joint management, with a view to their integration into the grid, within the concept of the Virtual Power Plant (VPP) [31,84–86].

Although the current version of the tool is not prepared for real-time analysis, it is possible to store within it monitoring data, for example on a daily and/or weekly basis, and to conveniently carry out data processing, close to the concept of Edge Computing [84], in order to supervise the operation of the plants under analysis.

After this introduction, which has described the context and objectives of the work to be established, Section 2 describes the methodology, which shows how this tool has been developed, Section 3 shows the developed databases, the application that allows access to the information, and an example of the analysis of the data stored in the management system. Finally, Section 4 includes the conclusions of the work.

2. Methodology

The tool developed in this work has consisted of a storage system for the information monitored in a set of PV plants, together with a system that allows us to automatically load in it the monitored data and the information corresponding to the plants, and an application that enables to easily access the information. This system is characterised by the fact that it adapts to a variety of PV plant configurations, different equipment models, recording of different types of parameters and even using different nomenclature or units of measurement. In addition, all the information measured in the plants was generated and/or organised in different ways and with files in different formats, such that the tool developed for its analysis allows information to be stored from files in different formats, both CSV, which was already done in previous versions, and XML, which is one of the improvements compared to previous versions of the system [79,80]. Furthermore, it has been designed in such a way that it is scalable so that the study of new installations can be incorporated into the system or elements can be modified in those already being studied, such as the incorporation or replacement of inverters or sensors. The tool, therefore, allows the data collected in different plants to be unified, which facilitates the joint processing of their data for their possible aggregation and joint management, with a view to their integration into the grid and to analyse their behaviour at different time intervals, both in terms of recording time and determining accumulated data or hourly, daily, monthly, or annual averages.

2.1. Characterisation and Modelling of the System Whose Information Is to Be Managed

In this paper, the tool developed to carry out the management of the data monitored in the PV plants will consist first in a Relational Database Management System (RDBMS) that allows storage in an orderly, accessible, fast way and without space limitations of all the information corresponding to the real PV systems to be studied, including the specifications of each and every one of its components or equipment, its configuration, and all the data that are recorded by the different measurement elements.

The transcription of a real system to an RDBMS first needs to be modelled by defining a logical and conceptual model of all the elements that compose it, for which it is essential to have correctly characterised the complete real system. Subsequently, an RDBMS must be designed to allow the model of the real system to be transferred to it and to represent the real system as faithfully as possible. This will allow the efficient management of all the information.

The conceptual model in this case corresponds to the description of the configuration of the different elements that constitute the various PV facilities to be analysed, and of the existing relationships between these elements, some of which will be involved in the generation of the information to be organised and analysed in the data storage system that has been developed. This conceptual model must be independent of the RDBMS to be used. In this case, the RDBMS must include the real entities shown in Table 1.

Table 1. Real entities included in the Relational Data Base Management System (RDBMS) implemented in this paper for the analysis of PV installations.

Real Entities Included in RDBMS
PV installations and their configuration and features.
PV modules and their technical specifications.
Inverters and their technical specifications.
Meteorological sensors
Other measuring devices, such as those that measure the current generated in strings of PV modules.
Monitored meteorological data.
Data from inverters.
Instantaneous parameters calculated using the parameters from monitoring.
Aggregated and averaged parameters, calculated from monitored and calculated parameters.

It is also necessary to consider the inverters in each of the plants, their technical specifications, and how this equipment is associated with the set of modules from which they receive the generated energy. If there are measurement sensors in the PV plant that record additional parameters to those measured by the inverters, such as meteorological parameters that affect the production of the plant, or the current generated in each string of PV modules (I_{grid}), the configuration of these sensors must also be considered in the RDBMS.

In addition to the different pieces of equipment that are part of the PV installations, it is necessary to characterise the data that are measured by the inverters or additional sensors in each one of them, the notation they use and their units. It is also necessary to consider how the measured data are configured and organised after monitoring, as their structure will condition how the automated migration of their content to the information storage system can be carried out. If we focus on the data measured by the inverters of the plants, these depend on the specific inverter model.

An example of the parameters that can be measured by these devices is shown in Table 2, which corresponds to some inverter models from the SMA company, which are the ones analysed in this work. However, not all inverters measure the same parameters, nor do they use the same notation or measurement units, and there may even exist differences between the different models of the same manufacturer.

Table 2. Parameters generally measured by inverters.

Parameter	Notation
DC current from PV modules	I_{DC}
Inverter input DC voltage from PV modules	V_{DC}
Inverter input DC power from the generator	P_{DC}
Inverter output AC power	P_{AC}
Inverter output total energy	E_{AC}
AC current injected into the grid	I_{AC}
Grid current	I_{grid}
Grid AC voltage in phase 1	V_{AC_1}
Grid AC voltage in phase 2	V_{AC_2}
Grid AC voltage in phase 3	V_{AC_3}
Grid frequency	F_{ac}
Grid impedance	Z_{grid}
Inverter operating status	<i>Status</i>
Error code (where appropriate)	<i>Error</i>
Number of times that the inverter finds the MPP	n_{MPP}
Inverter working total hours	h_{on}
Grid injecting total hours	h_{on_grid}
Total connections to the grid	n_{on_grid}
Installation boot-up number	n_b
Boot-up time	h_b
Inverter operating temperature	T_{inv}
Isolation resistance	R_{isol}

With respect to the monitored meteorological data, these may include horizontal and/or in-plane solar irradiance, ambient and module temperatures and, in some cases, wind velocity and relative humidity. The values of these parameters may be usually recorded at the same frequency as the data logged by inverters and saved in the same or different files. In the case of the PV plants analysed in this work, the data from meteorological sensors are sent, as well as the data from inverters, to a communication centre, such that all the information is collected and stored together in daily files or data files every 5 min. However, in the case of PV installations with a larger number of inverters, data from all inverters may not be collected in the same file but may be divided into different files corresponding to data from different groups of inverters. In this case, if there is only a single meteorological data record in the PV plant, these weather data will be stored in the files corresponding to the data register by one group of inverters, and in the data files corresponding to the rest of the inverters in the plant these weather measurements will not appear. However, once stored in the databases, the meteorological data measured in each installation must be available in such a way that, for example, if calculating the yields of PV modules associated with all the inverters of the plant is desired, it is necessary to have access to both the power production data of each inverter and the irradiance data. It is important to take these issues into account in order to be able to organise the information correctly.

Returning to Table 1, the last item listed refers to the parameters to be determined by calculation from the monitored data, which allows characterisation of the behaviour of the PV plant elements. These are those recommended by the standard IEC 61724, Photovoltaic system performance—Part 1: Monitoring, which indicates a series of parameters that must be calculated in order to perform an adequate analysis of the performance of the plant elements according to the type of PV installation [60]. All of these parameters are listed in a previous publication [80]. To calculate these parameters, in addition to the

monitored data, it is necessary to know the specifications of the modules and their configuration in the plant, information which, as mentioned above, must be stored in the database system, and which would have to be accessed to proceed with the calculation.

In order to create a complete model of the real system that allows its transcription to the databases that comprise the RDBMS, it is necessary, in addition to knowing the entities that compose and configure the system, to establish the existing relationships between them, which, in the case of the PV plants analysed in this work, are those shown in Table 3.

Table 3. Established relationships between real system entities included in the RDBMS.

Real Entities Included in RDBMS
An installation plant has at least one inverter.
An inverter model can be installed in one or several PV installations. But all installations do not necessarily have to be equipped with the same inverter model. In addition, inverters of different models can be installed in the same PV system.
A meter receives energy from one or more inverters.
Each piece of data that is generated in the monitoring system (with date and time), from inverters and other sensors, refers to a parameter.
Each piece of equipment that records information does so with a sampling frequency, which does not necessarily have to be the same as that of the rest of the equipment.
Each PV module manufacturer provides one or more models with different specifications.
In each PV plant, all PV modules associated with an inverter or string current metering device are of the same model.
In each PV plant, there is usually only one record of meteorological data. If there is no record of this type of data in a plant, meteorological data from another plant under study could be used for the calculations if it is located close to it.
The recorded weather data are stored in the data files corresponding to those recorded by some of the inverters in the plant. But there may be inverter data files that do not have this information.

2.2. Database Management System Used

Once an RDBMS is chosen, all the entities and their relationships collected in Tables 1 to 3 need to be transformed into real and related tables. The fields of each of the tables are the characteristics of the different entities they represent. In these tables, it is necessary to establish which data will be their primary keys, that is the major and unique identifiers for each table, as well as their foreign keys, that is a column or group of columns in a dependent table that refer to a column or group of columns in another table, known as the master table. The relationships between them must also be established. This allows the real configuration of the elements to be studied to be properly reflected in the RDBMS.

In this work, MariaDB [87] has been chosen for the developed system as a database management system because it has General Public Licence (GPL) and therefore it is a free management system and open source. It comes directly from MySQL and is widely used by important companies such as Google, Mozilla and the Wikimedia Foundation, and is supported by a large community of users. It stands out for being flexible and scalable, high performance and high availability, robust and with very complete support for most development environments, such as .NET, Java, Python, or PHP. Another important point to highlight about MariaDB is the variety of tools available for its administration, which are also free, such as MySQL WorkBench [88] or HeidiSQL [89]. Although this database manager allows the storage of data recorded with a higher sampling frequency, in this work data collected with a period of 5 min have been analysed, which is the period at which the inverters are configured by the company that operates them.

The structure that has been adopted for the RDBMS developed in this work has, on the one hand, a common database, in which the technical specifications of the elements and the configuration of all the PV installations to be studied are stored, and, on the other, it has been decided to have as many individual databases as installations to be studied, in which the data recorded in them will be stored. In the case of plants with many inverters and with multiple data sources, it is possible to divide them into several databases, and these databases, although independent, must share information corresponding, for example, to the measurement of meteorological parameters, given that, as previously indicated, in each PV installation studied they are usually recorded in a unique way.

The need to divide the RDBMS into several databases is due to the large volume of information that is recorded from the different PV installations and that needs to be stored, with the objective of reducing memory and processing time compared to what would be necessary if all the information were available in a single database, taking advantage of the flexibility feature of MariaDB that allows establishing relationships between tables and queries from different databases.

2.3. Procedure for Loading the Monitoring Data into the Database

Once the structure of the RDBMS system was designed, the next step was its development and the subsequent loading of information into the databases that constitute it. First, a common database must be created. The table structure of this database was done manually, directly working within the MariaDB administrator, adapting it to the needs that were found in the different PV installations to be analysed. Conversely, the structure of the individual databases was created following the procedure shown below.

When a new PV plant is going to be analysed, the first step is to enter manually in the common database only the name of this new plant and its location, so that there is a record in the database of a new plant for which no information previously existed. The next step is to create an individual database for this new plant. The process was implemented in such a way that the creation of this individual database is automated, and its structure is generated according to the characteristics of the new plant. For this purpose, a script was generated to carry out the migration process of the files containing the data monitored in this plant.

In this process, information about the set of elements that monitor information in the plant, inverters, string current metering devices and possible environmental sensors, is initially extracted from these files and, based on them, code in SQL language is automatically generated to create the physical model or table structure of the individual database, suitable for the data monitored by these elements to be stored in them. In addition, this process automatically transfers to the common database the identification codes of all the inverters and sensors that record information, and the set of parameters that are measured. Then, the migration process is interrupted, so that all the information corresponding to the elements of this installation is completed manually in the common database, making use of the manufacturers' specifications and the configuration of the plant modules associated with each inverter or string metering device, provided by the company that operates and maintains the plant. Once all this information is completed, the migration process continues, and now all the monitored data is automatically saved in the corresponding tables of the individual database that has been created. As indicated, free software resources were used for these developments, programming a series of Python scripts and using SQL language to work with the databases.

The entire structure of tables, views and procedures of the individual databases was prepared by a Python function named *create_db_structure* during the execution of the data migration process, where SQL files are created with everything necessary for its creation. To facilitate the migration process, a pure-Python MySQL client library named *pymysql* was used with which an auxiliary class was created with a series of functions and properties that facilitate the interaction with the database implemented with MariaDB. In this

way, functions were developed to check if a certain record is already included in the tables in order to avoid data duplication.

In the case of migrating data of an installation from which monitored data is already stored in the individual database, the process is simplified, and only the last step is carried out, in which the new data is transferred to the tables of the individual database that had to be created previously.

In order to carry out the data migration process described above, it is important to analyse the structure of the files generated in the monitoring of PV plants, which may vary from one inverter model to another. These files can have different formats, specifically, in this paper, both CSV and XML format files have been worked with. For each type, their structure is usually standardised, so it was possible to automate the reading and migration process.

Figure 1 shows an example of how the information from a set of inverters and string current metering devices of one of the PV installations analysed in this work is organised. In this case, the files are stored in a directory and they are identified by means of a code (wb150014229 in Figure 1), which identifies the plant and the inverter group to which the data refer, together with the year, month and day (yyyymmdd, 20220704 in Figure 1), and separated by a hyphen, the time at which the storage of such data took place (hhmmss). These files are zipped, and within them, there are other files, as can be seen in Figure 1. Some are log files, which are not used except in case of problems, and other files are called Mean, which in this case only have in the name allusion to the date and time at which the data they store were recorded. These files are also compressed, and if they are decompressed, an XML format file is obtained containing the information of interest recorded by the inverters and string current sensors. Depending on the group of inverters, it may also contain the measurements of the irradiance and temperature sensors that were recorded in the plant. Figure 1 shows that the Mean files corresponding to 14:45:39 and 14:50:39 were unzipped.

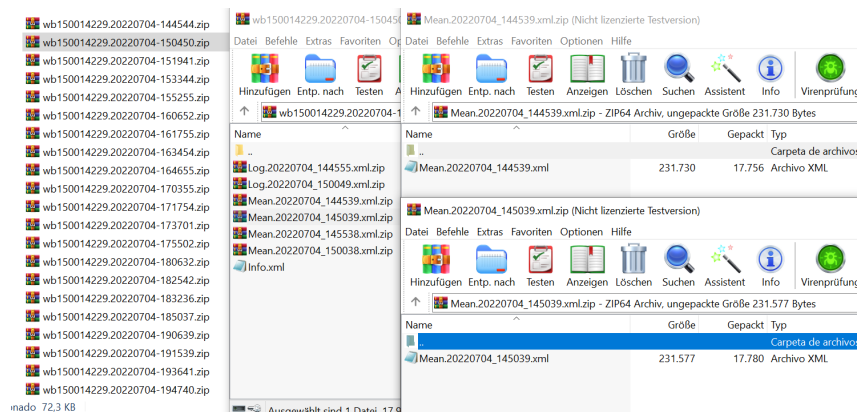


Figure 1. Organisation and format of the files corresponding to the monitoring of a sector of a PV plant during some hours of a specific day.

