

# Diagnosis of failures in Solar Plants based on Performance monitoring

Ana P. Talayero, Andrés Llombart, Julio J. Melero

Instituto Universitario de Investigación Mixto CIRCE (Fundación CIRCE – Universidad de Zaragoza), Zaragoza Spain  
Phone +0034 976 97 68 59 talayero@fcirce.es, llombart@fcirce.es, melero@unizar.es

## Abstract

Photovoltaic (PV) solar energy has become a reference in electrical generation. The plants currently installed, and those planned have a huge capacity and occupy large areas.

The increase in size of the plants presents new challenges in operation and maintenance areas, such as the optimization of the number of sensors installed, large data management and the reduction of the timework in maintenance.

The aim of this paper is to show a methodology, to diagnose failures, based on the measured data in the plant. The methodology used is supervised regression machine learning and comparison algorithms. This methodology allows the study of the sensors, the inverters, the joint boxes and the power reduction caused by soiling. The result would allow the detection of around 1-5% of production loss in the plant. The algorithms have been tested with real data of PV plants, and have detected common failures such as production drops in strings and losses due to soiling.

**Keywords:** PV Solar plants, Efficiency, failures and Diagnosis

## Introduction

Photovoltaic solar energy is growing by over 20% worldwide yearly [1]. The decrease in the price of the modules and the improvements in the technology, increasing the performance by up to 85% on average, make it competitive in the electricity generation sector [2]. These facts together with the policy incentives, at European [3] and national [4] levels, pushed the installation of new plants.

These new plants exceed the average power of the existing ones by 10 times, generating hundreds of megawatts and occupying thousands of hectares of extension. There is no previous experience, nor studies in these huge photovoltaic (PV) plants, where millions of modules are laid and connected in thousands of joint boxes that end up in power stations with inverters greater than a megawatt.

The monitoring of all modules and equipment of the plant is very costly as a consequence of the number of sensors needed and the cost of processing such an amount of information. So, one challenge in this kind of plants is finding the minimum number of sensors to control the plant and the most efficient way of working with millions of daily generated data.

The increase in the size of the plants requires the optimization of the operation and maintenance, which implies new challenges and trends for this task [2], such as minimising the timework in the maintenance tasks. It is mandatory to quickly identify the location of the faults in the plant in order to optimise the reparation time and its impact on production.

Machine Learning techniques (ML) combined with Big-Data methodologies are suitable to work with large time series of data streamed by reasonable number of sensors to model the plant. They can be used to solve the problem of the high number of sensors and to process the generated data.

Diagnosis is the way of determining the status and performance of the PV plant aiming at detecting failures in the plant components. Besides, it is not only useful in the maintenance tasks and to enhance efficiency, but also to avoid financial risks [5], and to decrease Operational Expenditures (OPEX) [6] and the Levelized Cost of Energy (LCOE) of the PV plant [7].

The diagnosis models using ML allow the location of the faults, and the control of the efficiency of the components. This helps to find the most appropriate way to address the failures, minimizing the number of man hours in maintenance, increasing availability of the plant and, therefore, the production.

In this study, a working methodology based on ML techniques, is presented. PV plants are modelled using several algorithms. This allows a quick identification of the main failures, such as sensor errors, string and inverter faults and performance drops.

In the following sections, a review of the faults appearing in the PV plants is carried out in order to assess those which are more convenient to work on, and to identify diagnostic models, adapted to these large plants. Finally, the diagnosis methodology is described, and two successful cases are presented.

## Background

Scientific literature shows a large number of papers describing and analysing failures in the existing PV plants. Most of the studies are focused on the direct

current side of the plant [8,9], specifically in the PV modules.

The study of the failures found along the life of the plant [10,11] reveal that different failures can be found depending on the operation time of the plant, start-up, mid-life or close to the final stages of the life of the plant.

During the first years of a plant, the failures are derived from manufacturing errors of the chosen material which determines its premature deterioration [5], or even from the production process failures.

In the midlife of the plant, the weather can affect the welds by thermomechanical fatigue [12], by oxidation, or by the appearance of "snailtrails" [13]. Other problems found are delamination, discoloration [9] and bubble formation [5].

