



Supervision and fault detection system for photovoltaic installations based on classification algorithms

Marc Castellà¹, Konstantinos Kampouropoulos¹, Eva M. Urbano² and Luis Romeral²

¹ Fundació Eurecat - Centre Tecnològic
Plaça de la Ciència 2, 08242 Manresa (Spain)

Phone: +34 938 777 373, e-mail: marc.castella@eurecat.org, konstantinos.kampouropoulos@eurecat.org

² MCIA Research Center, Department of Electronic Engineering
Universitat Politècnica de Catalunya
Rambla de Sant Nebridi 22, 08222 Terrassa (Spain)

Phone: +34 937 98519, e-mail: eva.maria.urban@upc.edu, luis.romeral@upc.edu

Abstract. This article presents a methodology for the supervision and fault detection on photovoltaic installation, through the information gathered by their SCADA system. The proposed methodology consists of the use of a multi-clustering approach to analyse and classify the operating behaviour of the photovoltaic installations, using information of their DC voltage, generated current (per string), as well as information related to the climatic conditions of the park (i.e. solar irradiance, temperature). The proposed methodology uses a supervised training algorithm, based on a decision tree learning algorithm, allowing to determine the appearance of anomaly behaviours in the installations of photovoltaic plants, including soiling detection, hot-spots, tracker deviations, electric connections and faults in sensors. The presented methodology has been developed in the framework of a CORFO R&D project and validated under real operating conditions in a utility-scale photovoltaic power plant of one axis mount, located in Chile, with a total power of 40 MWp.

Key words. Fault detection, predictive maintenance, photovoltaics, decision tree learning.

1. Introduction

In the past years, the renewable energy production has become a topic of high interest for the society. Due to climate change and the continues reduction of the fossil-fuel reserves, the society has been focused on more efficient solutions, which promote the green technologies and the sustainability. Some of the most important examples of the energy and climatic actions that are being carried out to reduce the greenhouse gases, are the 20/20/2020, 2030 and 2050 climate and energy packages, established by the European Union [1]-[3]. In this context, the use of renewable energy is considered one of the most promising solutions in order to reach the established sustainability goals [4].

While there is a continuous study and evolution of the technologies related to wind turbines, concentrated solar

power plants and photovoltaics parks, the last ones have presented an exponential evolution in the last 20 years, due to their high efficiency and their reduced OPEX costs. In the last 20 years the installation of new plants has been speedily growing worldwide, reaching in 2017 almost 100GW. This evolution represents an annual growth of almost 30%, expecting an even higher increment in a short-term. Currently, the total global production capacity is estimated at 402.5GW, which corresponds to 2.14% of the global electricity demand [5].

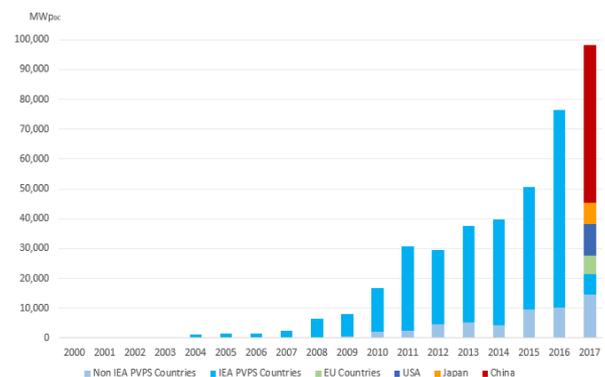


Figure. 1. Global annual evolution of new photovoltaic installation (MW) [5].

Nevertheless, despite the multiple benefits of these technologies, due to their accelerated growth, there are several new issues that are presented and have to be dealt. Part of them are related to the reliability of the electrical supply, while other with the need for detection of failures and anomalies in their behaviour. On one hand to maintain the energy efficiency of the installations, guaranteeing the expected production energy levels, while on the other hand, there is the need to supervise their operation and avoid possible disconnections, which could cause a power blackout of the power systems [6].