Each of these XML files corresponds to the record, in a set of inverters and string current sensors, of all the parameters they monitor in each sampling period, i.e., each XML file saves the data that were recorded in 5 min (the system was configured in such a way that the metering devices are recording information and save a value of each parameter recorded every 5 min). A fragment of the information stored in one of these XML files is shown in Figure 2, which shows, for the inverter model SCC100N7, with identification code 139028664, the recording of the parameters E_{total} , fac and $fehler$, at the time instant 14:45:39. The rest of the XML file contains the rest of the parameters recorded by this inverter, as well as by other inverters located in the same sector of the PV plant, and by string current sensors on the same sector, whose data are stored together in the same file.

```

Mean20220704_144539.xml: Bloc de notas
Archivo Edición Formato Ver Ayuda
</MeanPublic>
<MeanPublic>
  <Key>SCC100N7:139028664:E-Total</Key>
  <First>2416974.4</First>
  <Last>2416979.6</Last>
  <Min>2416974.4</Min>
  <Max>2416979.6</Max>
  <Mean>2416977.016667</Mean>
  <Base>6</Base>
  <Period>300</Period>
  <TimeStamp>2022-07-04T14:45:39</TimeStamp>
</MeanPublic>
<MeanPublic>
  <Key>SCC100N7:139028664:fac</Key>
  <First>50.000</First>
  <Last>50.000</Last>
  <Min>49.970</Min>
  <Max>50.000</Max>
  <Mean>49.985</Mean>
  <Base>6</Base>
  <Period>300</Period>
  <TimeStamp>2022-07-04T14:45:39</TimeStamp>
</MeanPublic>
<MeanPublic>
  <Key>SCC100N7:139028664:Fehler</Key>
  <First>0</First>
  <Last>0</Last>
  <Min>0</Min>
  <Max>0</Max>
  <Mean>0</Mean>
  <Base>6</Base>
  <Period>300</Period>
  <TimeStamp>2022-07-04T14:45:39</TimeStamp>
</MeanPublic>

```

Figure 2. Fragment of the information contained in an XML file with the information record of a group of inverters of a PV plant.

In these XML files the data are enclosed in *MeanPublic* tags, where *Key* includes the inverter and parameter information, separated by a colon. In addition, the file itself also includes the time at which the data were recorded, in the *TimeStamp* tag, which is very important when storing and processing the data.

Since this file corresponds to a PV plant with 17 inverters, similar files with data from other inverters and string current sensors in the plant will be found in another directory. However, as already mentioned, only in some of these files is the information corresponding to the meteorological parameter records stored. In the case of a small PV plant with few inverters, all the plant's inverter data are grouped in a single directory and will generate only one XML file every 5 min.

In general, it can be said that, for each day, 288 useful files will be available for each plant or set of inverters, in which information is recorded every 5 min throughout the day. Such disaggregation of the information makes it very difficult to process, so the tool developed in this work allows the automated loading of all this information and to have access to it in a much more orderly, selective and practical way.

However, not all inverters generate the information with this structure. There are cases of PV plants in which the metering devices directly generate the measured data in CSV format files. Moreover, in this case, the CSV files are not compressed and the structure of the directories that contain them is different, as is the structure inside the file (Figures 3 and 4). In this case, a single daily CSV file is generated, with all the records of each day, with the particularity that, inside the files, there is no reference to the day to which the data correspond; the date only appearing in the name of the file (Figure 3). Therefore, the script programmed to migrate the data from these files to the databases must be different from the one developed for the XML files.



Figure 3. Directory containing CSV files with the monitoring of a set of nine inverters corresponding to a sector of a PV plant.

	A	B	C	D	E	F	G	H
1	CSV-Export	Version: 1.01 Separator: Semicolon						
2								
3		SENS0500	SENS0500	SENS0500	SENS0500	SENS0500	SENS0500	WRSKA-03 V
4		4409	4409	4409	4409	4409	4409	2000412530 :
5	TimeStamp	ExtSolIrr	IntSolIrr	Optm	TmpAmb C	TmpMdul C	WindVel m/	E-Total F
6	hh:mm	W/m^2	W/m^2	h	°C	°C	m/s	kWh H
7	0:00	0	0	7256.49	12.81	12.53		0
8	0:05	0	0	7256.57	12.74	12.42		0
9	0:10	0	0	7256.65	12.62	11.95		0
10	0:15	0	0	7256.73	12.45	11.49		0
11	0:20	0	0	7256.82	12.27	11.11		0
12	0:25	0	0	7256.9	12.28	11.09		0
13	0:30	0	0	7256.98	12.41	11.31		0
14	0:35	0	0.06	7257.07	12.49	11.75		0
15	0:40	0	0.07	7257.15	12.39	11.98		0
16	0:45	0	0.19	7257.23	12.41	12.13		0
17	0:50	0	0.13	7257.32	12.47	12.23		0
18	0:55	0	0.2	7257.4	12.51	12.19		0
19	1:00	0	0.19	7257.48	12.44	11.86		0
20	1:05	0	0	7257.56	12.44	11.66		0

Figure 4. Part of the contents of a CSV file opened in Microsoft Excel from the monitoring of a PV plant.

In both formats, the migration scripts were coded in such a way that each piece of data must be assigned to the parameter to which it corresponds, the inverter or sensor that records it, as well as the day and time to which the data corresponds. As already indicated, how this process was carried out will depend on the internal structure of the file (Figures 2 and 4).

In general, to process the data, a series of steps will be required in the migration script in order to have access to read the directories where the data are located, decompress the files if necessary, read the files where the data are located (for example, in the case of XML files, only those beginning with Mean), extracting information from the headers or first lines of the file, which are usually standardised, to know which parameters are being measured and to which inverters or sensors they correspond, in the case of CSV files, or extracting this information through the structure of the XML format files. It is also essential, for each piece of data, to extract the date and time to which it corresponds, which, as has been seen, can be both in the name of the file and in the information stored in its internal structure, depending on the file format. It should be noted that one of the fields that are stored in the common database referring to the characteristics of each PV installation is the coordinates corresponding to the geographical location of the plant. This allows the date and time of the data to be associated with the local time of that geographical location; and when the measured data are stored in the individual databases of each installation, they are obviously stored with the local time. Once all this information is located, all the parameters of each inverter and sensor are distributed and entered in the corresponding databases, and specifically, in the fields of the tables assigned to them.

Since each PV plant to be studied has its own particularities, it is important to adapt the migration script to each one so that it corresponds to the file format they have and adapts to the characteristics of each so that there are no errors in the data migration. For example, whether or not they have weather data, and if not, from where to read these values. This avoids mixing configurations and minimises errors.

2.4. Tools for the Development of a Web Application for Accessing Information Stored in the Databases

Once all the data is stored in the common and individual databases, it will be necessary to be able to selectively access those data that are of interest depending on the analysis to be performed with them. To facilitate this task, a web application was developed in this work in order to be able to access all the information required. In a previous version, a desktop application was developed for this purpose [80]. However, it subsequently became necessary to develop a web application to be able to access the information in a much more versatile way. In this application, which was called *S-lar 2*, there are a series of fields in a form, which allows the user to include the parameters they want to select and the corresponding time period, and from there the information that is demanded will be generated. This information will be displayed both graphically and in tabular form.

This application was divided into the creation of a RESTful REST API as a back-end and a web client that makes use of this service. It was developed using Python on the server side, and HTML, CSS and Javascript on the client side.

The development of the RESTful API REST Web service was developed in Python using the micro framework named *Flask* [89], which is a tool prepared for the development of web applications using the MVC (model-view-controller) pattern, or for the design of REST APIs in a fast way, which is the reason why it was used. The MVC pattern presents a way of working that allows the differentiation and separation of the data model (the data that the application will have, which are usually stored in DB), the view (HTML page) and the controller (where the web application requests are managed).

The recipient of the REST API is a web client or front-end, developed with HTML, CSS and Javascript. VueJS [90] was used for this client part, which is a progressive Javascript framework for building user interfaces. It is implemented as a component architecture, which allows applications to be divided into blocks with independent functionalities. VueJS was the chosen option because it allows the development of projects of any size, is quite free in terms of organisation, and has a fast-learning curve, supported by its excellent documentation. The community behind the project, and the fact that it is open source, make it very attractive, along with its performance and low memory requirements.

2.5. Procedure for Data Processing

Once stored in their corresponding tables of the individual database of the installation to be analysed, there is also the option of calculating, from the monitored data, a series of parameters such as the efficiency of the plant's elements [80]. For this, since the data from the meteorological sensors are required, the procedure must consider the possible particularities that may exist in each plant to access this information. The values obtained from all these newly calculated parameters will also be stored in their corresponding tables in the individual databases of each plant.

In addition, all the parameters, both measured and calculated, recorded with the sampling period that was already indicated as 5 min, can be calculated at an hourly or daily level, determining, depending on the parameter, average or aggregate values, as appropriate. Again, this information will also be stored in the individual database of each plant.

The web application referred to in the previous section was developed to show, in the form of both graphs and tables, all this information, both the information recorded on a punctual basis and the average or aggregate values. However, if a more complex analysis of the data is required, it is possible, once the application gives us access to the data, to download them in the form of CSV files. The use of these, by means of a series of scripts also developed in Python, allows us to treat them and obtain from them information that can be useful for the management of the plant. Although dealing with this subject in detail is beyond the scope of this paper and only a few examples will be shown, it should be noted that, for the analysis of the data, small scripts were also developed in which the

numerous libraries that this programming language requires were used, such as Numpy [91] or Pandas [92] to process the selected data, Seaborn [93] or Matplotlib [94] for their graphical representation, or Scikit-Learn [95] to carry out modelling or classification processes [96].

3. Results

A schematic diagram representing the structure of the system developed in this work is shown in Figure 5.

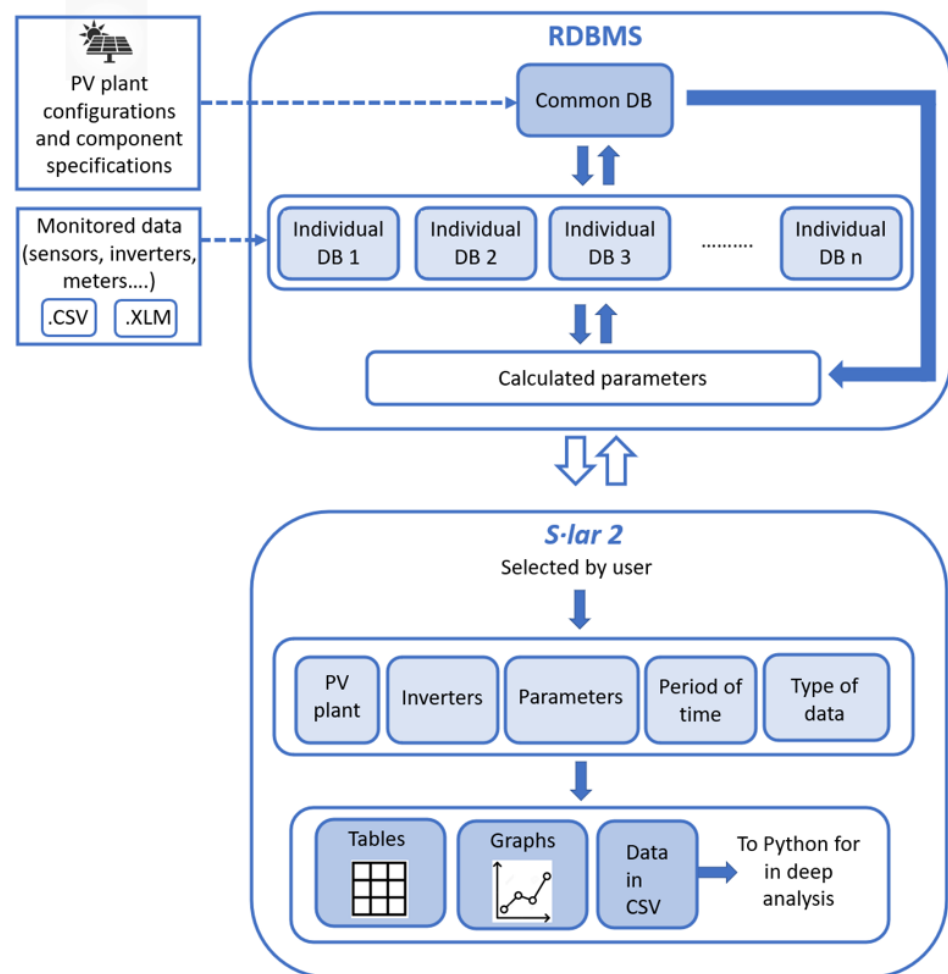


Figure 5. Schematic diagram representing the structure of the developed system for data analysis from PV plants.

3.1. Structure of the developed databases

As indicated in the previous section, the RDBMS has a common database and a series of individual databases corresponding to each PV installation analysed or each sector of them in the case of plants with a larger number of inverters. A schematic diagram of this RDBMS is shown in Figure 6.