As the plant ages, the most common problem studied is degradation. The studies related to the degradation analyse are based on the effect of the humidity, ultraviolet rays or temperature [5,9,14]. Other studies analyse the influence of the material used in the module regarding degradation [15].

One of the problems found throughout the whole life of the plant is the soiling in the PV modules. This phenomenon is mainly produced by the accumulation of dust [10,16] on the surface of the modules. Other biological remains as bird droppings, leaves, tree resin, adhered mosses, etc., also have an impact [5]. The effect of soiling decreases production and can also cause oxidation and a reduction of the lifetime in the modules.

Regarding the alternating current side of the plant, inverters are the most studied components, although they have a lower failure rate than the PV modules. Their errors are mainly related to high temperatures, lightning strikes and filter breakage [5]. The lifetime of the inverters is of around 10 years [10,17] which compared to the 30 years that the PV modules last, means that they reach the end of their useful life more than once during the operation of the plant.

The sensors used to control the performance of the plants, specially the pyranometers, also present failures. These failures must be detected because their measurements are very important for the proper operation of the plant [18,19].

The conclusions of the failure analysis are that the two more critical elements in PV plants are the PV modules and the inverters. Then, monitoring of the performance of those critical elements, control of the state of the sensors and continuous evaluation of the degradation and soiling are key for maintaining the plant in a good health state.

There are many approaches for the diagnostic and fault detection models that can be applied to PV plants [20]. Between them, those using SCADA data and plant performance to detect deviations from well-established thresholds are very interesting as they only use the Plant available information.

The diagnostic models, based on performance or efficiency, estimate the energy production and losses of the plant using machine learning algorithms [21,22]. These estimated values are then compared with measured values [23,24] and the deviations associated to faults in the plant.

The diagnostic models based on performance used to work with the modules control their efficiency. The efficiency is related to its influencing factors like shadows, soiling, temperature, degradation, faults, tracking system... [25,26].

Finally, the models focused on the inverters, relate the measurements, mainly power and currents, in the direct current side with those measured in alternating current side [27,28].

Working only with the data from the large inverter is not enough to locate failures in modules. In order to not deploy a large number of sensors, the solution would be working at string level. For this, it is necessary to obtain measurements in the junction box and then develop a methodology to work with all this information.

In this paper, simple framework with four objectives is proposed: identifying areas with possible failures in the modules, detect inverters whose operation is not adequate, discover sensor failures and reveal production drops. The result of the algorithms will allow an efficient planning in maintenance tasks.

## Methodology

The developed framework works with the data registered in the SCADA of the plant, such as the climatologic variables and electrical measurements. The available measurements in the reference plant, used for developing the models, are in Table 1. These meteorological variables have been measured in a met mast equipped with anemometer, wind vane, pyranometer, barometer, thermometer and humidity sensor, and situated in the middle of the plant. The temperature of the modules has been measured in the closest module to the met station, considered as reference module, and the electrical signals have been measured in the joint boxes and the inverters. A training period of one year of has been established for the determination of the parameters of the algorithms.

**Table1.** List of variables

Variable	
• Wind speed	• Input join box current
• Wind direction	• Output join box current
• Radiation	• Inverter current DC
• Pressure	• Inverter voltage DC
• Temperature	• Inverter current AC
• Humidity	• Inverter voltageAC
• Module Temperature	• Inverter Temperature
• DC Power	• AC Power

The methodology has been developed in the “R” programming language [29] and is organised in groups of algorithms according to its functionality, as can be seen in Figure 1. All algorithms developed can be run on a standard computer, as for example, in the same computer where the SCADA system has been installed.