In this aspect, in general, there are mainly 3 methodologies for fault diagnosis: i) through a visual inspection that implies a considerable human effort, ii) through image processing that can be divided into thermography, electroluminescence and UV fluorescence, that implies the use of drone technologies and the disconnection of the park from the grid to operate it to the maximum production; and iii) through the electrical measurements analysis, which can be obtained through the installed monitoring system [7]. Considering the limitations of the first two methodologies, this study is focused on the use of the available data of the photovoltaic farm to extract information related to its current condition.

In the case of the regulation of the generation power in centralized PV, it is necessary to consider that the PV plants do not operate always at their maximum available power, but it is the network manager who through systems of generation power limitations (such as the Automatic Reduction Scheme Generation ERAG in Chile) establish at all times the production power based on the installed power, the level of radiation and the demand of the network. This implies that the inverters do not operate in maximum power tracking mode, but instead they try to reach the reference established by an external manager. Consequently, the maximum power value (MPPT), used in the literature for fault detection, cannot be used when identifying deviations in the production of photovoltaic parks connected to the network

In this aspect, this work is focused on the development of new solutions for the supervision, classification and fault detection of the operation of the photovoltaic installations, through the use of the already available information, acquired through the SCADA system of the installation.

2. Methodology

The proposed methodology is based on a double stage process, as depicted in Figure 2.

- i. The modelling process, in which the mathematical model of the plants' behaviour is being made, based on historical information of the plant;
- ii. The supervision and prediction process, in which the system operates in real time, monitoring the behaviour of the panels with objective to detect and diagnose the appearance of anomalies or faults.

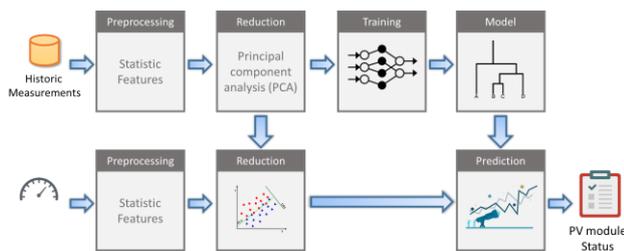


Figure 2. Block diagram of the proposed methodology.

As it can be observed in the figure, the modelling process consist of four main steps. Initially a pre-processing process is made over the historical data of the plant (i.e.

DC voltage of the panels, produced current by string, solar irradiance and surface temperature of the panels), in order to fix erroneous values, such as data gaps, infinite values and abnormal data registries, which can affect the accuracy of the training process. Additionally, during the pre-processing, an extraction of several statistic features is made, which are being used during the next steps of the modelling process. The static features that are extracted from the data are based on the type of faults that are being modelled, analysing their impact over the produced current of the panels. The following figures presents 3 examples of the current pattern of the panels, under the appearance of a hotspot fault, a bypass fault and a dc current sensor fault.

The healthy behaviour is assumed as the average instant current of all strings connected at the same DC bus of a central inverter.

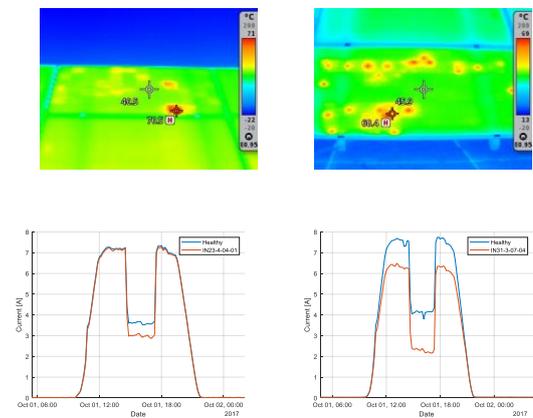


Figure 3. Comparison between current response of a healthy panel (blue) versus a hot spot fault (red), together with a thermography inspection of the faulty panel.