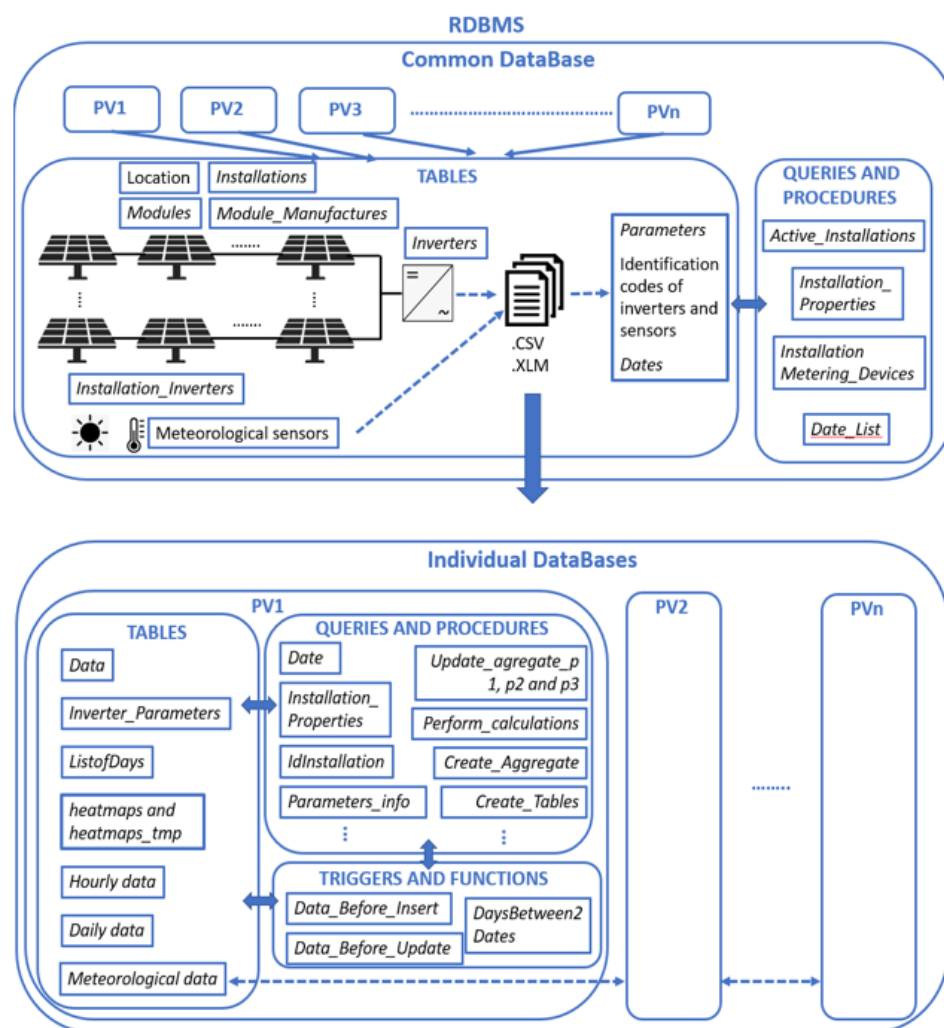


Figure 6. Schematic diagram representing the structure of the RDBMS.

The main tables included in the developed common database are shown in Table 4. As previously mentioned, the information in some of these tables, such as the *Installations*, *Modules* and *Inverters* tables, must be filled in manually, and the information corresponding to some of them, such as the *Parameters* and *Installation_Inverters* tables, is filled in automatically when the data file migration process takes place.

Table 4. Main tables included in the common database.

Table Name	Description
<i>Spanish_Municipalities_and_Provinces</i>	To facilitate entering the location of the PV installations to be analysed
<i>Installations</i>	Includes all fields that provide insight into all the characteristics of the installation.
<i>Module_Manufacturers</i>	List of trade names of manufacturers of PV modules.
<i>Modules</i>	List of the modules commercialised by different manufacturers, with their specifications: Material, Number of cells, Module surface, Nominal power of the module, Nominal Operating Cell Temperature (NOCT), Temperature coefficients, Open circuit voltage, Short circuit current, Voltage and Current for maximum power.

<i>Inverters</i>	Technical characteristics together with their commercial model and name, of inverters whose monitored data are inserted into the RDBMS during data migration.
<i>Installation_Inverters</i>	Identification of all inverters, among those in the <i>Inverters</i> table, that are in each of the PV plants to be studied, together with the number, orientation and tilt of PV modules associated with each one.
<i>Parameters</i>	List of the names of monitored parameters from inverters as well as all of the calculated parameters. The table is automatically completed by the <i>S-lar 2</i> application during the process of data migration. As well as the name of the parameters and their units, other fields included are Colour (store the colour code which graphically represents each record), Calculated (differentiates between monitored and calculated parameters), Aggregate (for the correct selection of the aggregate or average values of parameters measured and calculated).
<i>Dates</i>	Auxiliary table that stores the oldest and the most recent monitored data from each installation. It is updated when running the procedure for calculating aggregate data.

Several queries have been included in this common database, some of which are specifically prepared for the operation of the developed web application. One of them is called *Active_Installations*, which shows the list of PV installations for which data is stored in the databases and which will be accessible from the web application. Another important query is *Installation_Properties*, which shows a complete summary of the specifications of active installations. Another query worth mentioning is *InstallationMetering_Devices*, which generates a list of all the measuring devices that record the data in the different PV installations, with their identification codes, together with their characterisation, such as the associated modules, their nominal PV power, or the useful surface in m² of PV cells that modules connected to each monitoring device have. This is a query used by the web application to display all the information corresponding to the installation that has been selected by the user. It is also used to calculate parameters from the monitored data such as module or inverter efficiencies.

The common database also includes a procedure called *Date_List* in which, starting from two dates, you have the list of all the days between them.

The individual databases dedicated to storing the monitoring data of each PV plant or one of its sections are composed of the tables shown in Table 5. The *Installation_Parameters* table contains the list of parameters that are measured by each inverter. Its information facilitates the extraction of data from the database once they are already stored, since it allows the verification of whether a parameter is measured by an inverter, in order to proceed or not to export it. This is because not all inverters monitor the same parameters. The *ListofDays* table contains the list of all the days for which data are available for that installation, from a minimum date or the oldest date for which data are available, to a maximum or most recent date. This table makes it easy to extract the values of each parameter, once stored, in the form of arrays. These arrays will have, in one dimension, the days corresponding to the entire time period for which data are available and are to be analysed, and in the other dimension the hours of the records throughout each day. This type of array facilitates access to the entire set of values of each parameter in a given time period and makes it easier to perform daily or hourly statistics, as well as to perform a graphical representation of each parameter in the form of colour maps or heatmaps, representations that allow, in a practical way, the visualisation of the recorded value of a parameter at each instant of time throughout a given time interval. In case of any error in

the data recording, the missing days and hours will appear in the data array, although with empty data, so that the entire time period is complete in the array, regardless of whether there is data measured or not. This also facilitates that, if the value of two different parameters is to be compared or related, the time period of both parameters can be the same. The tables named *heatmaps_tmp* and *heatmaps* are in charge of storing the data in these arrays.

Table 5. Main tables included in the individual databases.

Table Name	Description
<i>Data</i>	Stores the values of the instantaneous parameters monitored by each inverter and additional sensors of the individual PV plant.
<i>Inverter_Parameters</i>	Stores the name of the parameters monitored by each particular inverter.
<i>ListofDays</i>	A list of all days between a minimum date and a maximum date between which data is recorded and stored in the individual database.
<i>heatmaps_tmp</i>	A temporal table with an array structure where all the records measured every 5 min of a given parameter in a given time range are stored.
<i>heatmaps</i>	Has the same structure as <i>heatmaps_tmp</i> . Its utility is to obtain a complete array in which all the records of the previous table parameter are contained, but including, with empty values, those instants in which there was no data record due to some failure. Needed for array data migration, using <i>heatmaps_tmp</i> and <i>ListofDays</i> tables, in order to have a complete range of data, even in the moments in which there could be no monitored data.
<i>Hourly_Data</i>	Stores aggregated or average hourly data
<i>Daily_Data</i>	Stores aggregated or average daily data.
<i>Meteorological_Data</i>	Stores the data monitored by the weather parameter sensors.

The data structures of the tables named *Data*, *Hourly_Data* and *Daily_Data* are variable and unique for each installation, since they depend on the parameters measured in each one. However, they all include the date (year, month, day) and hour and the identifier code of each inverter or sensor. These structures are defined exactly in the initial stages of data verification during the data migration process, according to the characteristics of each installation and the process by which two SQL scripts are generated to create all the tables, views and functions, thus completing the physical model of the databases.

The individual database may or may not have a table called *Meteorological_Data*, which is a table with a variable structure whose columns are the weather parameters that are recorded in the plant. Its usefulness is the weather data is made available to the rest of the inverters to make it possible to calculate some parameters such as the efficiency of the PV panels, for which the values of the irradiance received need to be known. When these parameters have not been measured in an installation, it is possible to share the meteorological data of another nearby installation (for example, through a query).

Each individual database of each PV plant or a section of it is completed with a series of queries or views involving tables and queries from other databases. These queries are listed in Table 6.

Table 6. Main queries included in the individual databases.

Table Name	Description
<i>Dates</i>	Checks the minimum and latest dates for which data is available in the installation database.
<i>Installation_Properties</i>	Provides the specifications of each individual plant, obtained for each plant from the query with the same name in the common database.
<i>IdInstallation</i>	Allows access to the installation's identification code.
<i>Parameters_info</i>	Allows the complete list of all the parameters (and their descriptions and characteristics) of the installation to be known. Extracted from the structure of the <i>Data</i> table, which is linked to the <i>Parameters</i> table of the common database.
<i>InstallationMetering_Devices</i>	Retrieves the specifications of all the metering devices that record data in the individual plant, from the query <i>InstallationMetering_Devices</i> of the common database.
<i>_idInstallationInverters</i>	Provides the identification code of each plant's metering device that is an inverter.
<i>InstallationInverters</i>	It provides the specifications of each plant's metering device that is an inverter.
<i>Weather_Data_Code</i>	Allows the identifier code of the measuring equipment that records the weather data to be known. In those installations that do not have such data, this query will refer to the inverter identifier code of the installation from which these weather data are taken.
<i>Original_Parameters</i>	Allows the list of all the parameters measured in the installation to obtain their identifier, name and unit to be known.
<i>p_irradiance</i>	Returns the name of the field that is identified with the irradiance. This is because not all measurement equipment records this parameter using the same notation. It is used for the calculation procedure.

The individual databases are completed with a series of procedures, which are listed in Table 7.

Table 7. Main procedures included in the individual databases.

Table Name	Description
<i>Update_Aggregate_p1</i>	This is the first of the three parts into which the process of updating calculations and aggregate data has been divided. In this p1 the minimum and maximum dates of data in the <i>Dates</i> table of the common database are updated. The <i>ListofDays</i> table is updated by calling a procedure called <i>Date_List</i> . And finally, if applicable, the new weather data is inserted into its corresponding table, before continuing with step 2.
<i>Update_Aggregate_p2</i>	Transfers the meteorological information to the rest of the inverters, in order to be able to contemplate the efficiency calculations. It ends with the call to step 3.
<i>Perform_Calculations</i>	Function that performs parameter calculation.
<i>Update_Aggregate_p3</i>	Inserts aggregated data at an hourly and daily level

<i>Create_Aggregate</i>	Creates the tables with the correct structure for hourly and daily aggregate data, starting from the Data table.
<i>Create_Tables</i>	Creates the tables necessary for the correct functioning of the application for data access. It must be executed only once, just after the data migration. Retrieves the installation identifier code, which will be used several times; it updates the minimum and maximum dates of the installation data, it creates the <i>ListofDays</i> table, and it creates the <i>Meteorological_Data</i> table for the meteorological data of the installation using the <i>Weather_Data_Code</i> query.
<i>Delete_Data_Date</i>	Responsible for deleting data from all tables after a certain date.
<i>Date_List</i>	Searches for the dates between which data is stored for a given inverter, and fills the <i>ListofDays</i> table with a list of all these days. It is used in the <i>heatmaps2</i> query.

Lastly, two *triggers* have been incorporated (*Data_Before_Insert* y *Data_Before_Update*) used during data insertion in the migration that perform the same function: fill in the time from the date data.

Regarding the *functions*, only the one called *DaysBetween2Dates* has been included. Its purpose is to generate a list of all the days between two dates, and it is used in the procedures called *Date_List* of each installation, where the dates needed are the oldest and most recent dates for which there are data, which are stored in the query called *Dates*.

3.2. Web Application for Accessing Information

As mentioned above, in order to have more convenient access to all the data stored in the different databases of the various PV installations, an application has been developed to facilitate the process of visualisation and extraction of information.

To open the web application, it is necessary to have the database server running and to run the web server. Both are local servers and can be accessed from any browser, preferably Chrome or Firefox.

The application has a start menu. In this menu, there are some pop-ups or scrolling lists that allow the user to choose the information to be extracted from the databases. First, at the top of the start menu, there is a list that enables the selection of the PV installation for which the analysis of some of its parameters is to be carried out. This selection will be made among those PV plants that are active in the application, which are those for which data are available in the common and individual databases.

Once the PV plant has been selected, the screen will display information about its characteristics, such as its nominal power, its location, the date the plant was created, and the minimum and maximum dates between which data for this plant is available in the database.

For the consultation of a specific parameter, the time period for which the information is to be consulted must be selected, and for this purpose, there are buttons to select the start date and the end date of the time period. In Figure 7, the time period chosen is from 01/11/2020 to 04/07/2022. In addition, there is a *Clear* button to clear the data form of the information previously selected.

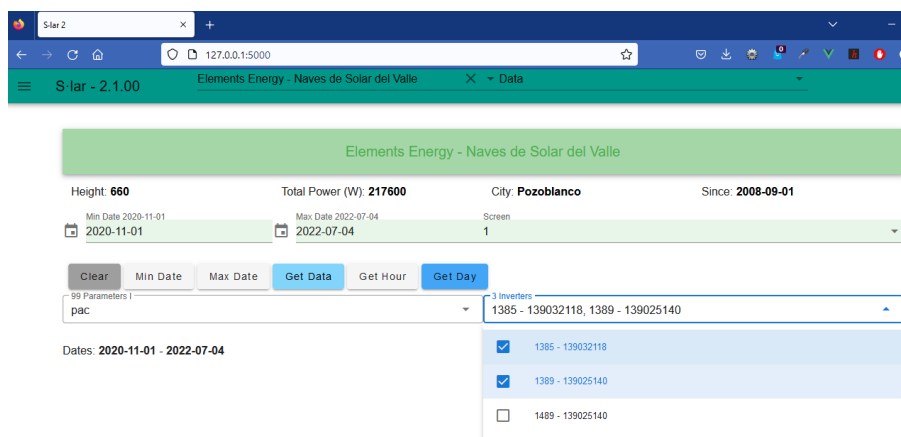


Figure 7. Image of the *S-lar 2* application screen, in which the selection of two inverters of the selected PV plant (named Element Energy) is performed.

