SIGNAL PROCESSING FOR FAULT DETECTION IN PHOTOVOLTAIC ARRAYS

H. Braun⋆*, S. T. Buddha*⋆*, V. Krishnan*⋆*, A. Spanias*⋆*, C. Tepedelenlioglu*⋆*, T. Yeider*†*, and T. Takehara*†

School of ECEE, SenSIP Center and Industry Consortium,
⋆ Arizona State University † Paceco Corporation, 25503 Whitesell Street, Hayward, CA , USA
{hcbraun, sbuddha, vkrish20, spanias, cihan}@asu.edu, {tyeider, ttakehara}@pacecocorp.com

ABSTRACT

Photovoltaics (PV) is an important and rapidly growing area of research. With the advent of power system monitoring and communication technology collectively known as the "smart grid," an opportunity exists to apply signal processing techniques to monitoring and control of PV arrays. In this paper a monitoring system which provides real-time measurements of each PV module's voltage and current is considered. A fault detection algorithm formulated as a clustering problem and addressed using the robust minimum covariance determinant (MCD) estimator is described; its performance on simulated instances of arc and ground faults is evaluated. The algorithm is found to perform well on many types of faults commonly occurring in PV arrays.

***Index Terms*—** Electrical Fault Detection, Photovoltaic Systems

1. INTRODUCTION

Although rapid progress has been made in the construction of photovoltaic (PV) modules and inverters (DC-AC converters), management of PV arrays remains an open problem. To ameliorate this, a monitoring system can be deployed within a PV array, providing high-resolution real-time measurements. These data are available to be used to evaluate the performance of the array, quantify the effects of PV module aging, and quickly identify faults and under-performing modules. In this paper we present an algorithm which applies clustering, robust estimation, and anomaly detection techniques from signal processing to the problem of PV array fault detection.

PV arrays are highly reliable: a typical new PV module is sold with a 25-year warranty guaranteeing performance degradation of no more than 1% of rated power output per year. Inverters are somewhat less reliable, typically sold with a 5-year warranty. Despite the high reliability of PV system components, outages do occur: in 2007 an International Energy Agency (IEA) survey found array availability of 99.5% and 95.5% for PV arrays built in 2003 and 1995, respectively [1].

Current ad-hoc methods of fault detection result in lengthy repairs and decreased array availability. In [2], the mean time to repair (MTTR) of several PV systems was calculated and no array achieved a MTTR of less than 19 days. A MTTR of 3.3 days is given in [3] for a large-scale PV array located in Japan; this is noted as an extremely short time. If PV is to become a significant part of world energy production, more effective monitoring is clearly needed to ensure reliability and a fast and highly reliable method of detecting faults in PV arrays is needed.

Several attempts have been made to automate fault detection in PV arrays. In [4] an ad-hoc method is presented which detects faults based on lower-than-expected string currents or sudden drops in one string's current. In [5], the expected behavior of the array is calculated based on current environmental conditions and compared against measured output. Finally, in [6] analysis of variance (ANOVA) and the Kruskal-Wallis test are used to locate faults to within one sub-array of a large array connected to multiple inverters. Despite these efforts, to the best of our knowledge, our work [7, 8] is the only fault detection technique in the literature which can detect faults in individual modules using robust clustering algorithms.

This paper is organized as follows: Section 2 briefly describes the electrical behavior of PV arrays. The problem of PV array fault detection is described in Section 3 and several types of commonly occurring faults are described along with their effect on array performance. The MCD-based fault detection algorithm is presented in Section 4 and its performance in detection of simulated ground and arc faults is described in Section 5.

2. BRIEF OVERVIEW OF PHOTOVOLTAICS

Fig. 1 shows the single-diode model, a simple and commonly used representation of a PV cell, module, or array [9]. As with a diode, the behavior of this circuit may be summarized by its current-voltage (I-V) characteristic. Fig. 2 shows the simulated I-V and power-voltage (P-V) curves of a Sharp NT-175UC1 PV module at standard test conditions (STC) of 1000 W/m^2 solar irradiance and 25 °C cell temperature. Note that the module achieves its maximum power output at 35.4 V and

This work is funded by PACECO Corporation (subsidiary of the MITSUI Group) and supported in part by the SenSIP site of the NSF NCSS I/UCRC under award 1035086.