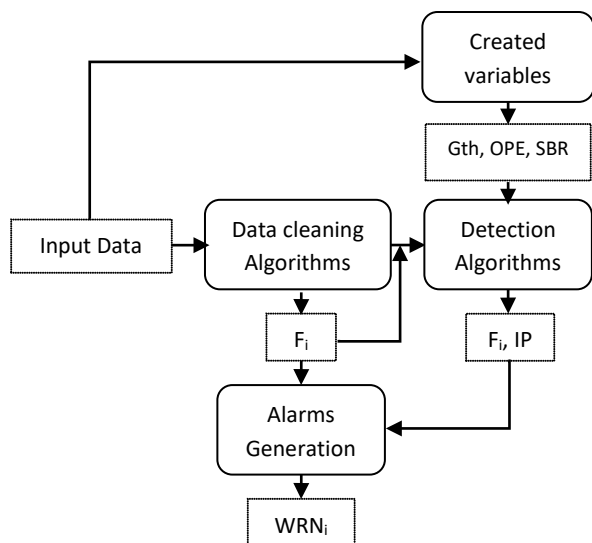


Figure 1. Framework description

The framework is divided in two main groups of algorithms. The first one, named as Data cleaning algorithm, is used to clean, organise and adequate the input data. The second group, named as detection algorithms, work with the input data in order to find malfunctions in the system and anomalies in the data. These algorithms will be explained later.

The results of these algorithms are binary variables indicating the anomalies. The combination, frequency and duration of these variables generate warnings that will be used to detect failures in the plant.

In addition to the input data, three more variables are calculated with a specific group of algorithms: "theoretical radiation" (Gth), "operation" (OPE) indicating if the plant is capable to operate, and "shadows" (SBR) that indicates the presence of shadows.

The theoretical radiation (Gth) is calculated following the clear sky model, used in the European Solar Radiation Atlas (ESRA) [30,31], corrected for inclined surfaces [32] and improved with the Linke Turbidity Factor, that accounts for the presence of aerosols and water vapour in the atmosphere [33]. This variable is used to verify the radiation measurement in the first group of algorithms and also to estimate the production of each inverter in the detection algorithm.

The binary variable OPE relates the radiation, the power, the module temperature and the ambient temperature taking the value 1 when it is in operation state. Finally, the SBR variable is determined from the hour and month with shadows and takes the value 1 when there is no shadow. These two variables are used to discriminate anomalies in the detection algorithm.

The Data cleaning Algorithms guarantee the quality of the information and detect anomalies in the measured input data. These algorithms work with key performance indicators (KPI), which determine the operating range of all the variables considered (see Table 1), allowing the detection and filtering of outliers in the measured data by means of Equation 1.

Persistence and variability of data are evaluated with Equation 2 which allows detecting when there are significant changes.

Finally, comparison of two variables representing the same physical magnitude, as is the case of measured and theoretical values of radiation, is performed using Equation 3, which allows comparing both values.

$$K_{jvar\_i} = \frac{Var_{ji}}{U_{var_j}} \quad \text{Equation1}$$

$$K_{jvar\_i} = \sum_{t=1}^N \left| \frac{Var_{ji} - Var_{ji-t}}{Var_{ji}} \right| \quad \text{Equation2}$$

$$K_{jvar\_i} = \frac{\widehat{Var}_{ji}}{Var_{ji}} = \frac{\sum_{k=1}^n a_k \cdot Var_{ki} + b}{Var_{ji}} \quad \text{Equation3}$$

$K_{jvar\_i}$ : KPI of variable (j) in instant (i)

Var: measured variable (j) in instant (i)

$U_{var_j}$ : reference limit of variable (j). Pre-defined value by technical limitation

N: instant of time

$a_k$  y  $b$ : coefficients

n: number of variables representing the same magnitude

The obtained KPIs for the input data are then compared with known thresholds, defined by the manufacturer of the components, by experience, or using the mean behaviour of the signals.

The result of the cleaning data algorithms is a set of binary flag variables  $F_i$ , linked to each KPI and indicating if the parameter is below the threshold (value “one”) or not (value “zero”). Whenever the threshold is surpassed, an anomaly in the corresponding measurement can be present. These anomalies in the measurements are not used in the next group of algorithms as they can stand for failures in sensors.

The second main group of algorithms detect anomalies in the operation of the plant. As mentioned in the background section, the main losses of production are related with failures in big components, PV modules and inverters, and with production drops. Then, these algorithms are focused on the detection of these events.

The algorithms to verify the inverters are based on Multivariable Linear Regression and use the measured and the calculated variables. The energy at the output of the inverter is estimated with several models based on weather and electrical variables together with other inverter data. Having different estimated results from different sources strengthens the model. The ratios

between the estimated and measured energy are indicators that can be used to compare with a threshold.