The available inverters in the selected plant are also displayed on the screen. The user has the option of selecting one, several, or all the inverters of the plant (Figure 7). The selection is carried out by means of the inverter identification codes, which are shown in the scrolling menu. In the PV plant selected in Figure 7, there are two inverters, but three are shown because, during the period for which data are available for this plant, one of the inverters was replaced by another, and the identification code of the new one was different from the identification code of the inverter that was replaced.

Once one or several inverters have been selected, another list or menu can be scrolled down, showing all the parameters that have been measured by the selected inverters (Figure 8). In this menu, it is possible to select the measured parameters in which the user is interested. In this selection, it is possible to choose one, several, or all the parameters displayed. However, the system is configured to be able to select either the same parameter measured by several inverters or different parameters from the same inverter. Since only those parameters that have been actually measured are shown in the pop-up menu, it is not possible to request from the database parameters for which no record is available, thus avoiding errors.

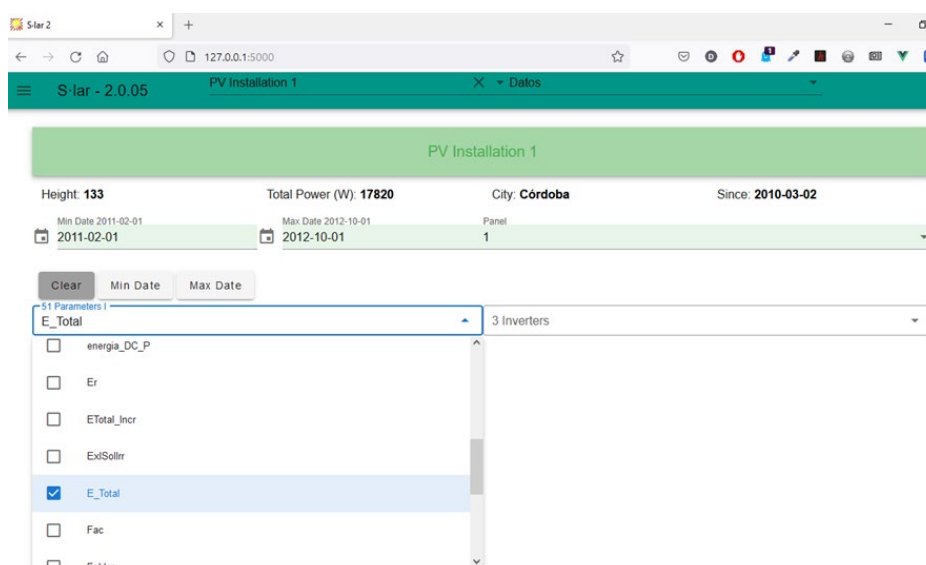


Figure 8. Image of the *S-lar 2* application screen, in which the parameters measured by the chosen inverters can be selected.

With the PV plant, inverter and parameter already selected, the different possibilities for displaying the information are shown, which can be the punctual values measured every 5 min (*Get Data* button in Figure 7), the hourly aggregated or averaged values (*Get Hour* button) or the daily aggregated or averaged values (*Get Day* button).

Once the data are received by the back-end of the application, the data can be displayed graphically, as shown in Figure 9. The *Get More Data* button allows more data to be displayed than is shown on the screen.

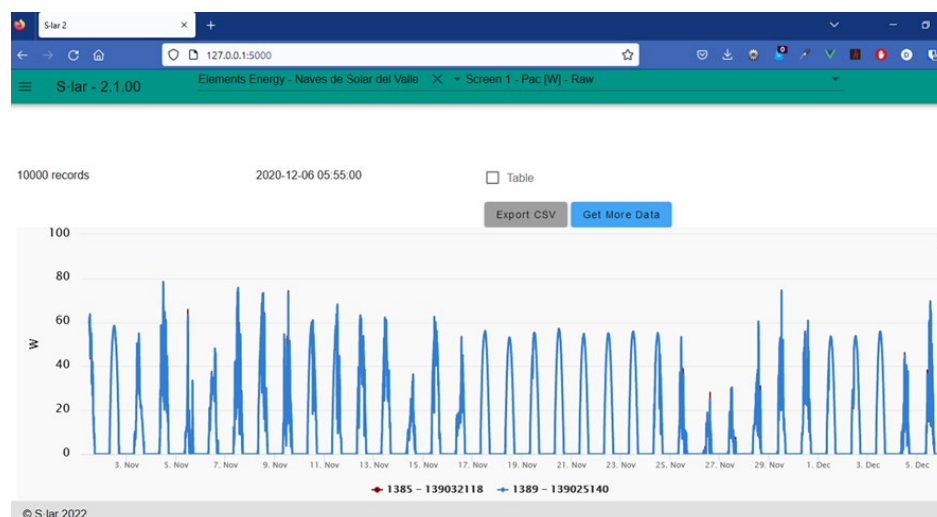


Figure 9. Image of the *S-lar 2* application screen, showing a graphical representation of the power in the output of the two selected inverters (P_{ac}) for the chosen time period.

The selected data can also be displayed on the application screen in table format (Figure 10). The application also gives the user the option to download the selected data set displayed on the screen in a CSV format file in order to perform a more in-depth analysis of the data than just displaying the monitored or averaged data on the screen of the application. This is possible using the button named *Export CSV* that appears when the selected data is shown in Table form (Figure 10). In this case, a CSV file will be generated for each parameter of each inverter that has been selected in the application, in which the data will be in the form of an array, with the days of the selected period in one dimension and in the other the 288 records of the parameter that is performed each day if it is punctual data. In the following section of the paper, some examples of analysis of the data downloaded through the application for some of the parameters measured by the inverters of two PV installations will be shown.

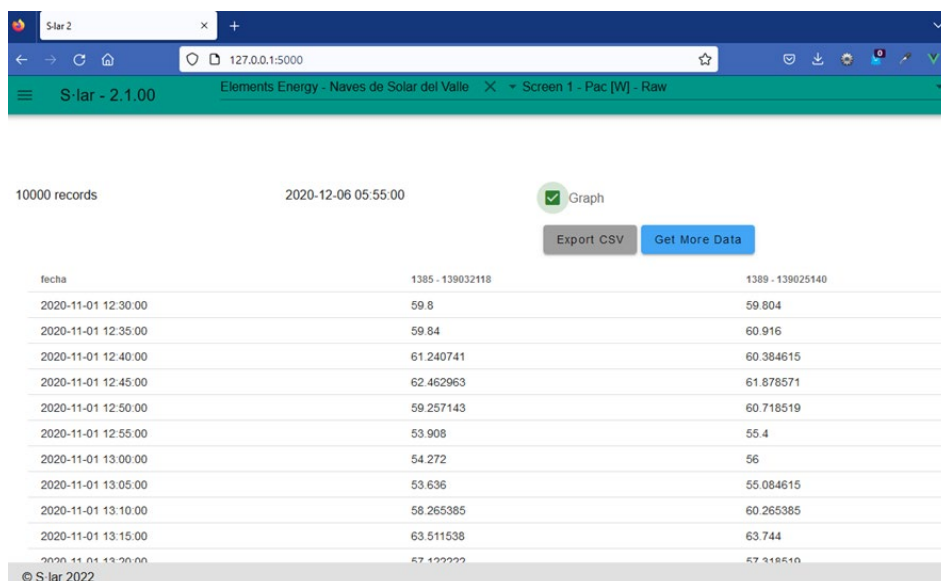
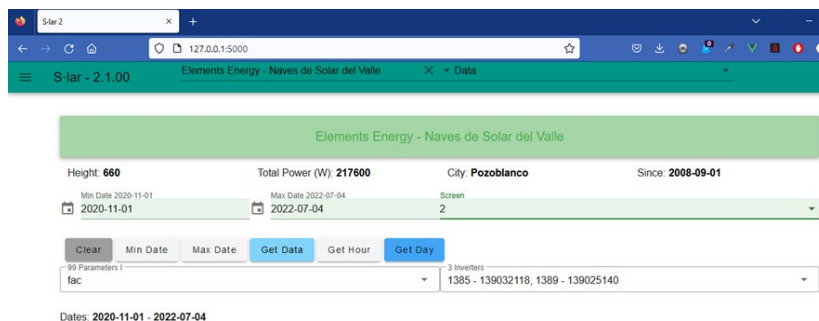
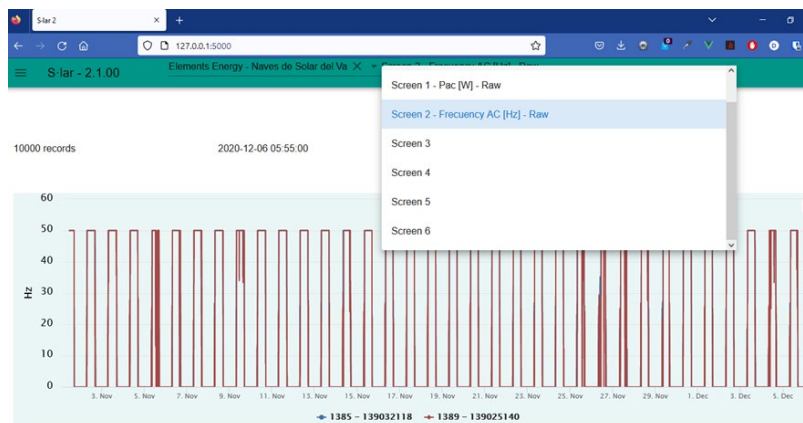


Figure 10. Image of the *S-lar 2* application screen, showing in tabular form the values of one of the parameters (P_{ac}) of the two selected inverters.

If the user wants to display data for other parameters, another screen or panel must be selected in the application. Then, the selection of the parameter in which the user is also interested is made again in the forms or menus of the application, and the data can also be displayed, as can be seen in Figure 11. Six screens are available for displaying different parameters



(a)



(b)

Figure 11. Image of the *S-lar 2* application in which the grid frequency f_{ac} has been selected for two inverters (a), and the values of this parameter are shown graphically but now on screen 2 (b).

Although a similar application was available in previous developments of the tool [80,97], this version has several improvements, among which the following may be mentioned. In the current version, it is possible to migrate data in XML format, which was not considered in previous versions. The calculation process of the calculated parameters, such as the efficiencies of the elements of the installations, has been optimised, for which it has been necessary to design a new structure for the individual database of each of the installations, and now this calculation is performed in the database itself when previously it was done simultaneously with the migration process, which made this process much slower. The data export routines have also been improved with the introduction of new tables, and, although each database now has a unique data structure, depending exclusively on its physical configuration, this avoids the occurrence of errors in the data migration and extraction process.

3.3. Use of the Developed Application in the Analysis of Some Parameters of Two PV Plants

Although it is beyond the scope of this paper to carry out a detailed analysis of the behaviour of the elements of PV installations, this section includes a brief analysis of some of the parameters recorded in two PV installations and which have been organised using the RDBMS developed in this work. The aim is to show that, once the data are correctly stored, it is easy and convenient to access them selectively and obtain useful information.

The technical characteristics of the PV plants analysed are shown in Table 8.

Table 8. Main technical characteristics of PV plants analysed.

Table Name	Plant 1	Plant 2
Total power (kW)	217.6	1843.2
Number of inverters	2	17
Inverter model	SMA SC-100	SMA SC-100
Number of modules per inverter	640	640
Module power connected to each inverter (kW)	108.8	105.6 and 108.8
Total number of modules	1280	10,880
Module model	BP-3170	BP-3165 and BP-3170
Module nominal power (W)	170	165 and 170
Single-phase or Three-phase	Three-phase	Three-phase
Rooftop or ground	Rooftop	Ground
Geographical location	Pozoblanco	Pozoblanco
Rural or city area	Rural, industrial area	Rural
Latitude	38°21'33" N	38°20'8"N
Longitude	4°50'40" O	4°49'41" O
Tilt	Adapted to the ground	9°
Orientation	South	10° deviation respect South

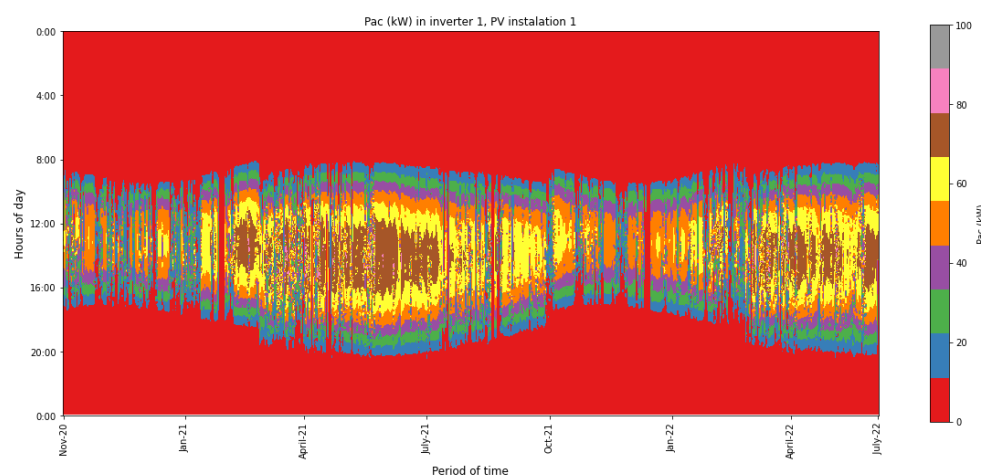
The data monitored at these plants were provided by the company that operates and maintains them, in the form of files in XML format organised as previously described in section 2.3. In the case of plant 1, the information recorded every 5 min by the two inverters is collected in the same XML file, while for plant 2, the monitored data is separated into two blocks, one containing the XML files with the data from inverters 1 to 8, and the other with the files containing the data from inverters 9 to 17. To organise the information, an individual database for the first PV plant and two databases for plant 2 have been set up to store the data from the two sets of inverters.