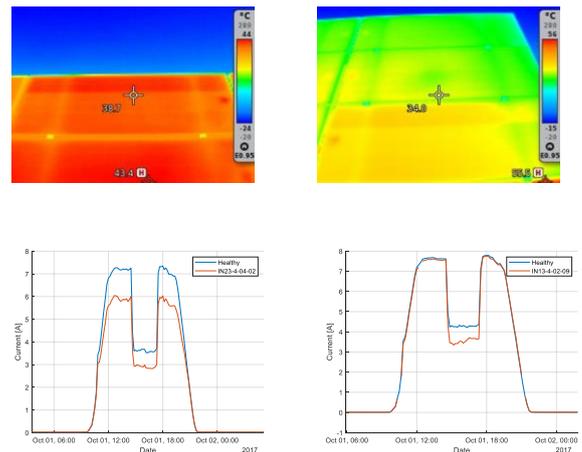


Figure 4. Comparison between current response of a healthy panel (blue) versus a by-pass fault (red), together with a thermography inspection of the faulty panel.

As it can be observed in the figures, the studied faults present a deviation of the produced current compared to its healthy state in several points of the daily operating profile. These points consist mainly on the maximum power production (midday) as well as at the times where the photovoltaic farm operates under power regulation

(with consequence to force a lower DV voltage for the string boxes). Additionally, in the case of a sensor fault, the measurements present a measurement of the half of the actual production, while there are registers of negative power generation during night time.

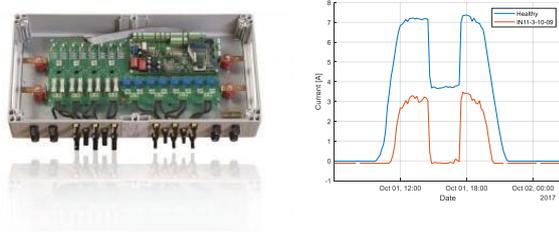


Figure. 5. Comparison between current response of a healthy panel (blue) versus a fault on the dc current sensor (red).

Based on the current behaviour over the different faults, and with objective to extract several indicators which can reveal the appearance of the different faulty states of the strings, five main indicators are extracted as described below.

- a) Percentage difference between the average current of the strings and the current generation of the specific string under study;

$$\frac{\sum(I_{mean} - I_{string})}{\sum(I_{mean})} * 100 \quad 1.$$

- b) Maximum percentage difference between average current and string current;

$$\max\left(\frac{I_{mean} - I_{string}}{I_{mean}} * 100\right) \quad 2.$$

- c) Average current variance and string current variance;

$$\sum_{i=1}^n (I_{string_i} - \overline{I_{string}}) \quad \sum_{i=1}^n (I_{mean_i} - \overline{I_{mean}}) \quad 3.$$

- d) Percentage difference between the average current and the string current at the point of maximum production;

$$\frac{I_{mean}^{t_{max}(I_{mean})} - I_{string}^{t_{max}(I_{mean})}}{I_{mean}^{t_{max}(I_{mean})}} * 100 \quad 4.$$

- e) Percentage difference between the average current and the string current at the point of minimum production.

$$\frac{I_{mean}^{t_{min}(I_{mean})} - I_{string}^{t_{min}(I_{mean})}}{I_{mean}^{t_{min}(I_{mean})}} * 100 \quad 5.$$

Once the (daily) features are extracted for each one of the strings, the process proceeds to their transformation to principle components, through the principle component

analysis (PCA). This stage permits to make a multivariate analysis of the statistical indicators and reduce the dimensional space for the training process.

After this PCA process, the training of the mathematical model of the strings takes place in order to correlate the operating conditions with the string's registered state. The information is obtained through the database of the SCADA system of the farm (i.e. generated current per string, solar irradiance, external temperature and DC voltage), together with a register of maintenance actions and observed faulty conditions. These observations are automatically set by the SCADA system to healthy state and faulty state, labelling also the type of faults over the faulty states.