In this case the threshold means unknown energy loss, and from a technical point of view, it is recommended to work with values equivalent to 3%-5% of the energy loss. Equation 4.

$$K_{E,i} = \frac{\hat{E}_i}{E_i}$$

$$\hat{E}_i = a \cdot G_i + b \cdot G_{th} + c \cdot T_{amb} + d \cdot T_{mod}, \quad \text{Equation 4}$$

$$\hat{E}_i = \sum_{k=1}^n e_k \cdot E_k$$

$$\hat{E}_i = \sum f \cdot I_i + g \cdot V_i \text{ with OPE, SBR, } F_i = 1$$

$\hat{E}_i$  y  $E_i$  estimated and measured energy

$G_i$  measured radiation

$G_{th}$  theoretical radiation

OPE, SBR, y  $F_i$  binary variables

$T_{amb}$  ambient temperature and  $T_{mod}$  temperature of module

$I_i$  current DC & AC and  $V_i$  voltage DC & AC

An easy and efficient methodology to verify the PV modules is the comparison of the input current measurement of the string junction boxes. These comparisons generate a correlation matrix between all the junction boxes inputs.

The Pearson correlation factor [34], obtained for the currents in the junction boxes, is the indicator used to determinate anomalies or faults in the PV modules, Equation 5.

$$P_{i,j} = \frac{\text{cov}(I_i, I_j)}{\sqrt{\text{var}(I_i) \cdot \text{var}(I_j)}} \quad \text{Equation 5}$$

$I_j$  junction box current

$P_{i,j}$  correlation coefficient

Working with the whole plant allows the comparison between strings, not only with those belonging to the same box where the correlation is normally better but also with the rest of the plant, in order to prevent failures in the detection when there are several strings with failure in the same box.

There are two threshold values to determine the anomalies in a string. One is defined as the average of the Pearson correlation coefficients  $P_{i,j}$ , estimated with the measurements of the training period for the whole plant assumed in good health status. The other threshold is obtained from the median of the average values of the correlation coefficients,  $P_{i,j}$ , of all the strings in each junction box.

When the indicator of a string measurement is below both thresholds, there is an anomaly in the string. If it is over them, it is running properly. In other cases, it may be necessary to analyse the problem in depth taking the result of the inverter algorithms into account.

The results of the detection failures algorithm in inverters and strings are also new  $F_i$  binary variables related with

the indicators, and in this case, they indicate anomalies in the behaviour of these components. When the KPIs defined do not fulfil the threshold, the value of the binary variable associated is "zero".

Production drops can be caused by soiling or degradation, as mentioned in the introduction section. Both decrease the whole plant production continuously and, in the case of soiling, a correct management of the cleaning process results in the reversion of the process and the increasing of the production.

Production drops are complicated to detect quickly enough due to the influence of factors difficult to estimate, such as cloudiness. To address this problem, the gain of the current in the inverters as a function of the radiation has been used.

This value is then constrained to a known range and smoothed, by using the inverse tangent function and taking the moving average in a predefined interval Equation 6.

Using this gain provides independence with the cloudiness and its transformation through the inverse tangent function allows to strength the results reducing the spurious values.

$$K_{d,t} = \frac{\sum_1^n \tan^{-1} \left( \frac{I_i}{G_i} \right)}{n} \quad \text{Equation 6}$$

$I_i$  the value of the input inverter current

$G_i$  the value of the radiation

$n$ : time

The threshold to determine when an intervention is necessary, is defined with a balance between its cost and the production loss cost. This is a decision of the manager of the plant. A decrease of the indicator in 2-3 units is equivalent to 4%-5% performance loss.

The whole set of binary flag variables defined in the data cleaning algorithms and in the failure detection algorithms, represent anomalies in the measurements or in the operational behaviour in the plant and are analysed in frequency and duration, to eliminate spurious data. The combination of activated flag variables with value "zero" determines the warning that indicates the need of the revision of a component or a sensor.

The study, performed component by component, allows detecting lack of production of around 5% in one string or inverter, which mean less than 1% considering the total power of the plant.

In the next section, these algorithms have been checked in two real common failures.

## Studied cases

A significant lack of production in a string and the detection of soiling are the cases selected to verify the framework.