The time period from 01/11/2020 to 04/07/2022 has been analysed, which corresponds to a total of 611 days. If, for each set of inverters, 288 XML files with information monitored at the plants are recorded every day, for this period this results in three sets of 175,968 files or a total of 527,904 files. This amount of data becomes unmanageable without a system to manage it optimally and efficiently.

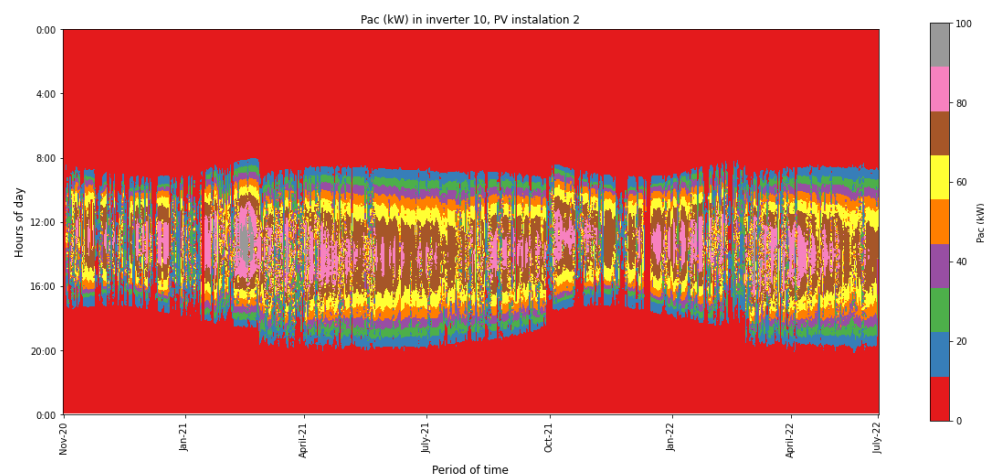
All these files were then loaded into their corresponding databases, saving each parameter in the corresponding tables. The specific parameters to be analysed were then extracted into CSV format files with an array structure using the previously described *S-lar2* application.

These arrays were imported, and the information was processed by means of small scripts developed in Python, which allow us to show their behaviour in the chosen period. It is important to remember that, in the arrays, in one dimension there are each of the days of the time interval of more than 600 days mentioned, and in the other, there are the 288 daily data records, from 00:00 to 23:55, with a frequency of 5 min.

Figure 12 shows directly the values of the arrays corresponding to the P_{ac} power recorded at the output of the inverters, in the case of inverter 1 of plant 1 (a) and inverter 10 of plant 2 (b). It can be seen that it is possible to conveniently observe the behaviour of all the values of this parameter recorded in the chosen time period, observing, in general, the variation of this magnitude, which during the night hours is null, and its value increases from sunset to the zenith, at midday, decreasing its value again until sunset. These graphs in Figure 12 show the greater number of hours of production in the summer months, and the presence of days with passing clouds, which causes the value of the power produced to fluctuate throughout the day. Although the PV panel power associated with the inverters represented for both plants is the same in this case, it can be observed that the deviation in the orientation of the panels on plant 1, conditioned by the fact that they are located on the roof of a building, with respect to the south orientation of the panels on plant 2, results in lower power production. The periods in which there were failures in the monitoring system were assigned a null value. The graphs obtained for the rest of the inverters on the two plants show very similar behaviour to that of the inverter on the same floor shown in Figure 12.



(a)

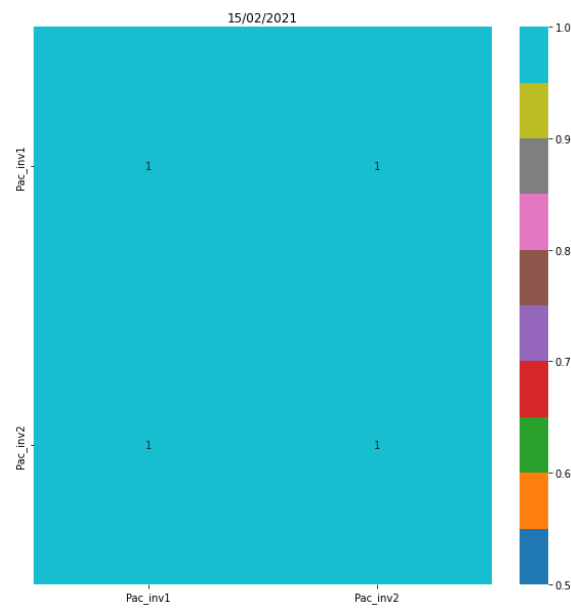


(b)

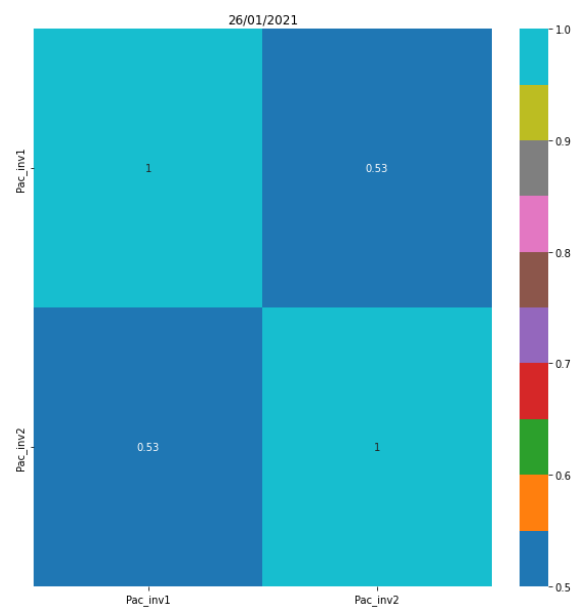
Figure 12. Instantaneous P_{ac} values recorded at the output of inverters 1 of plant 1 (a) and 10 of plant 2 (b), in the period from 1 November 2021 to 4 July 2022.

Although these graphs show the general behaviour of the P_{ac} production of each inverter, it is not always easy to visualise in them small deviations in the production of any inverter, which could alert us to problems in any of the elements of the plant, especially if the behaviour of numerous inverters must be supervised, as is the case of plant 2. Given that the inverters are close together and receive practically similar irradiance values, if their production is compared with each other, and it is not the same, this would be an indication of a failure or problem in one of the elements. To do this, a simple Python script was developed that asks the user for the date of the day to be analysed, and a graphical representation of a correlation matrix will be available, which, after correlating the production value of each inverter on the chosen day with that of the rest of the inverters in the plant, shows the value of the Pearson correlation index of these correlations. This coefficient measures the linear dependence of the production of each inverter with all the rest. The closer the value is to 1, the better the relationship between the output of pairs of inverters. If the value were 0, it would mean that there is no relationship between the output of any two inverters. The diagonal in these matrices appears by default with the value unity since they correspond to the comparison of the production of each inverter with itself. This type of graphical representation makes it easy to visually detect possible fault situations or small deviations that could lead to faults in the future.

As an example of these types of graphs, Figure 13 shows those obtained for two different days for inverters 1 and 2 of plant 1. On one of these days, 15 February 2021 (a), it can be seen that the correlation index of the production of the two inverters has the value of 1, which shows that the two inverters are producing power at the same rate. This is not the case on 26 January 2021 (b), where the value of the correlation coefficient is less than 1 (0.53), which indicates that one of the inverters did not produce the same amount as the other, indicating the presence of a fault in one of them. The colour variation in the graph also warns of this possible deviation.



(a)



(b)

Figure 13. Graphical representation of the correlation index matrix between P_{ac} values for inverters 1 and 2 of plant 1 on 15 February 2021 (a) and 26 January 2021 (b).

If once the fault has been detected by means of this low value of the correlation index, the P_{ac} power of the inverters on the days indicated is displayed, as can be seen in Figure 14, on the day on which the value of the correlation coefficient decreases (26 January 2021) it can be seen that there is a series of hours throughout that day in which there is no electricity production, in this case in inverter 2 (Figure 14b). In the case of the day on which the correlation indexes are equal to 1 (15 February 2021), it can be seen that there is the same power production in the two inverters of the plant (Figure 14a).

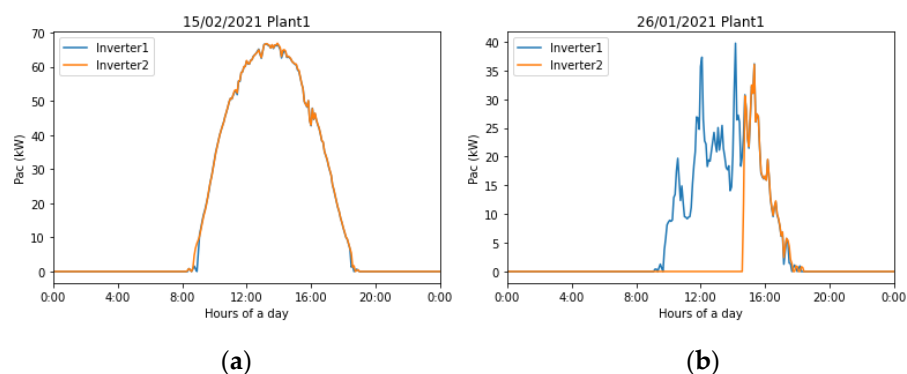
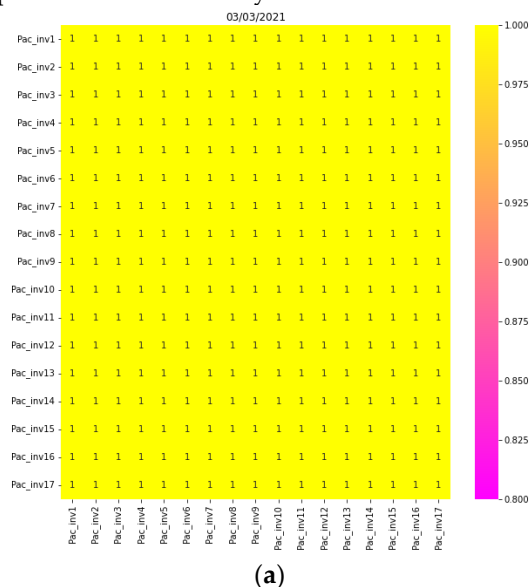


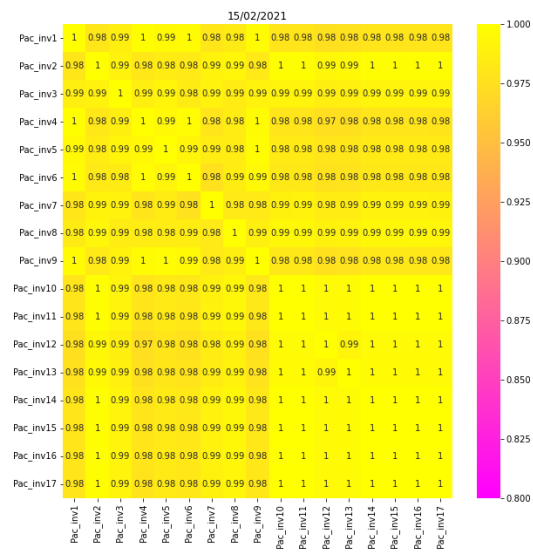
Figure 14. P_{ac} corresponding to inverters 1 and 2 of plant 1 on 15 February 2021 (a) and 26 January 2021 (b).

The correlation index matrix representation is even more useful in the case of installations where the number of inverters is higher, as in the case of plant 2. As an example, the representation of the correlation index matrices for the 17 inverters in the plant for three different days (3 March 2021 (a), 15 February 2021 (b) and 26 January 2021 (c)) is shown in Figure 13.

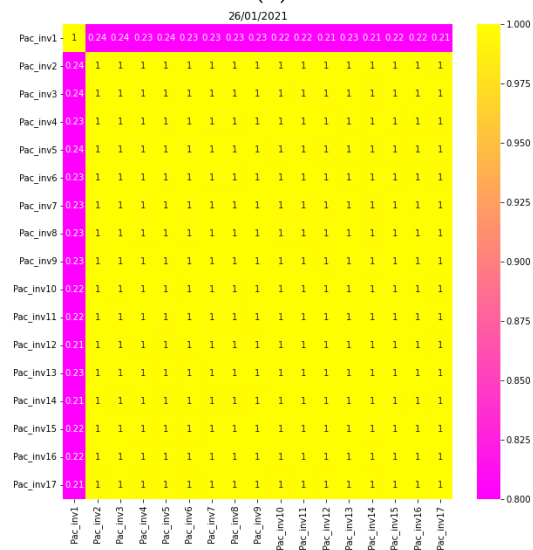
In Figure 15, it can be seen that, for 3 March 2021, the production of the 17 inverters is perfectly correlated, since the correlation indices all have the value of 1. This means that all 17 inverters in the plant are producing at the same rate, as shown in Figure 16, where the P_{ac} power produced that day by the 17 inverters is shown (Figure 16a,b). If we look at the correlation index matrix for the day 15 February 2021 (Figure 15b), we find that there are small deviations in the production of some inverters, but none of them are very significant. The colour variations of the representation also highlight these small deviations. This behaviour corresponds to the P_{ac} production of the 17 inverters shown for that day in Figure 16c,d. There are small deviations in the production, which result in the Pearson correlation index values of some inverters not having the unity value, but these are very occasional variations. Conversely, for 26 January 2021, it can be seen in the correlation matrix (Figure 15c) that one of the inverters, inverter 1, does not correlate its production with that of the rest of the inverters, given that the value of the Pearson correlation index is equal to 0.24. The colour in this case also warns of this low value. This is a sign of the presence of some kind of fault in this inverter, as it is evident from the graphical representation of the power of the inverters on that day (Figure 16e,f), in which inverter 1 had practically no power production on that day.



(a)



(b)



(c)

Figure 15. Graphical representation of the matrix of the correlation indexes of the power production at the output of the 17 inverters (P_{ac}) of plant 2 for 3 March 2021 (a), 15 February 2021 (b) and 26 January 2021 (c).

