

Outlier Detection Rules for Fault Detection in Solar Photovoltaic Arrays

Ye Zhao, Brad Lehman*

Department of Electrical and Computer Engineering
Northeastern University
Boston, MA, US
zhao.ye@husky.neu.edu

Roy Ball, Jerry Mosesian, Jean-François de Palma

MERSEN USA Newburyport-MA, LLC
Newburyport, MA, US

Abstract— Solar photovoltaic (PV) arrays are unique power sources that may have uncleared fault current when utilizing conventional overcurrent protection devices. To monitor the PV operation and detect these unnoticed faults, outlier detection rules have been proposed for fault detection based on instantaneous PV string current. This paper discusses three rules in detail: 3-Sigma rule, Hampel identifier, and Boxplot rule. Unlike other methods, the proposed methods do not require weather measurement or efforts in model training. Our experimental results show that Hampel identifier and Boxplot rule may be recommended for PV fault detection. Furthermore, the proposed models become more reliable as the number of PV measurements increases. The developed methods may be integrated with PV monitoring system for real-time operation.

I. INTRODUCTION

Fault detection and health monitoring are important to increase efficiency, reliability and safety of solar photovoltaic (PV) systems. Different from traditional power sources, solar PV array is unique because it may have undetectable and uncleared fault current by conventional overcurrent protection devices (OCPD). This may be caused by non-linear output characteristics of solar PV arrays, low irradiance, MPPT of the PV inverters, high fault impedance, or degradation of solar cells [1]. Without proper fault detection, uncleared faults in PV arrays not only cause power losses, but also might lead to safety issues and fire hazards [2]. While the topic of fault tolerance and detection has received considerable attention for microgrids [3], drives [4], [5], and power electronic converters/inverters [6], [7], there has been less analysis specifically for PV installation.

Several fault detection models and monitoring systems have been studied for PV systems [8]–[14]. PV monitoring and fault detection models based on energy yield and power losses have been proposed in [8]–[12]. Fault detection model based on PV module temperature, current and voltage using the Kalman filter has been proposed in [13]. A decision-tree model has been proposed for fault detection and classification for PV systems [14].

However, these methods may either require a costly and long-term training process (e.g. a few days or even weeks), or need a complex PV simulation model, which could hinder their effectiveness [8]–[14]. Besides, the trained fault detection model may be unique to the specific PV installation, which means it cannot be easily adopted in a new PV system. The models must be modified due to different PV capacity, solar cell technology, or installation location. Also, since a fault can only be detected on PV array or sub-array level, the fault detection's accuracy and sensitivity might be low [8]–[12]. For example, the fault cannot be noticed until the whole array/sub-array has a significant underperformance. In addition, requirement of weather conditions, such as solar irradiance and/or temperature, could bring complexity and uncertainty into the fault detection methods [8]–[10], [12]–[14].

To overcome the aforesaid limitations, this paper proposes a new fault detection method in PV arrays using outlier detection rules. If one PV string is faulted, it is likely to behave differently from others, resulting in anomalous electrical parameters (called outliers). In most cases, a PV string outlier is a reduced string current. Thus, by applying outlier detection rules, a normal range for PV string currents can be simply found based on instantaneous measurement of each string current. Then the fault detection method becomes straightforward: the string current out of the normal range will be identified as an outlier. Note that the normal range is updated continuously and it is determined by all PV string currents. For proof of concept, the developed methods have been tested on a small-scale PV system under real fault conditions. The experimental PV faults include line-line faults with and without fault impedance, open-circuit faults, degradation faults and partial shadings.

In summary, this paper presents the following research contributions:

- For the first time, outlier detection rules are developed for PV fault detection based on instantaneous measurements (e.g., PV string current). The detection method has several advantages, such as no model training procedure, independence with

*The author gratefully acknowledges the support through grants by Mersen USA and the National Science Foundation (under grant 0901439).

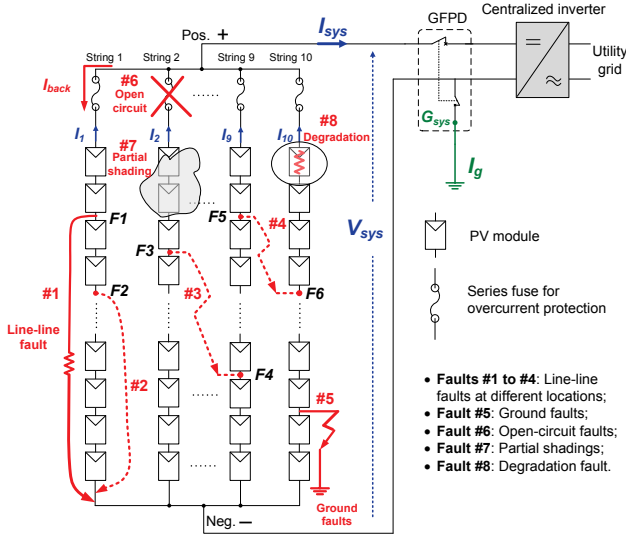


Figure 1. Typical faults in a grid-connected PV system [1]

irradiance and temperature sensors, inexpensive computational cost, real-time operation, and easy implementation in a software/algorithm;

- In addition to PV string level, the proposed methods can be simply extended to PV module, or sub-array levels. Besides using PV string current, the proposed methods may also include other readily available PV parameters, such as module insulation impedance, output power or energy yield. This could bring more flexibility into the fault detection methods.