### Faulty string detection

The detection of the lack of production in the strings, in such large plants is complicated since they can be negligible over the whole production of the plant, but these small losses can result in important economic losses along the lifetime of the plant. These string faults, as explained in the methodology section, are detected comparing the production of the different strings.

Figure 2, shows the correlation coefficients between the string productions of only 4 joint boxes of the plant, for each one is showing only 9 input strings of the box. This sample allows it to better detect the lack of production in the strings, and how the box is affected.

Because of the proximity between the strings of the same box, they have higher correlation among them than when they are compared with those of different boxes. This can be identified in Figure 2 by the darker big squares along the diagonal.

It can also be seen how string 4 in box 2 and inverter 1 present a substantial lack of production, showing a lower correlation compared with all the others, and how this anomaly in one string also affects its joint box having a worse correlation between the rest of the boxes.

It is worth to note that faults of production in one string cause less than 1% of production drop regarding the total plant power, which is very difficult to detect without a specific study.

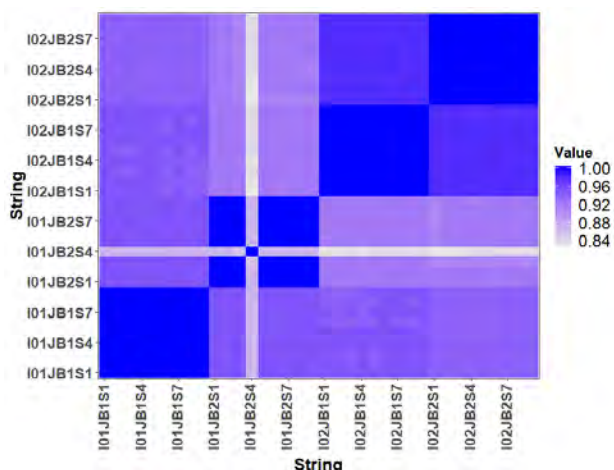


Figure 2. Comparative study of string production

### Soiling detection

The cleaning of the PV modules in big plants is a task that involves several days of work. An excess of cleaning in such large plants implies unnecessary costs, while a scarce cleaning can cause a production loss. The identification of the optimum cleaning time, based on the

measured data, is an important aid to improve the maintenance cost-effectiveness.

As has been described in the methodology to identify the soiling, a specific KPI is defined. An example in Figure 3, shows a continuous decrease of the parameter in one inverter of a plant with close to 100 MW and 3 years of age and how it has recovered after the cleaning. In this specific case, the action taken resulted in a decrease of 5 points (equivalent to 12% performance loss) in the inverter study.



Figure 3. Soiling indicator before and after cleaning

The soiling in this plant is mainly caused by bird droppings so not all inverters are equally affected. The loss of production in relation to the total power of the plant was less than 10% when the cleaning was decided. This decision followed an economical criterion. Nevertheless, the detection of the loss was done when the production had dropped by 5% in the worse inverter and the affection in the plant was less than 3%.

## Conclusions

The increase in the PV plants size makes it necessary to diagnose and locate the faults which influence the availability and efficiency, such as errors in the sensors, production drops, or soiling.

This paper has presented a framework which combines different algorithms based on comparisons and ML techniques working with measurements available in the plant, without the need to install additional instrumentation or use external information. This way the CAPEX of the project does not need to be increased.

The two application cases which have been presented are the detection of a fault in a string and the identification of soiling, showing how the framework, and the proposed algorithms, allow the identification of common failures, impossible to identify optimally using previous know how.

This improves of the profit, and the reduction of the losses caused by the failures being able to detect lacks in production of around 1-5%.