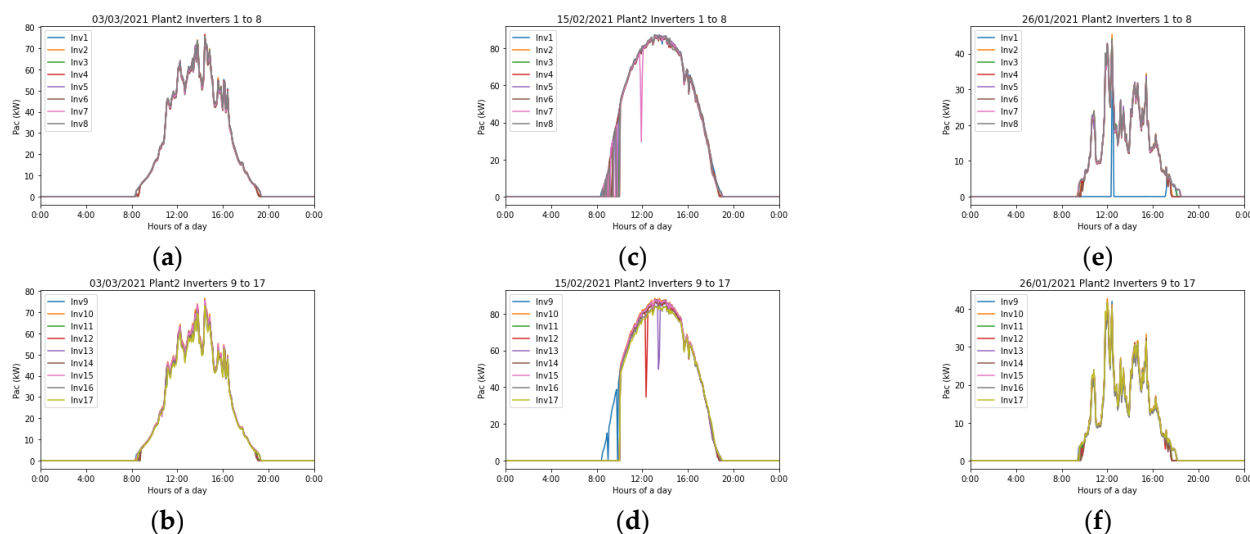


Figure 16. P_{ac} corresponding to inverters of plant 2 on 3 March 2021 ((a) for inverters 1 to 8, and (b) for inverters 9 to 17), 15 February 2021 ((c) for inverters 1 to 8, and (d) for inverters 9 to 17) and 26 January 2021 ((e) for inverters 1 to 8, and (f) for inverters 9 to 17).

Therefore, the representation of the correlation index matrices allows visual access to a comparison of the operation of all the inverters in a plant, and it can be a very useful tool, especially in plants with a larger number of inverters, detecting possible deviations in production and showing the greater or lesser magnitude of these possible deviations. Thus, just by visualising a series of numbers, it is possible to see whether the inverters are working correctly or whether there is some kind of problem. It would only be necessary to proceed to analyse in more detail the daily power values if these graphs show the presence of some kind of deviation in order to have more information about when the problem occurred.

In addition to the behaviour of the PV plant elements, another very important aspect to take into account during operation is whether their production is having any kind of impact on the electricity grid. In most plants, this is an aspect that is not recorded or analysed, but it is essential in order to be able to correctly integrate the production of this type of plant into the electricity grid without producing disturbances that could affect users [96]. Although a complete analysis of their impact on the electricity network would require grid analysers at the injection point of the plants (PCC), the inverters themselves can provide information on some parameters of the grid and their relationship with the production of the plants. For example, the inverters record the value of the voltage signal of the electricity grid (V_{ac}), so it is possible, with the values measured by inverters, and without the need for additional instrumentation, to carry out an analysis of the value of this parameter in the different phases. This parameter can be recorded by the inverters 24 h a day, even when no production is being injected into the grid, as is the case with the inverters analysed in this study. As an example, the graphs in Figure 17 show the value of this voltage V_{ac} in phase 1 measured every 5 min by two inverters, one from plant 1 (a) and the other from plant 2 (b), during the entire period analysed (01/11/2020 to 04/07/2022). The values between 225 and 250 V have been represented in order to better observe the behaviour of this parameter with the colour degradation. In both cases, the highest voltage values generally occur in the central hours of the day, and to a greater extent in the spring and summer months, which is when the production of the PV plants is higher in their geographical location. This behaviour is even more evident in the case of plant 2, which has a higher nominal power and therefore a higher total production injected into the grid than plant 1.

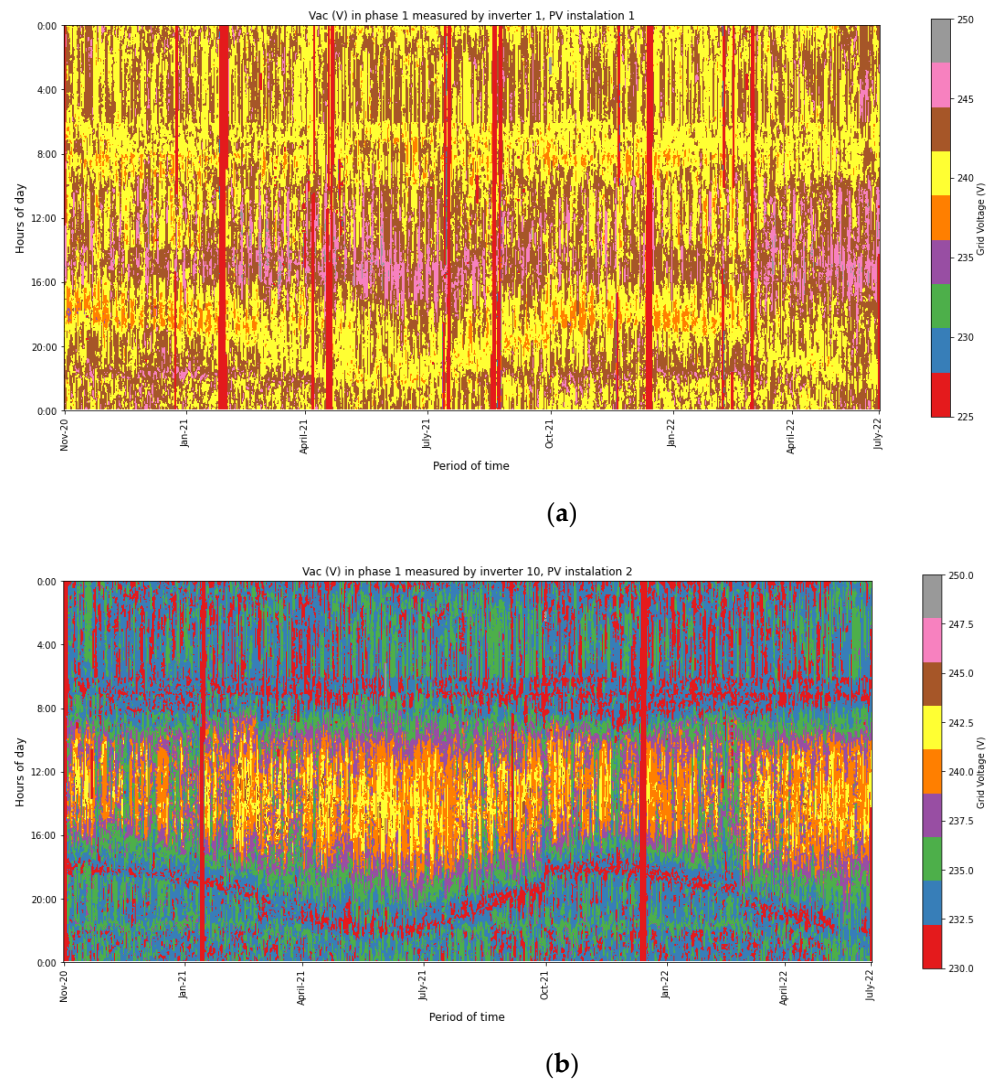


Figure 17. Grid voltage values (V_{ac}) measured every 5 min on phase 1 by inverters 1 (a) and 10 (b) of plants 1 and 2 respectively during the period 1 November 2020 to 4 July 2022.

For a more complete analysis, in addition to the instantaneous values of the V_{ac} voltage, it has been possible, using the tool developed, to easily determine statistical values of this magnitude, and in general of all the parameters recorded in the PV plants. The maximum, minimum and average values of the voltage measured from the grid have therefore been determined for each day. The obtained results are shown in Figure 18. The two graphs in this figure show the average daily values of the V_{ac} grid voltage measured in phase 1 by an inverter of PV plant 1 (a) and by an inverter of PV plant 2 (b). Also shown here in blue is the range of values corresponding to the difference between the maximum and minimum daily values of the recorded grid voltage. It can be seen that, in the case of plant 1, the value of the grid voltage measured at its PCC is higher than at the second plant. Even though plant 1 is smaller in size or power rating, the value of the V_{ac} voltage will depend primarily on the state of the grid at that point, which will be conditioned by all the elements connected to it, and not only by the PV installation. It can be also observed that there are days in which lower voltage values were registered, and occasionally, cases of overvoltage values. The variability throughout the year is not very significant, although a slight increase can be observed during the summer months.

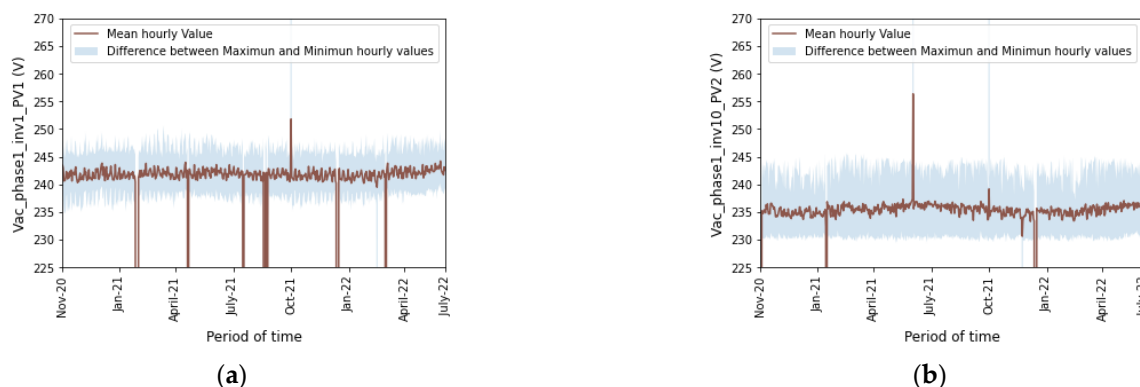


Figure 18. Average values (brown) and the difference between daily maximum and minimum (blue) of the V_{ac} grid voltage measured on phase 1 by inverter 1 of plant 1 (a) and inverter 10 of plant 2 (b), in the period from 1 November 2020 to 4 July 2022.

The maximum, minimum and average voltage values have also been determined for the entire period analysed, but in this case for one of the 288 moments of the day in which the data is recorded, i.e., every 5 min. The results obtained can be seen in the graphs in Figure 19, which show the V_{ac} values in phase 1 measured by inverters 1 (a) and 10 (b) of plants 1 and 2, respectively. It can be seen in this case that the overvoltage values recorded and shown in the graphs in Figure 16 occur in this case at times of the day when the PV plants are not in operation. If these overvoltage values had occurred during production hours, the inverter would have interrupted the injection of energy until the voltage values had returned to the established ranges.

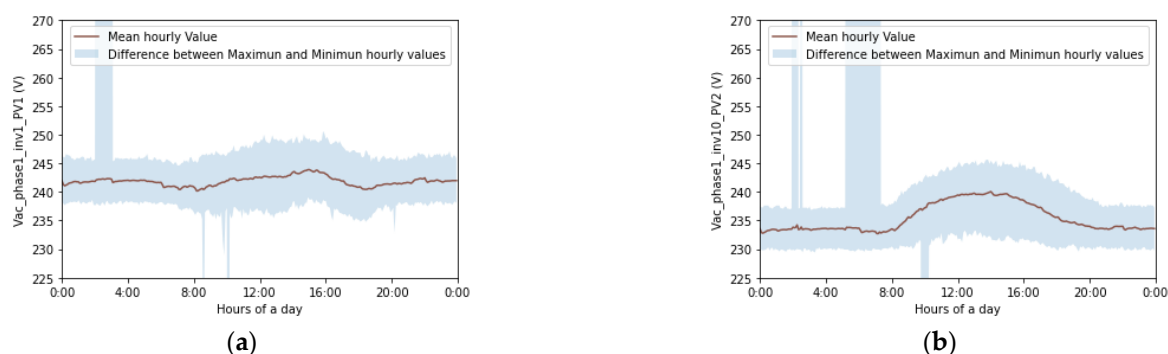


Figure 19. Average values (brown) and the difference between daily maximum and minimum (blue) for the period from 1 November 2020 to 4 July 2022 of the V_{ac} grid voltage measured in phase 1 by inverter 1 of plant 1 (a) and inverter 10 of plant 2 (b), in each of the 288 daily records, with a frequency of 5 min.

Furthermore, the graphs in Figure 19 show very clearly that the grid voltage value shows a slight increase during the hours when the PV plants are in operation. This behaviour is more significant in the case of installation 2 (Figure 19b), as already shown in Figure 17, as it has already been indicated that it is larger. Although the V_{ac} voltage values are within the margins accepted by the standards, the grid at these points will have a hosting capacity, so that if the power values of the installations increase, so would the value of the grid voltage, so that the installed power could never be greater than that which would cause the value of the V_{ac} voltage to exceed the value recommended by the standards.

These examples show that, once the data values measured by the different elements of the PV plants have been correctly ordered and stored using the database management system developed, it is possible to quickly analyse the measured data to obtain information on both the behaviour of the plant elements and their influence on the electricity

grid in order to make decisions regarding their O&M for the benefit of both the plant owners and the grid users.

4. Conclusions

The increasing number of PV installations connected to the grid requires that their operation be perfectly characterised in order to improve their efficiency and control their impact on the electricity grid. For this to be possible, the plants must be monitored and generate data that, after processing, can be converted into useful information to help in the O & M of these renewable sources and the management of their injection into the grid. However, the large amount of data generated in the monitoring of even small plants makes it essential to use tools that help in the management and processing of all this data.

For this purpose, in this work, firstly, an RDBMS was implemented to store in an orderly manner all the information corresponding to the technical specifications of the plants to be analysed, and a process was developed to automatically load all the data measured in them into the database tables created. A web application was also developed, which, when connected to the databases, allows us to have selective access to the stored information.