This paper proceeds as follows: Section II briefly introduces the limitation of conventional protection devices in PV systems; Section III develops the outlier detection rules in solar PV systems; The experimental results by different outlier detection rules are discussed in Section IV; Section V concludes the paper.

II. FAULT DETECTION CHALLENGES IN TYPICAL SOLAR PV ARRAYS

A. Typical Grid-Connected PV Systems

A typical grid-connected PV system (see Fig. 1) consists of several major components, including a solar PV array, a centralized inverter with MPPT algorithm, electrical connection wirings, and protection devices, such as overcurrent protection devices (OCPD) and ground fault protection devices (GFPD) [15]. As the fundamental building block of PV systems, m number of PV modules can be assembled in series to build a PV string. Then, n number of PV strings can be in parallel to create a PV array with $m \times n$ PV modules.

Fig. 1 also shows several typical faults that may occur in the PV array, including line-line faults, ground faults, and mismatch faults (e.g., open-circuit faults, partial shadings, and degradation faults [16]). Among these faults, line-line faults and mismatch faults are focused in this paper, since

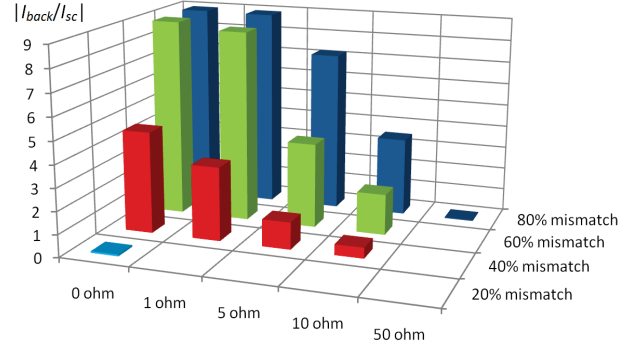


Figure 2. Transient peaks of $|I_{back}/I_{sc}|$ under various fault locations and fault impedances [1]

they are more difficult to detect by conventional protection devices than ground faults.

B. "Blind Spots" of OCPD

The US National Electrical Code (NEC) requires that a single OCPD (e.g., fuse) shall be used in series-connected strings to protect PV modules and conductors. The rating current of OCPD (I_N) shall be no less than 156 percent of PV module's rated short-circuit current (I_{sc}) at standard test condition [15]. Furthermore, fuses only melt according to their current vs. melting time characteristics. For instance, the minimum breaking capacity ($I_{min-break}$) of fuses is rated at $1.35I_N$, according to UL standard 2579-7 [17]. Therefore, the OCPD (i.e., PV fuses) is only able to clear faults and isolate faulty circuits if they carry a fault current larger than $2.1I_{sc}$ ($I_{min-break} = 1.35 \times 1.56I_{sc}$). However, the fault current may be greatly reduced below $I_{min-break}$ by many factors, such as fault location, fault impedance, solar irradiance and even the MPPT of the inverter [1]. This could lead to the protection failure and "blind spots" of OCPD in the PV systems.

For example in Fig. 1, a large PV system (~ 17.6 kW) consisting of 10×10 PV modules ($m=10$ and $n=10$) is simulated in MATLAB/Simulink [1]. Line-line faults with four typical fault locations are studied: 1) Between $F1$ and negative conductor ($F1-Neg$); 2) Between $F2$ and negative conductor ($F2-Neg$); 3) Between $F3$ and $F4$ ($F3-F4$); and 4) Between $F5$ and $F6$ ($F5-F6$). Line-line fault #1 is defined as "80% location mismatch", since it involves 8-module mismatch between the fault points in the faulted string (normally 10 modules per string). Similarly, line-line fault #2, #3, and #4 are defined as 60%, 40% and 20% location mismatches, respectively. The faulted string current may become negative after the fault, which is usually called the backfed current (I_{back}).

Fig. 2 has shown that fault location and fault impedance can influence the reduction of $|I_{back}|$ to become below the minimum breaking capacity of PV fuses. Also, the faulted string may still have positive current under low mismatch faults and high fault impedance. Thus, the fault may remain in the PV array so that the protection "blind spots" occurs. Moreover, the uncleared fault might bring potential safety hazards to PV systems, even though $|I_{back}|$ is small. For instance, I_{back} in the PV array could sustain the DC arcs on the fault path and bring fire hazards.

Our research goal is to develop a new fault detection system to eliminate the aforesaid “blind spots” in PV protection schemes.

III. OUTLIER DETECTION RULES IN SOLAR PV ARRAYS

In data mining, an outlier is usually defined as anomalous data with respect to the majority of the dataset [18]. For PV systems, an underperforming/faulty PV-component could be viewed as an outlier when its electrical parameter deviates markedly from other normal ones. As shown in Fig. 1, a PV system consisting of parallel strings is the most dominant type in solar market [19]. Since each string shares the same array voltage, their current measurement might be convenient to build the outlier detection rules.

