Renewable Energy and Power Quality Journal (RE&PQJ), ISSN 2172-038 X, No.9, 12th May 2011, Reference Paper: 598

European Association for the Development of Renewable Energies, Environment and Power Quality (EA4EPQ)	International Conference on Renewable Energies and Power Quality (ICREPQ'11) Las Palmas de Gran Canaria (Spain), 13th to 15th April, 2011

Cumulative Statistical Analysis to Monitor the Energy Performance of PV Plants

S. Vergura

Dipartimento di Elettrotecnica ed Elettronica
Politecnico di Bari
Via E. Orabona 4, 70100 Bari (Italy)
Phone/Fax number:+39 080 5963590, e-mail: vergura@poliba.it

Abstract. This paper proposes a procedure based on statistical tools for diagnosis of PhotoVoltaic (PV) plants. As the data are acquired, statistical analyses are realized. At every new loop other data are added to the previous ones, implementing a cumulative statistical analysis. In this manner it is possible to follow the trend of some specific parameters and to understand the real operation of the PV plant as the environmental conditions change during the year. The proposed approach, based on ANOVA and Kruskal-Wallis tests, is effective in detecting and locating abnormal operating conditions. The proposed algorithm has been applied to a real case and results are presented.

Keywords
Photovoltaic plant, statistics, ANOVA, Kruskal-Wallis.

1. Introduction

In the design of PV plant a crucial problem is the strong dependence of the system response on many extrinsic factors, such as irradiance intensity, ambient temperature, cell temperature, air velocity, humidity, cloudiness and pollution. Successively, when a PV plant has been set up, a monitoring of the system to ensure an optimal performance with respect to the change of environmental conditions is needed.

Standard benchmarks [1], called "final PV system yield", "reference yield" and "Performance Ratio" (PR), are currently used to assess the overall system performance in terms of energy production, solar resource, and system losses. They are defined as follows.

a) Final PV system yield:

$$Y_f = \frac{E}{P_o}[kWh/kW] \qquad (1)$$

It is the net energy output E divided by the DC power P_0 of the installed PV array. It represents the number of hours that the PV array would need to operate at its rated power to provide the same energy.

b) Reference yield:

$$Y_r = \frac{H}{G}[hours] \qquad (2)$$

is the total in-plane irradiance H divided by the PV's reference irradiance G. It represents an equivalent number of hours at the reference irradiance. If G equals 1 kW/m^2, then Yr is the number of peak sun-hours. It is a function of the location, orientation of the PV array, and month-to-month and year-to-year weather variability.

c) Performance Ratio (PR):

$$PR = \frac{Y_f}{Y_r} \qquad (3)$$

it is related to the overall effect of losses on the rated output due to: a) inverter inefficiency, wiring, mismatch, and other losses when converting from d.c. to a.c. power; b) PV module temperature; c) incomplete use of irradiance by reflection from the module front surface; d) soiling or snow; e) system down-time; f) component failures.

Unfortunately, they exhibit two drawbacks: a) they supply a rough information about the performance of the overall PV plant; b) they do not allow any assessment of the behavior of the PV plant single parts.

Some authors have considered the use of the statistics for assessing solar PV plant [2], while a monitoring and decision algorithm based on two main theoretical branches of statistical science, namely *descriptive* and *inferential* statistics, has been developed in [3]. The former one is useful to characterize the data population by assigning a proper descriptive model or distribution family to it. The latter one, adopted when the entire set of data is unknown, consists of a data producing process trying to infer the behavior of the entire population from a sub-set of sample data. The idea at the basis of the procedure in [3] is to predict mis-operation events whatever the amount of field measurements is.

This paper proposes an algorithm able to analyze the operation of a PV plant as the data are acquired, implementing a cumulative statistical analysis. This approach allows to monitor also the *trend of some benchmarks* in order to evaluate the operation trend.

For this aim, the algorithm proposed in this paper is based on the whole population of the energy, even if, for a first stage of analysis, it could be applied to sampled data in order to verify if important failures are present.

The paper is structured as follows: Section II introduce the proposed algorithm, Section III describes the PV plant under test and finally Section IV presents the results.