The system makes it possible to homogenise the data recorded in different plants, which, as mentioned above, have different specifications, recorded parameters, nomenclature, units of measurement, or format of the recorded data. This makes it possible to treat data from different types of plants together and to perform the same type of analysis in order to compare results.

The system was developed with open-source software tools, which avoids the payment of software licenses, making the system much more affordable.

Although the manufacturers of equipment for PV installations have applications to visualise recorded parameters, the system developed in this work allows access to the data in a much more broadly and aggregated form, in order to perform more comprehensive and varied analyses than the limitation presented to users by other applications.

As an example, this work includes the analysis, using the developed tools, of some of the parameters recorded in two PV plants during a period of 20 months. The system allows access to the temporal variation of each parameter in order to obtain information on deviations that allow maintenance actions to be improved, to avoid reaching the point of corrective maintenance. For example, a mechanism is presented that allows us to make a daily comparison of the power generated by the different inverters of the same plant such that a deviation in the production of one or more inverters can be detected. Colour maps with the Pearson correlation indexes corresponding to the linear dependence relationships of the production of all the inverters of the plant, easily generated with the use of Python, allow us to visualise these possible deviations.

As a further example, in the analysis of the grid voltage values measured by the inverters of the plants, the increase in grid voltage that occurs due to the production and injection into the grid of the power generated in the plants was easily seen.

The developed system is flexible and scalable, such that the type and number of PV plants analysed with it can be increased.

Although the system does not work in real-time, it allows the data monitored in the plants to be uploaded daily, so that production can be analysed in the shortest possible time in order to detect possible deviations from correct operation as soon as possible.

There are several aspects in which this line of work could be continued. On the one hand, although the user interface of the application has been designed to be easy to use, it is always possible to improve it in order to make it more attractive to the user. Also, the system has been optimised compared to previous versions in order to make it more flexible and optimal so that the loading time of the data is lower, but this is always an aspect that can be improved, especially for loading data from PV plants with a large number of

inverters. On the other hand, the system could also be adapted to new formats of monitored data, as the current version includes CSV and XML file formats, or to include data from additional measuring sensors that could be installed in the plants.

Author Contributions: Conceptualization, D.T.-M. and I.S., Methodology, D.T.-M., J.J.L.-R, I.S., Software and Data Curation, D.T.-M., S.C.-L., M.G.-R., I.S., Writing—Original Draft Preparation and Writing—Review and Editing, D.T.-M., M.G.-R., J.J.L.-R, I.S., Supervision, I.S., Funding acquisition, M.G.-R. and I.S. 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.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the company Solar del Valle SL, which, via its collaboration agreements with our Research Group, provided data monitored in PV installations.

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

References

1. Department of Economic and Social Services. Sustainable Development. United Nations The Sustainable Goals. Available online: <https://sdgs.un.org/es/goals> (accessed on 15 November 2022).
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 November 2022)).
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 November 2022).
4. UNEF (Spanish Photovoltaic Union). Anual Report UNEF 2022, Photovoltaic Solar Energy. Oportunity for sustainability | UNEF. Available online: <https://www.unef.es/es/recursos-informes?idMultimediaCategoria=18> (accessed on 20 October 2022).
5. UNEF (Spanish Photovoltaic Union). Anual Report UNEF 2022, Solar Energy. A Sure Bet for Economic Recovery | UNEF. Available online: <https://www.unef.es/es/recursos-informes?idMultimediaCategoria=18> (accessed on 20 October 2022).
6. The International Energy Agency (IEA). *Solar Photovoltaic: More Efforts Need*; The International Energy Agency (IEA): Paris, France, 2022.
7. The European Solar Initiative. Available online: <https://europeansolarinitiative.eu/> (accessed on 4 November 2022).
8. Solar Power Europe. Available online: <https://www.solarpowereurope.org/> (accessed on 4 November 2022).
9. PVinsights, 2020. Solar PV Module Weekly Spot Price. Available online: <http://pvinsights.com/> (accessed on 20 October 2022).).
10. Schmela, M. *Global Market Outlook. For Solar Power 2018–2022*; Solar Industry Reports; SolarPower Europe: Brussels, Belgium, 2018.
11. Schmela, M. *Global Market Outlook. For Solar Power 2019–2032*; Solar Industry Reports; SolarPower Europe: Brussels, Belgium, 2019.
12. Muellejans, E.; Zaiman, H.; Galleano, W.; Pavanello, R.; Salis, D.; Sample, E.; Dunlop, T.; Bardizza, G.; Garcia, J.K.R.L.; Shaw, D.; et al. *State-of-the-Art for Assessment of Solar Energy Technologies 2019 2020*; JRC Technical Reports; European Commission: Rome, Italy, 2018; pp. 1–57.
13. Zafoschnig, L.A.; Nold, S.; Goldschmidt, J.C. The Race for Lowest Costs of Electricity Production: Techno-Economic Analysis of Silicon, Perovskite and Tandem Solar Cells. *IEEE J. Photovolt.* **2020**, *10*, 1632–1641. <https://doi.org/10.1109/JPHOTOV.2020.3024739>.
14. Ahmad, L.; Khordehghah, N.; Malinauskaite, J.; Jouhara, H. Recent advances and applications of solar photovoltaics and thermal technologies. *Energy* **2020**, *207*, 1–12. <https://doi.org/10.1016/j.energy.2020.118254>.
15. Alkhalayfeh, M.A.; Aziz, A.A.; Pakhuruddin, M.Z.; Katubi, K.M.M. Recent Advances of Perovskite Solar Cells Embedded with Plasmonic Nanoparticles. *Phys. Status Solidi* **2021**, *218*, 2100310. <https://doi.org/10.1002/pssa.202100310>.
16. Buerhop, C.; Stroyuk, O.; Pickel, T.; Hauch, J. Impact of the Module Backsheet Components on the Electrical Field Performance of PV-Plants. In Proceedings of the 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), Calgary, AB, Canada, 15 June–21 August 2020; pp. 320–323.
17. Raza, K.; Akhtar, S.S.; Arif, A.F.M.; Hakeem, A.S. Computational design and development of high-performance polymer-composites as new encapsulant material for concentrated PV modules. *Sci. Rep.* **2020**, *10*, 5304. <https://doi.org/10.1038/s41598-020-62191-9>.

18. Ahmad, A.; Navarro, H.; Ghosh, S.; Ding, Y.; Roy, J.N. Evaluation of New PCM/PV Configurations for Electrical Energy Efficiency Improvement through Thermal Management of PV Systems. *Energies* **2021**, *14*, 4130. <https://doi.org/10.3390/en14144130>.
19. Aghaei, M.; Fairbrother, A.; Gok, A.; Ahmad, S.; Kazim, S.; Lobato, K.; Oreski, G.; Reinders, A.; Schmitz, J.; Theelen, M.; et al. Review of degradation and failure phenomena in photovoltaic modules. *Renew. Sustain. Energy Rev.* **2022**, *159*, 112160. <https://doi.org/10.1016/J.RSER.2022.112160>.
20. Windarko, N.A.; Nizar Habibi, M.; Sumantri, B.; Prasetyono, E.; Efendi, M.Z. A New MPPT Algorithm for Photovoltaic Power Generation under Uniform and Partial Shading Conditions. *Energies* **2021**, *14*, 483. <https://doi.org/10.3390/en14020483>.
21. Debnath, A.; Olowu, T.O.; Parvez, I.; Sarwat, A. Step-size Optimization of New Straight Line Approximation-Based MPPT Algorithm for Photovoltaic Systems. In Proceedings of the 2021 IEEE 12th International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Chicago, IL, USA, 28 June–1 July 2021; pp. 1–6.
22. Nivedha, S., [Vijayalaxmi, M.](#) Performance Analysis of Fuzzy based Hybrid MPPT Algorithm for Photovoltaic System. In Proceedings of the 2021 International Conference on Communication, Control and Information Sciences (ICCIsc), Idukki, India, 16–18 June 2021; Volume 1, pp. 1–4.
23. Kumar, M.; Kumar, A. Performance assessment and degradation analysis of solar photovoltaic technologies: A review. *Renew. Sustain. Energy Rev.* **2017**, *78*, 554–587. <https://doi.org/10.1016/J.RSER.2017.04.083>.
24. dos Santos, S.A.A.; João, J.P.; Carlos, C.A.; Lameirinhas, R.A.M. The impact of aging of solar cells on the performance of photovoltaic panels. *Energy Convers. Manag. X* **2021**, *10*, 100082. <https://doi.org/10.1016/J.ECMX.2021.100082>.
25. Dhimish, M.; Tyrrell, A.M. Power loss and hotspot analysis for photovoltaic modules affected by potential induced degradation. *npj Mater. Degrad.* **2022**, *6*, 11. <https://doi.org/10.1038/s41529-022-00221-9>.
26. Guo, J.; Liu, X.; Yu, J.; Xu, C.; Wu, Y.; Pan, D.; Senthil, R.A. An overview of the comprehensive utilization of silicon-based solid waste related to PV industry. *Resour. Conserv. Recycl.* **2021**, *169*, 105450. <https://doi.org/10.1016/j.resconrec.2021.105450>.
27. Hernandez, G.O.; Ariolli, D.M.G.; Paez, P.S.E.; Chiantore, P.V. Trends and innovations in photovoltaic operations and maintenance. *Prog. Energy* **2022**, *4*, 042002. <https://doi.org/10.1088/2516-1083/AC7C4F>.
28. Status of PV Module Recycling in Selected IEA PVPS Task12 Countries—IEA-PVPS. Available online: <https://iea-pvps.org/key-topics/status-of-pv-module-recycling-in-selected-iea-pvps-task12-countries/> (accessed on 27 October 2022).
29. Preliminary Environmental & Financial Viability Analysis of Circular Economy Scenarios for Satisfying PV System Service Lifetime—IEA-PVPS. Available online: <https://iea-pvps.org/key-topics/preliminary-environmental-financial-viability-analysis-of-circular-economy-scenarios-for-satisfying-pv-system-service-lifetime/> (accessed on 27 October 2022).
30. Oprea, S.-V.; Băra, A. Ultra-short-term forecasting for photovoltaic power plants and real-time key performance indicators analysis with big data solutions. Two case studies—PV Agigea and PV Giurgiu located in Romania. *Comput. Ind.* **2020**, *120*, 103230. <https://doi.org/10.1016/j.compind.2020.103230>.
31. de Arquer Fernández, P.; Fernández, M.Á.F.; Candás, J.L.C.; Arbolea, P.A. An IoT open source platform for photovoltaic plants supervision. *Int. J. Electr. Power Energy Syst.* **2021**, *125*, 106540. <https://doi.org/10.1016/J.IJEPES.2020.106540>.
32. Rediske, G.; Michels, L.; Cezar Mairesse Siluk, J.; Rigo, P.D.; Rosa, C.B.; Bortolini, R.J.F. Management of operation and maintenance practices in photovoltaic plants: Key performance indicators. *Int. J. Energy Res.* **2022**, *46*, 7118–7136. <https://doi.org/10.1002/ER.7737>.
33. Livera, A.; Theristis, M.; Micheli, L.; Fernandez, E.F.; Stein, J.S.; Georghiou, G.E. Operation and Maintenance Decision Support System for Photovoltaic Systems. *IEEE Access* **2022**, *10*, 42481–42496. <https://doi.org/10.1109/ACCESS.2022.3168140>.
34. Ge, L.; Li, Y.; Li, Y.; Yan, J.; Sun, Y. Smart Distribution Network Situation Awareness for High-Quality Operation and Maintenance: A Brief Review. *Energies* **2022**, *15*, 828. <https://doi.org/10.3390/EN15030828>.
35. Hojabri, M.; Kellerhals, S.; Upadhyay, G.; Bowler, B. IoT-Based PV Array Fault Detection and Classification Using Embedded Supervised Learning Methods. *Energies* **2022**, *15*, 2097. <https://doi.org/10.3390/EN15062097>.
36. Rahman, M.M.; Khan, I.; Alameh, K. Potential measurement techniques for photovoltaic module failure diagnosis: A review. *Renew. Sustain. Energy Rev.* **2021**, *151*, 111532. <https://doi.org/10.1016/J.RSER.2021.111532>.
37. Navid, Q.; Hassan, A.; Fardoun, A.A.; Ramzan, R.; Alraeesi, A. sustainability Fault Diagnostic Methodologies for Utility-Scale Photovoltaic Power Plants: A State of the Art Review. *Sustainability* **2021**, *13*, 1629. <https://doi.org/10.3390/su13041629>.
38. The Use of Advanced Algorithms in PV Failure Monitoring—IEA-PVPS. Available online: <https://iea-pvps.org/key-topics/the-use-of-advanced-algorithms-in-pv-failure-monitoring/> (accessed on 27 October 2022).
39. Livera, A.; Paphitis, G.; Montes-Romero, J.; Pikolos, L.; Papadopoulos, I.; Montes-Romero, J.; Lopez-Lorente, J.; Makrides, G.; Sutterlueti, J.; Georghiou, G.E. Intelligent Cloud-Based Monitoring and Control Digital Twin for Photovoltaic Power Plants. In Proceedings of the 2022 IEEE 49th Photovoltaics Specialists Conference (PVSC), Philadelphia, PA, USA, 5–10 June 2022; pp. 1–8.
40. Liao, K.C.; Wu, H.Y.; Wen, H.T. Using Drones for Thermal Imaging Photography and Building 3D Images to Analyze the Defects of Solar Modules. *Inventions* **2022**, *7*, 67. <https://doi.org/10.3390/INVENTIONS7030067>.
41. Sornek, K.; Filipowicz, M.; Vergura, S. Criticalities of the Outdoor Infrared Inspection of Photovoltaic Modules by Means of Drones. *Energies* **2022**, *15*, 5086. <https://doi.org/10.3390/EN15145086>.
42. Fares, E.; Aissa, B.; Isafan, R.J. Inkjet printing of metal oxide coatings for enhanced photovoltaic soiling environmental applications. *Glob. J. Environ. Sci. Manag.* **2022**, *8*, 485–502. <https://doi.org/10.22034/GJESM.2022.04.03>.