A. Parameter Estimation of Normal Distribution

If we assume each parallel PV string has identical electrical ratings and similar environmental conditions, each string should have similar current under normal conditions. Also, we assume the PV strings operate at their maximum power point (MPP) with the help of the centralized inverter. Therefore, given certain environmental conditions, each string current becomes a univariate variable x_i ($i=1\dots n$), where n is the number of strings in parallel. Furthermore, the entire set of string currents can be viewed as random samples coming from a normal distribution $N(\mu, \sigma^2)$, with two parameters: mean μ and variance σ^2 (σ is the standard deviation). Also we assume $\{x_1\dots x_n\}$ is independent and identically distributed (i.i.d.). The normal probability density function $p(x)$ is defined as:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} \quad (1)$$

It is also necessary to mention that $p(x)$ is not only dependent on PV installation, but also is strongly related to irradiance level (less related to temperature) [13]. As solar irradiance is changing continuously during a day, the parameters of $p(x)$ may vary as well.

However, true values μ and σ are unknown in real-world conditions because of limited samples and measurement noise. Instead, we can estimate $p(x)$ using the sample mean $\hat{\mu}$ and the sample variance $\hat{\sigma}^2$, based on maximum likelihood estimation [20]. Since the sample variance in (3) is biased, we use an unbiased estimator for the sample variance, described in (4).

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (3)$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (4)$$

For demonstration purposes only, a normal distribution of PV string currents under certain environmental conditions is

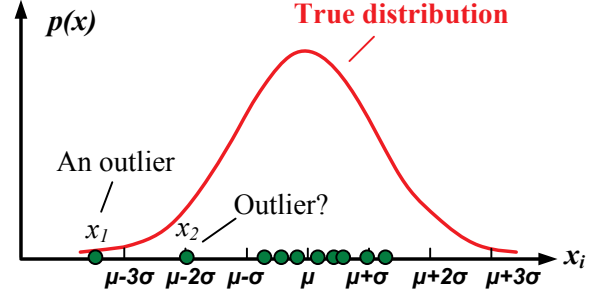


Figure 3. Univariate normal distribution of PV string currents

drawn in Fig. 3. Every green dot represents a measurement of string current at specific time $t=t_l$. Notice that most of values are located near the true mean μ , except for two values x_1 and x_2 . By common sense, we may pick x_1 as an outlier, since it is far away from μ . But it is difficult to identify x_2 . This leads to the need of a quantitative approach to set the outlier detection rules.

The accuracy of parameter estimation of $p(x)$ is limited by the number of samples $\{x_1\dots x_n\}$. The intuition is that if more samples are available, the true distribution will be better estimated. In addition, the parameter estimation is greatly related to the contamination level of $\{x_1\dots x_n\}$, which is defined as the fraction of outliers in the dataset [18]. Higher contamination level tends to bring more difficulty to parameter estimation.

B. 3-Sigma Rule

The 3-Sigma rule states that nearly all (99.73%) of values lie within 3 standard deviations of the mean. If the sample x_i lies out of that range as described in (5), it is a good chance to say x_i is an outlier.

$$|x_i - \hat{\mu}| > 3\hat{\sigma} \quad (5)$$

The 3-Sigma rule might be the best-known criterion to detect an outlier, but it is likely to have poor performance in practice, especially when the sample size is small (e.g., one outlier out of 10 samples). The main reason is that the sample mean and the sample variance are so vulnerable to outliers that they may remarkably deviate from the true values. Therefore, 3-Sigma rule may break down at contamination level greater than 10%.

C. Hampel Identifier

Instead of the sample mean, the Hampel identifier uses the sample median \tilde{x} as the reference value, defined as:

$$|x_i - \tilde{x}| > \alpha S \quad (6)$$

$$S = \frac{1}{0.6745} \text{median}\{|x_i - \tilde{x}|\} \quad (7)$$

where α is a threshold parameter (typical value is 3 [18]), and S is the median absolute deviation (MAD) scale estimator, defined in (7). Since the sample median is relative outlier-resistant, S in Hampel identifier can tolerate any outlier less than 50% contamination level.

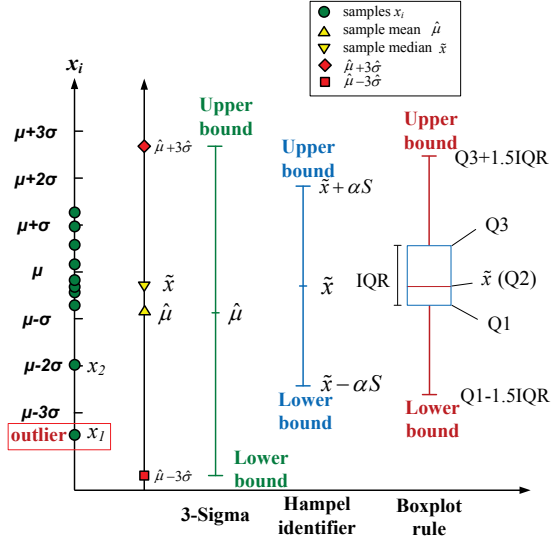


Figure 4. An illustrative example of outlier detection rules

D. Boxplot Outlier Rule

Another outlier detection rule uses the quartile-based boxplot, which uses the lower quartile (Q1, 25th percentile), the median (\tilde{x} or Q2, 50th percentile), the upper quartile (Q3, 75th percentile), and the interquartile range ($IQR=Q3-Q1$). Then Boxplot outlier rule can be expressed as:

$$x_i > Q_3 + 1.5IQR \cup x_i < Q_1 - 1.5IQR \quad (8)$$

Unlike either 3-Sigma rule or Hampel identifier, Boxplot outlier rule is extended to asymmetric data distributions. This may help its performance in outlier detection in PV string current, since most of the outliers are reduced PV string current, which are located on the left side of the sample mean (see Fig. 3). However, a practical disadvantage of Boxplot rule is that IQR is not as outlier-resistant as MAD in Hampel identifier. For example, IQR breaks down at contamination levels greater than 25%.

E. Illustrative Example

An illustrative example is shown in Fig. 4 to give some intuition about three outlier detection rules. The leftmost axis shows the samples and true distribution parameters (same as Fig. 3). Fig. 4 shows that the sample mean deviates from the true mean because of the outlier x_1 . Therefore, x_1 still lies in the normal range of 3-Sigma rule, resulting in the failure of outlier detection. On the other hand, Hampel identifier and Boxplot rule can identify x_1 as an outlier, since it drops below the lower bound of these two rules. Meanwhile, the value x_2 is identified as a normal sample by all rules. Moreover, the contamination level is 10% in this particular case, which can explain the sample mean's breakdown in 3-Sigma rule.

IV. EXPERIMENTAL RESULTS

A. Experimental Set Up

A small-scale grid-connected PV system has been set up to create and record each PV string current ($I_{pv-string}$) under

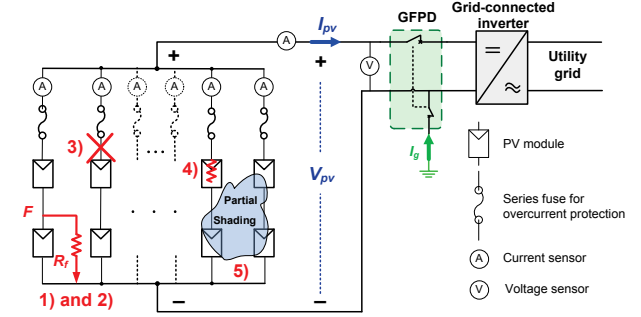


Figure 5. Schematic diagram of the experimental PV system

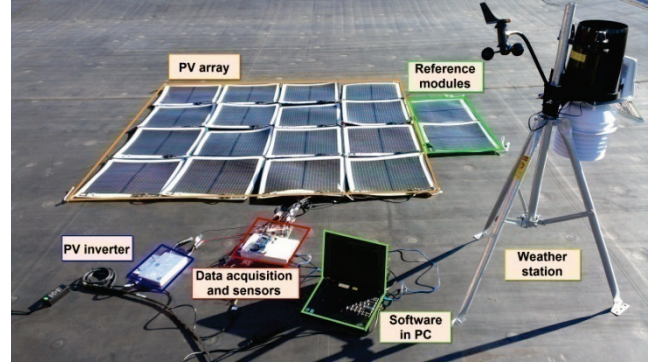


Figure 6. Photo of the experimental PV system

real working conditions. The schematic diagram and the photo are shown in Fig. 5 and Fig. 6, respectively. The parameters of PV components are summarized in Table I. The sampling frequency is arbitrarily chosen as 100Hz. A moving average filter (the span is 20 samples) is used to smooth the measurement. According to the weather station, the irradiance level during experiments is varying from $430W/m^2$ to $530W/m^2$. Note that the proposed fault detection algorithm does not need weather information. The solar irradiance provided by the weather station (see Fig. 6) is only for reference.

Five types of faults mentioned previously have been created (see Fig. 5): 1) Line-line fault between the middle of one string and negative bus bar; 2) The same line-line fault with a fault impedance $R=20$ ohm; 3) Open-circuit faults on a string; 4) Degradation fault on a PV module, emulated by inserting a series resistor R_{insert} ; 5) Partial shadings on PV modules. It is necessary to mention that the fault current of all faults is so small that it cannot be detected by OCPD.