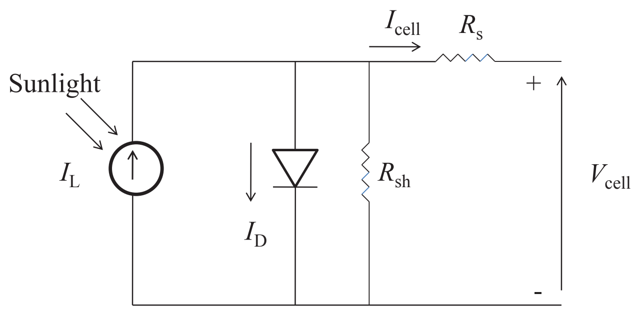

Fig. 1. Single-diode model of a PV cell.

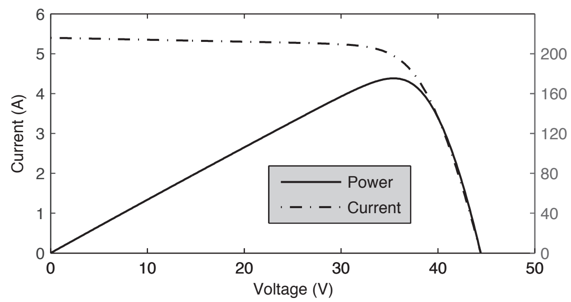

Fig. 2. I-V and P-V curves of Sharp NT-175UC1 module at STC.

4.95 A, achieved when the module is connected to a load of 7.15 Ω. This is known as the maximum power point (MPP) of a PV cell, module, or array. Modern inverters dynamically adjust the load they present to the array in order to maintain operation at the MPP, a process known as maximum power point tracking.

The component values of the single-diode model vary with environmental conditions, most notably temperature and solar irradiance. Light-generated current I_L is almost perfectly proportional to solar irradiance and is modeled as such. Temperature primarily affects the operation of the diode: I_D increases with increasing temperature, reducing the power output of the cell. Bypass diodes (not shown) are generally connected across several PV cells; this diode conducts large currents when the cell is reverse biased ($V_{cell} < 0$), effectively bending the I-V curve upward at negative voltages and minimizing power dissipation and heating of the cell.

PV modules generally consist of many cells connected in series strings, with strings occasionally connected in parallel to increase current output. A similar connection scheme is used to form arrays, with series strings of modules connected in parallel to achieve the desired voltage and current. Larger arrays are divided into sub-arrays, with each sub-array feeding a different inverter. When all modules in an array are perfectly electrically matched, available array power is simply the power of each module multiplied by the number of modules and the entire array's I-V characteristic can be adequately represented by the single-diode model in Fig. 1.

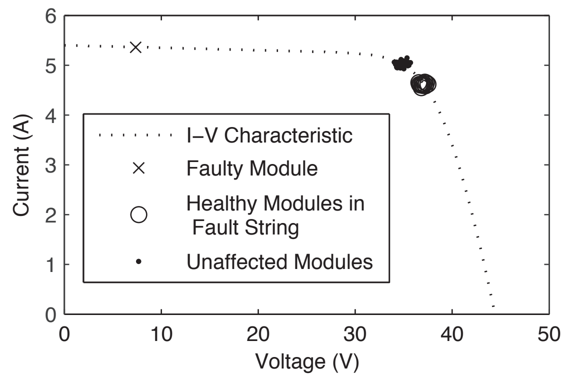

Fig. 3. Operating points of an array under ground fault conditions.

However, when mismatch between modules exists, this is no longer accurate, and array maximum power may not be the sum of individual module maximum powers. Such mismatch is often due to a fault in the array.

3. PROBLEM STATEMENT

Several conditions impair the performance of PV arrays, including partial shading, module soiling, inverter failure, and mismatch due to variation in manufacturing or aging of PV modules. In general, faults result in a characteristic pattern in which the faulty module(s) form a cluster in I-V space, while the rest of the string in which the fault occurs forms another cluster, and the remaining unaffected modules in the other strings form a third cluster. This paper focuses on the detection of ground and arc faults.