43. Murtadha, T.K.; Dil Hussein, A.A.; Alalwany, A.A.H.; Alrwashdeh, S.S.; Al-Falahat, A.M. Improving the cooling performance of photovoltaic panels by using two passes circulation of titanium dioxide nanofluid. *Case Stud. Therm. Eng.* **2022**, *36*, 102191. <https://doi.org/10.1016/J.CSITE.2022.102191>.
44. Nabil, T.; Mansour, T.M. Augmenting the performance of photovoltaic panel by decreasing its temperature using various cooling techniques. *Results Eng.* **2022**, *15*, 100564. <https://doi.org/10.1016/J.RINENG.2022.100564>.
45. Siah Chehreh Ghadikolaie, S. Solar photovoltaic cells performance improvement by cooling technology: An overall review. *Int. J. Hydrogen Energy* **2021**, *46*, 10939–10972. <https://doi.org/10.1016/J.IJHYDENE.2020.12.164>.
46. Zhao, Y.; Gong, S.; Zhang, C.; Ge, M.; Xie, L. Performance analysis of a solar photovoltaic power generation system with spray cooling. *Case Stud. Therm. Eng.* **2022**, *29*, 101723. <https://doi.org/10.1016/J.CSITE.2021.101723>.
47. Udayakumar, M.D.; Anushree, G.; Sathyaraj, J.; Manjunathan, A. The impact of advanced technological developments on solar PV value chain. *Mater. Today Proc.* **2021**, *45*, 2053–2058. <https://doi.org/10.1016/j.matpr.2020.09.588>.
48. de Freitas Viscondi, G.; Alves-Souza, S.N. A Systematic Literature Review on big data for solar photovoltaic electricity generation forecasting. *Sustain. Energy Technol. Assess.* **2019**, *31*, 54–63. <https://doi.org/10.1016/J.SETA.2018.11.008>.
49. Hasan, K.; Yousuf, S.B.; Tushar, M.S.H.K.; Das, B.K.; Das, P.; Islam, M.S. Effects of different environmental and operational factors on the PV performance: A comprehensive review. *Energy Sci. Eng.* **2022**, *10*, 656–675. <https://doi.org/10.1002/ESE3.1043>.
50. Roumpakias, E.; Stamatelos, A. Performance analysis of a grid-connected photovoltaic park after 6 years of operation. *Renew. Energy* **2019**, *141*, 368–378. <https://doi.org/10.1016/j.renene.2019.04.014>.
51. Murat Ates, A.; Singh, H. Rooftop solar Photovoltaic (PV) plant—One year measured performance and simulations. *J. King Saud Univ. Sci.* **2021**, *33*, 101361. <https://doi.org/10.1016/J.JKSUS.2021.101361>.
52. Gopi, A.; Sudhakar, K.; Ngui, W.K.; Kirpichnikova, I.M.; Cuce, E. Energy analysis of utility-scale PV plant in the rain-dominated tropical monsoon climates. *Case Stud. Therm. Eng.* **2021**, *26*, 101123. <https://doi.org/10.1016/J.CSITE.2021.101123>.
53. Bansal, N.; Pany, P.; Singh, G. Visual degradation and performance evaluation of utility scale solar photovoltaic power plant in hot and dry climate in western India. *Case Stud. Therm. Eng.* **2021**, *26*, 101010. <https://doi.org/10.1016/J.CSITE.2021.101010>.
54. Ihaddadene, R.; El hassen Jed, M.; Ihaddadene, N.; De Souza, A. Analytical assessment of Ain Skhouna PV plant performance connected to the grid under a semi-arid climate in Algeria. *Sol. Energy* **2022**, *232*, 52–62. <https://doi.org/10.1016/J.SOLENER.2021.12.055>.
55. Daher, D.H.; Gaillard, L.; Ménéz, C. Experimental assessment of long-term performance degradation for a PV power plant operating in a desert maritime climate. *Renew. Energy* **2022**, *187*, 44–55. <https://doi.org/10.1016/J.RENENE.2022.01.056>.
56. Kumar, M.; Chandel, S.S.; Kumar, A. Performance analysis of a 10 MWp utility scale grid-connected canal-top photovoltaic power plant under Indian climatic conditions. *Energy* **2020**, *204*, 117903. <https://doi.org/10.1016/j.energy.2020.117903>.
57. Srivastava, R.; Tiwari, A.N.; Giri, V.K. An overview on performance of PV plants commissioned at different places in the world. *Energy Sustain. Dev.* **2020**, *54*, 51–59. <https://doi.org/10.1016/j.esd.2019.10.004>.
58. Adar, M.; Najih, Y.; Gouskir, M.; Chebak, A.; Mabrouki, M.; Bennouna, A. Three PV plants performance analysis using the principal component analysis method. *Energy* **2020**, *207*, 118315. <https://doi.org/10.1016/j.energy.2020.118315>.
59. AL-Rasheedi, M.; Gueymard, C.A.; Al-Khayat, M.; Ismail, A.; Lee, J.A.; Al-Duaj, H. Performance evaluation of a utility-scale dual-technology photovoltaic power plant at the Shagaya Renewable Energy Park in Kuwait. *Renew. Sustain. Energy Rev.* **2020**, *133*, 110139. <https://doi.org/10.1016/j.rser.2020.110139>.
60. IEC 61724-1:2021; Photovoltaic System Performance—Part 1: Monitoring. International Electronic Comision IEC: Geneva, Switzerland, 2021. Available online: <https://webstore.iec.ch/publication/65561> (accessed on 20 October 2022).
61. Communication and Control for High PV Penetration under Smart Grid Environment—IEA-PVPS. Available online: <https://iea-pvps.org/key-topics/communication-and-control-for-high-pv-penetration-under-smart-grid-environment/> (accessed on 27 October 2022).
62. Roumpakias, E.; Stamatelos, T. Prediction of a Grid-Connected Photovoltaic Park’s Output with Artificial Neural Networks Trained by Actual Performance Data. *Appl. Sci.* **2022**, *12*, 6458. <https://doi.org/10.3390/APP12136458>.
63. Data Model and Data Acquisition for PV Registration Schemes and Grid Connection—IEA-PVPS. Available online: <https://iea-pvps.org/key-topics/data-model-and-data-acquisition-for-pv-registration-schemes-and-grid-connection-best-practice-and-recommendations/> (accessed on 27 October 2022).
64. Torres, J.F.; Troncoso, A.; Koprinska, I.; Wang, Z.; Martínez-Álvarez, F. Big data solar power forecasting based on deep learning and multiple data sources. *Expert Syst.* **2019**, *36*, e12394. <https://doi.org/10.1111/exsy.12394>.
65. Cordeiro-Costas, M.; Villanueva, D.; Eguía-Oller, P.; Granada-Álvarez, E. Machine Learning and Deep Learning Models Applied to Photovoltaic Production Forecasting. *Appl. Sci.* **2022**, *12*, 8769. <https://doi.org/10.3390/APP12178769>.
66. Aleem, S.A.; Suhail Hussain, S.M.; Ustun, T.S. A Review of Strategies to Increase PV Penetration Level in Smart Grids. *Energies* **2020**, *13*, 636. <https://doi.org/10.3390/EN13030636>.
67. Bacha, M.; Terki, A. Diagnosis algorithm and detection faults based on fuzzy logic for PV panel. *Mater. Today Proc.* **2022**, *51*, 2131–2138. <https://doi.org/10.1016/J.MATPR.2021.12.490>.
68. EN 50160: 2011; Voltage Characteristics of Electricity Supplied by Public Electricity Networks. AENOR: Madrid, Spain, 2011.
69. Santiago, I.; García-Quintero, J.; Mengibar-Ariza, G.; Trillo-Montero, D.; Real-Calvo, R.J.; Gonzalez-Redondo, M. Analysis of Some Power Quality Parameters at the Points of Common Coupling of Photovoltaic Plants Based on Data Measured by Inverters. *Appl. Sci.* **2022**, *12*, 1138. <https://doi.org/10.3390/APP12031138>.

70. Mahato, G.C.; Choudhury, T.R.; Nayak, B.; Debnath, D.; Santra, S.B.; Misra, B. A Review on High PV Penetration on Smart Grid: Challenges and its Mitigation Using FPPT. In Proceedings of the 2021 1st International Conference on Power Electronics and Energy (ICPEE), Bhubaneswar, India, 2–3 January 2021. <https://doi.org/10.1109/ICPEE50452.2021.9358474>.
71. Ikić, M.; Mikulović, J. Experimental Evaluation of Distortion Effect for Grid-Connected PV Systems with Reference to Different Types of Electric Power Quantities. *Energies* **2022**, *15*, 416. <https://doi.org/10.3390/EN15020416>.
72. Rahman, S.; Saha, S.; Islam, S.N.; Arif, M.T.; Mosadeghy, M.; Haque, M.E.; Oo, A.M.T. Analysis of Power Grid Voltage Stability with High Penetration of Solar PV Systems. *IEEE Trans. Ind. Appl.* **2021**, *57*, 2245–2257. <https://doi.org/10.1109/TIA.2021.3066326>.
73. Kalay, M.Ş.; Kılıç, B.; Sağlam, Ş. Systematic review of the data acquisition and monitoring systems of photovoltaic panels and arrays. *Sol. Energy* **2022**, *244*, 47–64. <https://doi.org/10.1016/J.SOLENER.2022.08.029>.
74. Moreno-Garcia, I.M.; Palacios-Garcia, E.J.; Pallares-Lopez, V.; Santiago, I.; Gonzalez-Redondo, M.J.; Varo-Martinez, M.; Real-Calvo, R.J. Real-Time Monitoring System for a Utility-Scale Photovoltaic Power Plant. *Sensors* **2016**, *16*, 770. <https://doi.org/10.3390/S16060770>.
75. Rahman, M.M.; Selvaraj, J.; Rahim, N.A.; Hasanuzzaman, M. Global modern monitoring systems for PV based power generation: A review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 4142–4158. <https://doi.org/10.1016/J.RSER.2017.10.111>.
76. Dabou, R.; Bouraiou, A.; Ziane, A.; Necaibia, A.; Sahouane, N.; Blal, M.; Khelifi, S.; Rouabhia, A.; Slimani, A. Development of autonomous monitoring and performance evaluation system of grid-tied photovoltaic station. *Int. J. Hydrogen Energy* **2021**, *46*, 30267–30287. <https://doi.org/10.1016/J.IJHYDENE.2021.06.204>.
77. Kabalci, E.; Kabalci, Y. A wireless metering and monitoring system for solar string inverters. *Int. J. Electr. Power Energy Syst.* **2018**, *96*, 282–295. <https://doi.org/10.1016/J.IJEPES.2017.10.013>.
78. Madeti, S.R.; Singh, S.N. Monitoring system for photovoltaic plants: A review. *Renew. Sustain. Energy Rev.* **2017**, *67*, 1180–1207. <https://doi.org/10.1016/J.RSER.2016.09.088>.
79. Santiago, I.; Montero, D.T.; Rodríguez, J.J.L.; Garcia, I.M.M.; Garcia, E.J.P. Graphical diagnosis of performances in photovoltaic systems: A case study in southern Spain. *Energies* **2017**, *10*, 1964. <https://doi.org/10.3390/en10121964>.
80. 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. <https://doi.org/10.1016/J.ENCONMAN.2014.02.026>.
81. Zhou, K.; Fu, C.; Yang, S. Big data driven smart energy management: From big data to big insights. *Renew. Sustain. Energy Rev.* **2016**, *56*, 215–225. <https://doi.org/10.1016/J.RSER.2015.11.050>.
82. Hu, J.; Vasilakos, A.V. Energy Big Data Analytics and Security: Challenges and Opportunities. *IEEE Trans. Smart Grid* **2016**, *7*, 2423–2436. <https://doi.org/10.1109/TSG.2016.2563461>.
83. Ceci, M.; Corizzo, R.; Fumarola, F.; Ianni, M.; Malerba, D.; Maria, G.; Masciari, E.; Oliverio, M.; Rashkovska, A. Big Data Techniques for Supporting Accurate Predictions of Energy Production from Renewable Sources. In Proceedings of the IDEAS '15: Proceedings of the 19th International Database Engineering & Applications Symposium, Yokohama, Japan, 13–15 July 2015; pp. 62–71. <https://doi.org/10.1145/2790755.2790762>.
84. Bhuiyan, E.A.; Hossain, M.Z.; Mueyen, S.M.; Fahim, S.R.; Sarker, S.K.; Das, S.K. Towards next generation virtual power plant: Technology review and frameworks. *Renew. Sustain. Energy Rev.* **2021**, *150*, 111358. <https://doi.org/10.1016/J.RSER.2021.111358>.
85. Hannan, M.A.; Abdolrasol, M.G.; Mohamed, R.; Al-Shetwi, A.; Ker, P.; Begum, R.; Muttaqi, K. ANN-Based Binary Backtracking Search Algorithm for VPP Optimal Scheduling and Cost-Effective Evaluation. *IEEE Trans. Ind. Appl.* **2021**, *57*, 5603–5613. <https://doi.org/10.1109/TIA.2021.3100321>.
86. Popławski, T.; Dudzik, S.; Szlag, P.; Baran, J. A Case Study of a Virtual Power Plant (VPP) as a Data Acquisition Tool for PV Energy Forecasting. *Energies* **2021**, *14*, 6200. <https://doi.org/10.3390/EN14196200>.
87. MariaDB. Available online: <https://mariadb.org/> (accessed on 24 December 2022).
88. MySQL Workbench. Available online: <https://www.mysql.com/products/workbench/> (accessed on 24 December 2022).
89. HeidiSQL. Available online: <https://www.heidisql.com/> (accessed on 24 December 2022).
90. Vuejs. Available online: <https://vuejs.org/> (accessed on 24 December 2022).
91. 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. <https://doi.org/10.1038/s41586-020-2649-2>.
92. McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010*; van der Walt, S., Millman, J., Eds.; Publisher: Austin, Texas, USA 2010; pp. 56–61.
93. Waskom, M.L. seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. <https://doi.org/10.21105/joss.03021>.
94. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
95. 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. <https://doi.org/10.1038/s41592-019-0686-2>.

-
96. 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>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.