TABLE I. EXPERIMENTAL PV COMPONENTS

Equipment	Parameters	
	Type	Detailed parameters
PV module	Power Film (amorphous silicon)	At STC: $V_{oc}=18V$, $I_{sc}=0.9A$, $V_{mpp}=14V$, $I_{mpp}=0.75A$, $P_{mpp}=10.5W$
Entire PV array	2×8 modules configuration	At STC: $P_{sys mpp}=168W$, $V_{sys mpp}=28V$, $I_{sys mpp}=6A$
Grid-connected inverter	Enphase microinverter M190	Max. output power 190W, min. start voltage: 28V; MPPT voltage range: 22V to 40V

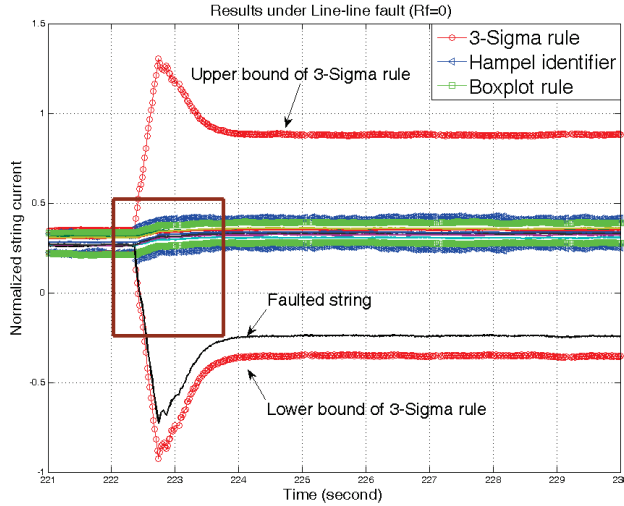


Figure 7. Experimental result: a line-line fault ($R_f=0$) in 8 strings

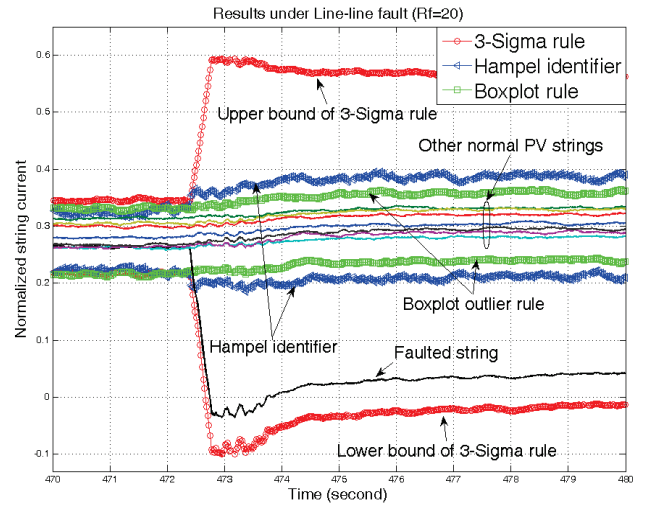


Figure 9. Experimental result: a line-line fault ($R_f=20$ ohm) in 8 strings

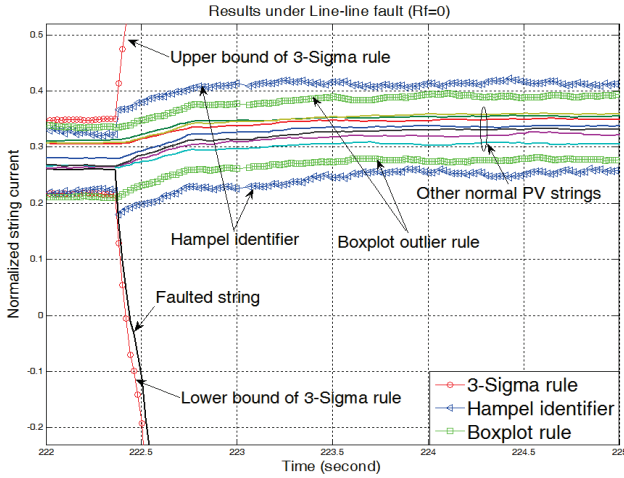


Figure 8. Experimental result: zoom in of a line-line fault ($R_f=0$ ohm) in 8 strings

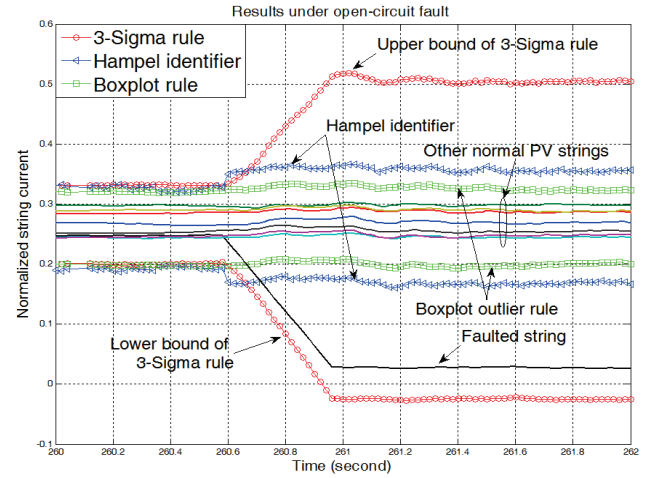


Figure 10. Experimental result: an open-circuit fault in 8 strings

In the following experimental results, two sets of samples with different contamination levels (12.5% and 33.3%) are used to study the outlier detection rules. In the following figures, the current axis is normalized by the module's short-circuit current (I_{sc}) for clarity of explanation, i.e., $I_{pv-string}/I_{sc}$ in the vertical axis. As a result, the outlier detection rules are normalized by I_{sc} as well. For comparison among three different outlier rules, their upper bound and lower bound are plotted in the same figure.