## References

- [1] Anpier. Asociación nacional de productores de energía fotovoltaica, Anuario Fotovoltaico, (2018). <http://proyectotransicionenergetica.org/AnuarioFotovoltaico2018.pdf>.
- [2] A. Woyte, S. Goy, 11 - Large grid-connected photovoltaic power plants: Best practices for the

- design and operation of large photovoltaic power plants, Elsevier Ltd., 2017.  
<https://doi.org/10.1016/B978-1-78242-336-2.00011-2>.
- [3] Diario Oficial de la Union Europea, DIRECTIVA (UE) 2018/2001 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 11 de diciembre de 2018 relativa al fomento del uso de energía procedente de fuentes renovables (versión refundida), D. Of. La Union Eur. 2018 (2018) 128.
- [4] Ministerio para la Transición Ecológica, Borrador del plan nacional integrado de energía y clima 2021-2030, Gob. España. (2019).
- [5] R. Iea-pvps, Assessment of Photovoltaic Module Failures in the Field, 2017.
- [6] R. Iea-pvps, Technical Assumptions Used in PV Financial Models Review of Current Practices and Recommendations, 2017.
- [7] J.-W. Arnulf, PV Status Report 2017 PV Status Report 2017, 2017. <https://doi.org/10.2790/682995>.
- [8] A. Livera, M. Theristis, G. Makrides, G.E. Georghiou, Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems, *Renew. Energy*. 133 (2019) 126–143.  
<https://doi.org/10.1016/j.renene.2018.09.101>.
- [9] D.C. Jordan, T.J. Silverman, J.H. Wohlgemuth, S.R. Kurtz, K.T. VanSant, Photovoltaic failure and degradation modes, *Prog. Photovoltaics Res. Appl.* 25 (2017) 318–326. <https://doi.org/10.1002/pip.2866>.
- [10] M. Köntges, S. Kurtz, C.E. Packard, U. Jahn, K. Berger, K. Kato, T. Friesen, H. Liu, M. Van Iseghem, Review of Failures of Photovoltaic Modules, 2014. <https://doi.org/978-3-906042-16-9>.
- [11] D. DeGraaff, R. Lacerda, Z. Campeau, S. Corp, Degradation mechanisms in Si module technologies observed in the field; their analysis and statistics, NREL 2011 Photovolt. Modul. Reliab. Work. (2011) 1–25. [http://www.irishellas.com/files/Degradation-Mechanisms-in-Si-Module.pdf%0Ahttps://www1.eere.energy.gov/solar/pdfs/pvmrw2011\\_01\\_plen\\_degraaff.pdf](http://www.irishellas.com/files/Degradation-Mechanisms-in-Si-Module.pdf%0Ahttps://www1.eere.energy.gov/solar/pdfs/pvmrw2011_01_plen_degraaff.pdf).
- [12] N. Bosco, T.J. Silverman, S. Kurtz, Climate specific thermomechanical fatigue of flat plate photovoltaic module solder joints, *Microelectron. Reliab.* 62 (2016) 124–129.  
<https://doi.org/10.1016/j.microrel.2016.03.024>.
- [13] T. Takashima, J. Yamaguchi, K. Otani, T. Oozeki, K. Kato, M. Ishida, Experimental studies of fault location in PV module strings, *Sol. Energy Mater. Sol. Cells*. 93 (2009) 1079–1082.  
<https://doi.org/10.1016/j.solmat.2008.11.060>.
- [14] D.C. Jordan, S.R. Kurtz, K. VanSant, J. Newmiller, Compendium of photovoltaic degradation rates, *Prog. Photovoltaics Res. Appl.* 24 (2016) 978–989.  
<https://doi.org/10.1002/pip.2744>.
- [15] D.C. Miller, J. Bath, M. Köhl, T. Shioda, IEC Quality Assurance Task Group 5 : UV , Temperature , and Humidity, 2014.
- [16] G. Corbellini, S. October, Tools for field testing Gianluca Corbellini - SUPSI, (2015).
- [17] NREL, Useful Life, (2019).
- [18] S. Mohanty, P.K. Patra, S.S. Sahoo, Prediction and application of solar radiation with soft computing over traditional and conventional approach - A comprehensive review, *Renew. Sustain. Energy Rev.* 56 (2016) 778–796.  
<https://doi.org/10.1016/j.rser.2015.11.078>.