3.1. Ground Fault

A ground fault occurs when a PV array develops an unintentional low-resistance path to ground, such as when wire insulation is compromised, leaving it in contact with adjacent conductors. This shorts one or more modules in a string, leaving the rest of the modules to make up the difference in voltage. The overall effect is demonstrated in Fig. 3.

3.2. Arc Fault

An arc fault is the unintended flow of current through air or another dielectric. Arc faults are generally divided into two categories: series and parallel arcs. A series arc occurs when a short break is created in a conductor, such as when a connector pulls apart due to aging or thermal stress. A parallel arc occurs when two conductors of widely differing voltage are placed near one another, for instance when the insulation on a cable running over a steel frame is compromised. The current through a series arc is generally limited by the circuit with which it is in series, while a parallel arc can theoretically consume as much power as a source is able to supply. Fig. 4 shows the effect of a small (5 V) series arc within a module.

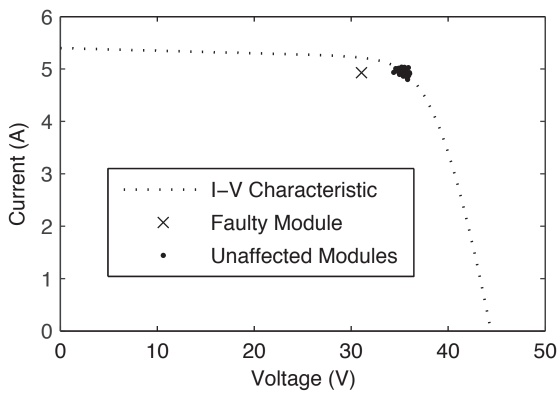

Fig. 4. Operating points of an array under series arc fault conditions.

This low-voltage arc causes only the faulty module to appear as an outlier in I-V space. However, larger arcs affect the entire string in which the arc occurs, just as in the ground fault case.

Several models exist which attempt to describe the behavior of an electric arc. One extremely simple approach is the Cassie model, which describes an arc as a time-varying conductance with constant steady-state voltage . The Mayr model takes a similar approach but assigns an arc a constant steady-state power dissipation. The Habedank model [10] consists of Cassie and Mayr arcs in series with one another.

4. ALGORITHM DESCRIPTION

The problem of identifying the presence of a fault in a PV array may be formulated as an outlier or anomaly detection problem: the presence of an outlier in I-V data indicates a fault. A large body of work exists in anomaly detection; however, the presence of multiple clustered outliers due to PV faults (e.g. Fig. 3) presents a unique challenge. The algorithm presented here uses a classical approach to outlier detection, employing more recent work in robust statistics to overcome the problem of multiple clustered anomalous observations.

The algorithm accepts pairs of voltage and current measurements from N individual PV modules as input, denoted $\mathbf{x}_i(t) = \begin{bmatrix} V_i(t) & I_i(t) \end{bmatrix}^T$, where $1 \le i \le N$ is the index of an individual module and t is time in seconds. Since values of V and I change quickly with weather variation, only data from a single time t is considered in the algorithm; dependence on t will be omitted from the following discussion. Thus, the input data may be visualized as a set of points in I-V space as in Figs. 3 and 4. The algorithm is not restricted to 2-dimensional $\mathbf{x}_i$ and can easily be adapted to incorporate a wider variety of measurements, for instance module temperature.

The algorithm computes a test statistic $T(\mathbf{x}_1, ..., \mathbf{x}_N)$ given by

$$T(\mathbf{x}_1, ..., \mathbf{x}_N) = \max_i \left(\sqrt{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_x)^T \hat{\mathbf{C}}_x^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_x)} \right) \tag{1}$$

in which $\hat{\boldsymbol{\mu}}_x$ and $\hat{\mathbf{C}}_x$ are estimates of sample mean and covariance matrix computed using the MCD estimator. The MCD (discussed below) identifies the $N/2$ most tightly clustered observations from the set of $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ and uses these values to estimate $\boldsymbol{\mu}_x$ and $\mathbf{C}_x$. Finally, $T(\mathbf{x}_1, ..., \mathbf{x}_N)$ is compared against a threshold γ and if $T(\mathbf{x}_1, ..., \mathbf{x}_N) > \gamma$ a fault is determined to be present. The threshold γ is currently chosen to achieve a desired false alarm rate.