B. Experimental Results #1 – Contamination Level=12.5%

In this data set, since only one string is faulted out of 8 parallel strings, the outlier contamination level is 12.5% ($=1/8$). This might lead to the breakdown of sample mean in 3-Sigma rules. But Hampel identifier and Boxplot should tolerate and identify the outlier.

1) *Line-line faults ($R_f=0$):* As shown in Fig. 7, the fault occurs at $t=222.4$ s. The faulted string current drops below zero immediately and becomes a negative backfed current (I_{back}). With the help of the MPPT of the inverter, $|I_{back}|$ is reduced gradually and becomes a stable value after $t=224$ s [21]. Notice that this $|I_{back}|$ (the peak is $\sim 0.9I_{sc}$) is not high

enough to be detected by OCPD ($I_{min-break}$ is $2.1I_{sc}$), leading to potential safety issues.

The proposed outlier detection rule is calculated and updated continuously. It shows that 3-Sigma rule fails to detect the fault, as the faulted string current is always in the normal range of 3-Sigma rule. The reason is that, as we discussed in Section III.B, 3-Sigma rule is prone to fail at contamination level greater than 10%. The sample mean in 3-Sigma rule is significantly deviated by the outlier, and so does its lower bound. The rectangular area in Fig. 7 is zoomed in and plotted in Fig. 8 for easy explanation. It clearly shows that both Hampel identifier and Boxplot rule can successfully detect the fault at $t\sim 222.4$ s, when the faulted string current drops below the lower bound of these two outlier rules. In addition, Boxplot rule has a tighter normal range than Hampel identifier.

2) *Line-line faults ($R_f=20$ ohm):* The same line-line fault occurs with a 20 ohm fault impedance. Fig. 9 shows that R_f reduces the peak of $|I_{back}|$ at the moment of the fault. Also, the faulted string settles down to a small positive current ($0.04I_{sc}$). For the same reason in the previous fault, 3-Sigma

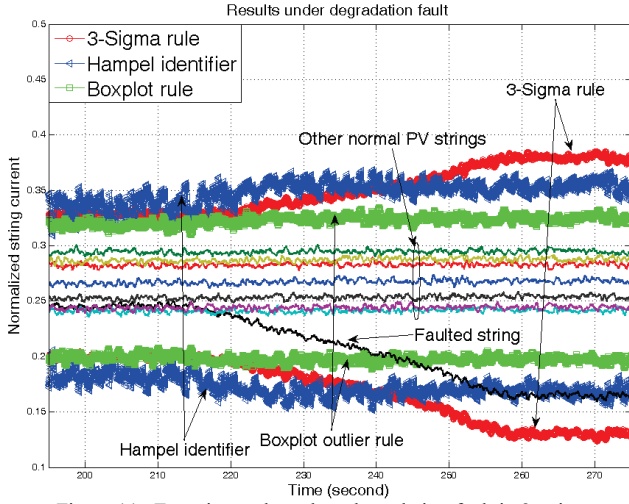


Figure 11. Experimental result: a degradation fault in 8 strings

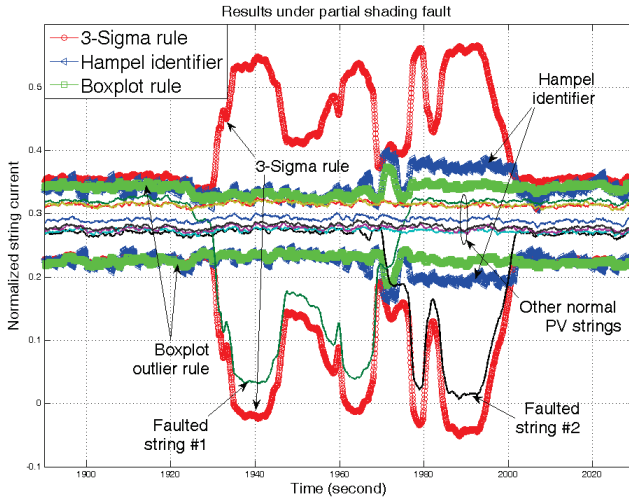


Figure 12. Experimental result: a partial shading in 8 strings

rule fails in fault detection, while the other two rules identify the fault at $t \sim 472.4$ s successfully. Similar to the previous case, Boxplot rule has a narrower normal range than Hampel identifier.

3) *Open-circuit faults*: As shown in Fig. 10, one string becomes open-circuit at fault ($t \sim 260.6$ s), which might be caused by cracked solar cells, solder joint failure, or blown fuse. This fault can be detected by Hampel identifier and Boxplot rule, but remains undetected by 3-Sigma rule.