- [19] L. Menyhart, A. Anda, Z. Nagy, A new method for checking the leveling of pyranometers, *Sol. Energy*. 120 (2015) 25–34.  
<https://doi.org/10.1016/J.SOLENER.2015.06.033>.
- [20] A. Livera, M. Theristis, G. Makrides, G.E. Georghiou, Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems, *Renew. Energy*. 133 (2019) 126–143.  
<https://doi.org/10.1016/j.renene.2018.09.101>.
- [21] A. Mellit, G.M. Tina, S.A. Kalogirou, Fault detection and diagnosis methods for photovoltaic systems: A review, *Renew. Sustain. Energy Rev.* 91 (2018) 1–17.  
<https://doi.org/10.1016/J.RSER.2018.03.062>.
- [22] S. Stettler, P. Toggweiler, E. Wiemken, W. Heidenreich, A.C. De Keizer, S. Feige, M. Schneider, G. Heilscher, A. Drews, D. Heinemann, Failure Detection Routine for Grid Connected PV Systems as Part of the PVSAT2 Project, in: 20th Eur. Photovolt. Sol. Energy Conf. Exhib., 2005: pp. 4–7.
- [23] P. Guerriero, L. Piegari, R. Rizzo, S. Daliento, Mismatch Based Diagnosis of PV Fields Relying on Monitored String Currents, *Int. J. Photoenergy*. 2017 (2017). <https://doi.org/10.1155/2017/2834685>.
- [24] M. Davarifar, A. Rabhi, A. El-Hajjaji, M. Dahmane, Real-time model base fault diagnosis of PV panels using statistical signal processing, *Proc. 2013 Int. Conf. Renew. Energy Res. Appl. ICRERA 2013*. (2013) 599–604.  
<https://doi.org/10.1109/ICRERA.2013.6749826>.
- [25] S.A.S. Eldin, M.S. Abd-Elhady, H.A. Kandil, Feasibility of solar tracking systems for PV panels in hot and cold regions, *Renew. Energy*. 85 (2016) 228–233. <https://doi.org/10.1016/j.renene.2015.06.051>.
- [26] J.C. Sánchez Barroso, N. Barth, J.P.M. Correia, S. Ahzi, M.A. Khaleel, A computational analysis of coupled thermal and electrical behavior of PV panels, *Sol. Energy Mater. Sol. Cells*. 148 (2016) 73–86.  
<https://doi.org/10.1016/j.solmat.2015.09.004>.
- [27] C. Ventura, G.M. Tina, Development of Models for On-line Diagnostic and Energy Assessment Analysis of PV Power Plants: The Study Case of 1 MW Sicilian PV Plant, *Energy Procedia*. 83 (2015) 248–257.  
<https://doi.org/10.1016/J.EGYPRO.2015.12.179>.
- [28] C. Ventura, G.M. Tina, Utility scale photovoltaic plant indices and models for on-line monitoring and fault detection purposes, *Electr. Power Syst. Res.* 136 (2016) 43–56.  
<https://doi.org/10.1016/J.EPSR.2016.02.006>.
- [29] R.C. Team, R: A Language and Environment for Statistical Computing, (2019).
- [30] C. Rigollier, O. Bauer, L. Wald, On the clear sky model of the ESRA - European Solar Radiation Atlas - With respect to the Heliosat method, *Sol. Energy*. 68 (2000) 33–48. [https://doi.org/10.1016/S0038-092X\(99\)00055-9](https://doi.org/10.1016/S0038-092X(99)00055-9).
- [31] A.B. Sproul, Derivation of the solar geometric relationships using vector analysis, *Renew. Energy*. 32 (2007) 1187–1205.  
<https://doi.org/10.1016/j.renene.2006.05.001>.
- [32] D. Yang, Solar radiation on inclined surfaces: Corrections and benchmarks, *Sol. Energy*. 136 (2016) 288–302.  
<https://doi.org/10.1016/J.SOLENER.2016.06.062>.
- [33] Linke Turbidity factor, Ozone, Water Vapor and Angstroembeta - [www.soda-pro.com](http://www.soda-pro.com). (n.d.). <http://www.soda-pro.com/es/web-services/atmosphere/linke-turbidity-factor-ozone-water-vapor-and-angstroembeta> (accessed September 14, 2017).
- [34] Y. Mu, X. Liu, L. Wang, A Pearson ' s correlation coefficient based decision tree and its parallel implementation, *Inf. Sci. (Ny)*. 435 (2018) 40–58.  
<https://doi.org/10.1016/j.ins.2017.12.059>.