Finally, a classification process takes place by means of a decision tree learning algorithm, operating in two levels. Initially, a single classification is made to split the data into the healthy and faulty clusters. Then it isolates only the data that correspond to the faulty class, and proceeds to a second clustering based on their label (type of fault). The following figure presents the two stages of clustering.

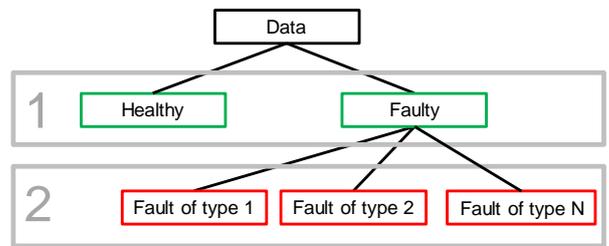


Figure. 6. Two stage process of the data clustering.

3. Experimental Results

This section presents a part of the obtained results, applying and validating the proposed methodology on a photovoltaic plant of 40MW, located in Atacama Desert in Chile. For the shake of example, the following figures present some of the type of faults that were used for the testing and validation of the training and diagnosis algorithm, extracted from the historical data of the photovoltaic farm.

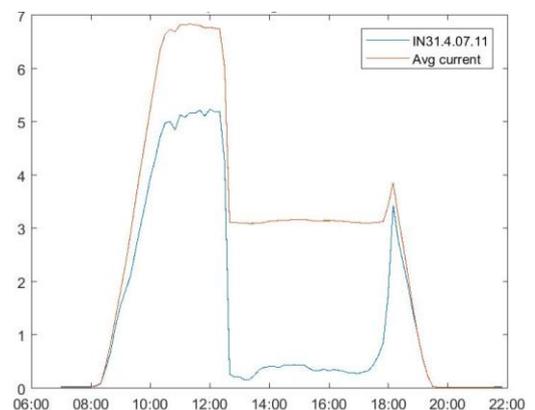


Figure. 7. Example of faulty behavior during power regulation.

4. Conclusions

This article presents a methodology for the supervision and fault detection of anomaly behaviour of photovoltaic panels, in installation of high power. It is based only on information gathered by the commercial SCADA systems of the photovoltaic plants, consisting of the climatic and operating conditions, which are the solar radiation, the string's temperature, the DC voltage and the produced current of each string (or stringbox, based on the available information).

The proposed method has been tested and validated under real operating conditions in a utility-scale photovoltaic power plant located in the north of Chile, in a total of over 15000 strings for a period of 9 months. During the testing, it permitted to identify soiling deposition and growth over the installation, sensor faults as well as hotspots in several panels, which were later validated by means of visual inspection of the maintenance staff.

The method was developed as a calculation algorithm in Matlab platform and it was integrated in a web server with direct communication with the installation's SCADA and databases for its real time operation.

References

- [1] European Commission, "20 20 by 2020, Europe's climate change opportunity. Brussels," Communication from the Commission to the European Parliament, Brussels, Jan. 2008.
- [2] European Commission, "A policy framework for climate and energy in the period from 2020 to 2030," Communication from the Commission to the European Parliament, Brussels, Jan. 2014.
- [3] European Commission, "A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy," Communication from the Commission to the European Parliament, Brussels, Nov. 2018.
- [4] L. Pietrosevoli and C. Rodríguez Monroy, "The impact of sustainable construction and knowledge management on sustainability goals. A review of the Venezuelan renewable energy sector," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 683–691, 2013.
- [5] International Energy Agency, "2018 Snapshot of Global Photovoltaic Markets - IEA PVPS," pp. 1–16, 2018.
- [6] K. Rahimi, S. Mohajeryami and A. Majzoobi, "Effects of photovoltaic systems on power quality," 2016 North American Power Symposium (NAPS), Denver, 2016, pp. 1-6. doi: 10.1109/NAPS.2016.7747955
- [7] M. Köntges, S. Kurtz, C.E. Packard et al, Review of Failures of Photovoltaic Modules, Ed. IEA-Photovoltaic Power Systems Programme 2014