4) *Degradation faults ($R_{insert}=0$ to 50 ohms)*: A variable resistor R_{insert} is inserted into PV string to emulate the degraded PV module. This is a challenging fault to OCPD since it always has a positive current. Fig. 11 shows that the faulted string current decreases as R_{insert} increases during $t=200$ s to 260s. Notice that the fault current can only be detected by Boxplot rule.

5) *Partial shading faults*: As shown in Fig. 12, a moving shadow is passing string #1 and #2, which causes partial (non-uniform) irradiance and a reduced string current. Notice that 3-Sigma rule deviates significantly with the reduced PV

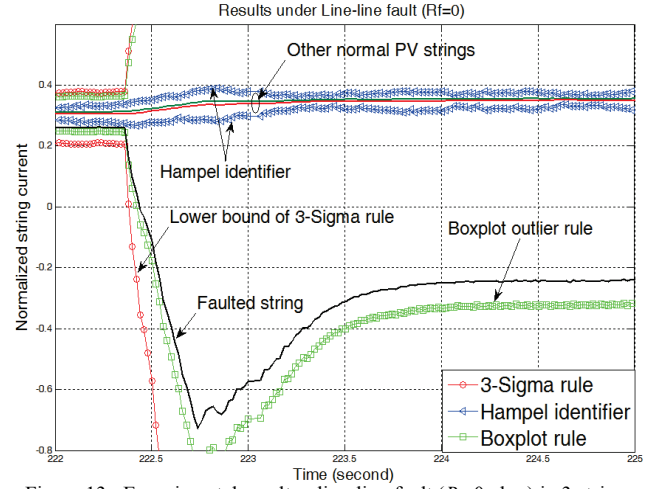


Figure 13. Experimental result: a line-line fault ($R_f=0$ ohm) in 3 strings

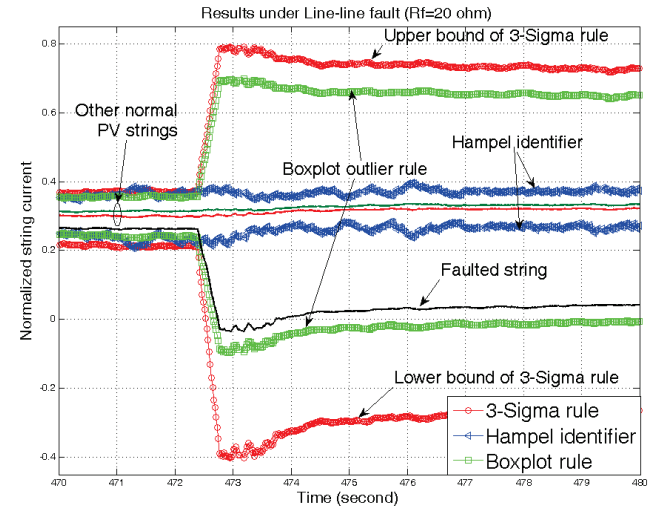


Figure 14. Experimental result: a line-line fault ($R_f=20$ ohm) in 3 strings

string current. The other rules can detect the partial shading fault successfully. On the other hand, Hampel identifier may be so aggressive at normal conditions (during $t=1900$ s to 1920s) that it could cause false alarm by its upper bound.

C. Experimental Results #2 – Contamination Level=33.3%

The same fault scenarios are created as the ones in Experimental Results #1, but we only consider the current of three PV strings (other 5 normal strings are ignored). If one string is faulted, the contamination level becomes as high as 33.3% ($= 1/3$). As we discussed in previous sections, this might lead to the breakdown of sample mean in 3-Sigma rule and failure of IQR in Boxplot rule. But in theory, MAD scale estimator in Hampel identifier should tolerate this high contamination level.

1) *Line-line faults ($R_f=0$)*: As shown in Fig. 13, the fault occurs at $t=222.4$ s. As we expected, the fault can not be detected by 3-Sigma rule and Boxplot rule, because of the high contamination level. On the other hand, Hampel identifier has a narrow normal range that is less affected by the outlier. Therefore, the fault can be identified by Hampel identifier. However, Hampel identifier might be aggressive