The test statistic was selected to measure the maximum distance in I-V space between a module and the center of the modules' distribution. The distance measure used is a quadratic distance closely related to the well-known Mahalanobis distance, but uses different estimators for the mean and covariance matrix.

Selection of estimators for $\hat{\boldsymbol{\mu}}_x$ and $\hat{\mathbf{C}}_x$ is non-trivial: faults in PV arrays often cause the appearance of multiple clusters of module data in I-V space as in Fig. 3. These clusters prevent the use of conventional estimators of sample mean and covariance, since clusters of faulty modules and strings will mask the characteristics of the non-faulty data. To combat this masking effect, several robust estimators exist which discard large numbers of outlying observations and return estimates only of the cluster to which the majority of observations belong.

The minimum covariance determinant (MCD) estimator was used for this purpose. The MCD estimator [11] first determines the subset $S \subset X = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ of h observations (with $N/2 \le h \le N$) such that the determinant of the sample covariance matrix $|\hat{\mathbf{C}}_S|$ is minimized. This step effectively discards $N - h$ points as outliers; we chose $h = N/2$ for maximum robustness. Next, the estimated mean $\hat{\boldsymbol{\mu}}_x = \hat{\boldsymbol{\mu}}_S$ and covariance matrix $\hat{\mathbf{C}}_x = \alpha \hat{\mathbf{C}}_S$ are calculated, where

$$\alpha = \frac{\mathrm{med}_i \left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_S)^T \hat{\mathbf{C}}_S^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_S) \right)}{m} \tag{2}$$

is a consistency factor to correct for non-anomalous points excluded from S. med is the sample median and m is the population median of a χ_2^2 random variable. This is used because if $\mathbf{x}_i$ are 2-dimensional and normally distributed, $(\mathbf{x}_i - \boldsymbol{\mu}_x)^T \mathbf{C}_x^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_x)$ will be χ_2^2 distributed. Determining the exact subset S is prohibitively expensive to compute for large N. However, the more recent FAST-MCD estimator [11] computes good approximate results with acceptable complexity.

5. ALGORITHM PERFORMANCE

The performance of the MCD-based algorithm presented in Section 4 was evaluated using Monte Carlo analysis with data obtained from SPICE simulations of a PV array under

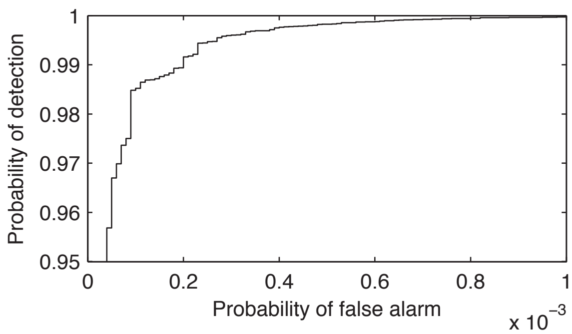

Fig. 5. ROC of detector for 5 V series arc fault (Cassie model) .

arc fault conditions. Faults were simulated on an array of 52 Sharp NT-175UC1 PV modules, arranged in 4 parallel-connected strings of 13 series-connected modules each. This configuration was chosen because it corresponds to a small utility-operated array in the vicinity of the ASU campus on which a prototype monitoring system has been installed. The PV module performance model given in [9] was used to calculate the SPICE circuit parameters for operation at standard test conditions (STC).

The metric used to evaluate algorithm performance was the Receiver Operating Characteristic (ROC), which quantifies the trade-off between sensitivity and specificity in a detector. The x-axis plots the probability of false alarm, while the y-axis plots probability of detection. Simulations were performed under the assumption of additive Gaussian measurement noise of variance $0.13\ \mathrm{V}^2$ and $0.0025\ \mathrm{A}^2$ for voltage and current, respectively; these values correspond to a standard deviation of 1% of the true (noiseless) value. Errors in voltage and current measurement were assumed to be independent of one another; in measurements from the prototype monitoring system this appears to be the case.

Series arc fault simulation was performed using the Cassie arc model for steady-state arc voltage of 5 V, corresponding to a relatively small arc, but one which is fully capable of destroying a PV module. This is the same arc fault shown in Fig. 4. Fig. 5 shows the ROC of the MCD-based detector operating on this arc fault at STC. It can be seen that the detector achieves a detection rate of 98% at a false alarm rate of only 0.01%.

6. CONCLUSIONS AND FUTURE WORK

The fault detection algorithm presented here promises the ability to detect a wide range of conditions affecting array output. The algorithm may be deployed as part of a comprehensive monitoring system which improves array efficiency and availability with a minimum of human operator involvement.

Although the algorithm shows good performance in simulations, several opportunities for improvement exist. First, the sampling period of the monitoring system may be increased so that voltages and currents may be treated as quasi-stationary. The algorithm may then be altered to consider data over a short time window rather than a single snapshot. Another approach is to incorporate measurements of irradiance and/or module temperature in predicting array output. A graphical user interface (GUI) for PV array fault detection and monitoring has been developed by SenSIP Center researchers in MATLAB [12] and an online GUI is also being developed in Java-DSP [13,14].

7. REFERENCES

[1] Luzi Clavadetscher and Thomas Nordmann, "Cost and performance trends in grid-connected photovoltaic systems and case studies," Tech. Rep., International Energy Agency: Photovoltaic Power Systems Programme, Dec. 2007.

[2] A.B. Maish, C. Atcitty, S. Hester, D. Greenberg, D. Osborn, D. Collier, and M. Brine, "Photovoltaic system reliability," in *Photovoltaic Specialists Conference, 1997., Conference Record of the Twenty-Sixth IEEE*, Sep. 1997, pp. 1049 –1054.

[3] K. Otani, T. Takashima, and K. Kurokawa, "Performance and reliability of 1 MW photovoltaic power facilities in AIST," in *Photovoltaic Energy Conversion, Conference Record of the 2006 IEEE 4th World Conference on*, May 2006, vol. 2, pp. 2046 –2049.

[4] Huang Zhiqiang and Guo Li, "Research and implementation of microcomputer online fault detection of solar array," in *Computer Science Education, 2009. ICCSE '09. 4th International Conference on*, Jul. 2009, pp. 1052 –1055.

[5] D. Stellbogen, "Use of PV circuit simulation for fault detection in PV array fields," in *Photovoltaic Specialists Conference, 1993., Conference Record of the Twenty Third IEEE*, May 1993, pp. 1302 –1307.

[6] S. Vergura, G. Acciani, V. Amoruso, and G. Patrono, "Inferential statistics for monitoring and fault forecasting of PV plants," in *ISIE 2008. Industrial Electronics, 2008, IEEE International Symposium on*, Jun. 2008, pp. 2414 –2419.

[7] Santoshi Tejasri Buddha, Henry Braun, Venkatachalam Krishnan, Cihan Tepedelenlioglu, Andreas Spanias, Ted Yeider, and Toru Takehara, "Signal processing for photovoltaic arrays," in *Int. Conf. on Emerging Signal Processing Applications*, Jan 2012.

[8] Henry Braun, Santoshi Tejasri Buddha, Venkatachalam Krishnan, Andreas Spanias, and Cihan Tepedelenlioglu, *Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization*, Morgan & Claypool, 2012, In preparation.

[9] W. De Soto, S. Klein, and W. Beckman, "Improvement and validation of a model for photovoltaic array performance," *Solar Energy*, vol. 80, no. 1, pp. 78 – 88, Aug. 2006.

[10] U. Habedank, "Application of a new arc model for the evaluation of short-circuit breaking tests," *Power Delivery, IEEE Transactions on*, vol. 8, no. 4, pp. 1921 –1925, Oct. 1993.

[11] Peter J. Rousseeuw and Katrien van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, Aug. 1999.

[12] Venkatachalam Krishnan, Santoshi Buddha, Henry Braun, Cihan Tepedelenlioglu, and Andreas Spanias, "A graphical user interface for photovoltaic fault detection and monitoring," in *IEEE Transactions on Industry Applications*, 2012, In Preparation.

[13] Andreas Spanias, *Digital Signal Processing; An Interactive Approach*, Andreas Spanias, May 2011.

[14] A. Spanias and V. Atti, "Interactive online undergraduate laboratories using J-DSP," *Education, IEEE Transactions on*, vol. 48, no. 4, pp. 735 – 749, nov. 2005.