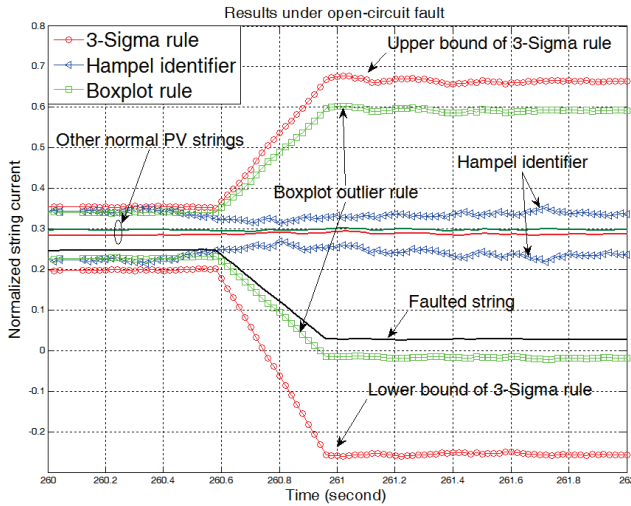


Figure 15. Experimental result: an open-circuit fault in 3 strings

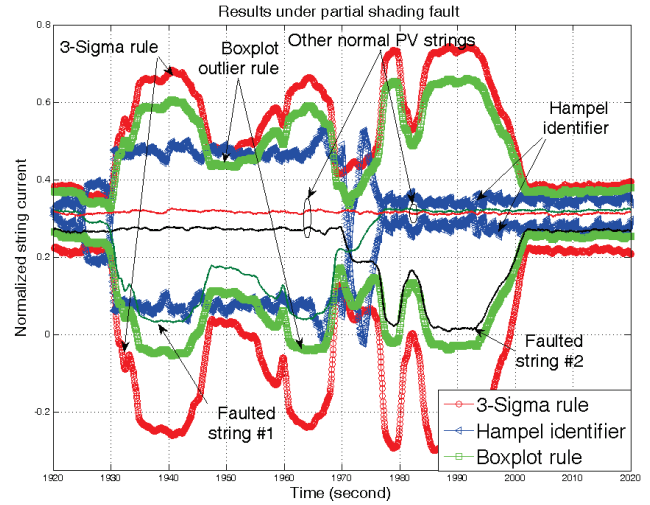


Figure 17. Experimental result: a partial shading fault in 3 strings

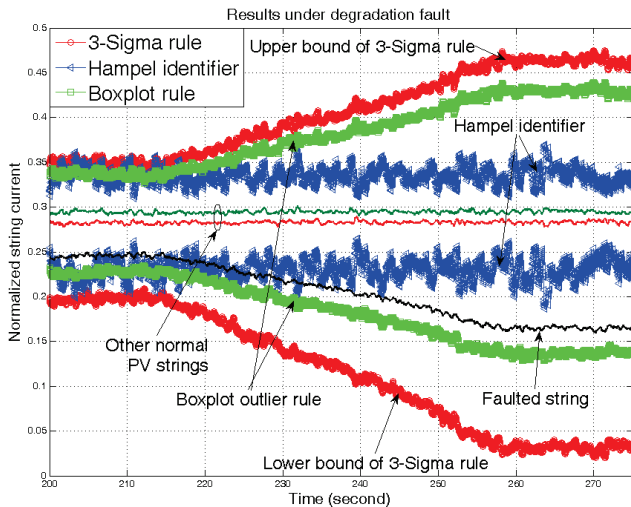


Figure 16. Experimental result: a degradation fault in 3 strings

under normal conditions (before the fault) that it can trigger the false alarm.

2) *Line-line faults* ($R_f = 20 \text{ ohm}$): The same line-line fault occurs with a 20 ohm fault impedance (see Fig. 14). For the same reason in the previous fault, 3-Sigma rule and Boxplot fail in fault detection, since the faulted string current is always between their normal range. Notice that Hampel identifier has a stable normal range that can identify the fault at $t \approx 472.4\text{s}$ successfully.

3) *Open-circuit faults*: Shown in Fig. 15, this fault can be detected by Hampel identifier, but remains undetected by 3-Sigma rule and Boxplot rule.

4) *Degradation faults* ($R_{insert} = 0 \text{ to } 50 \text{ ohms}$): Fig. 16 shows that the fault current can only be detected by Hampel identifier. However, before the fault occurs at $t = 210\text{s}$, the faulted string current may drop below the lower bound of Hampel identifier occasionally, leading to false alarms.

5) *Partial shading faults*: As shown in Fig. 17, 3-Sigma rule and Boxplot rule deviate significantly as the shaded string current changes. As a result, the faulted string current is always between the upper and lower bound of these two rules. On the other hand, Hampel identifier may detect the partial shading fault successfully. However, Hampel identifier may be so aggressive at normal conditions (during $t = 1920\text{s}$ to 1930s , and $t = 2005\text{s}$ to 2020s) that it could cause false alarms by its lower bound.

D. Discussion

In summary, the experimental results of 5 particular fault scenarios at 12.5% contamination level (from Fig. 7 to Fig. 12) and 33.3% contamination level (from Fig. 13 to Fig. 17) have shown that:

- *3-Sigma rule*: It fails to detect the fault in all cases, since it is vulnerable to outliers and breaks down quickly at contamination level greater than 10%. Thus, it is not recommended for PV fault detection.
- *Hampel identifier*: it can detect all faults, except for degradation faults at 12.5% contamination level. It has the best performance under 33.5% contamination level, because it has the best tolerance of outliers. Sometimes it may be too aggressive at normal conditions that may trigger false alarms (e.g. partial shading faults in Fig. 12 and Fig. 17, line-line faults with $R_f = 0 \text{ ohm}$ in Fig. 13, and degradation faults in Fig. 16).
- *Boxplot rule*: At 12.5% contamination level, Boxplot rule has the best performance among three outlier rules. It can detect all faults successfully. Also, it exhibits low outlier-sensitivity compared to other rules, and it has the most narrow normal range in all fault cases. However, Boxplot rule fails in all fault cases at 33.3% contamination level, since it breaks down at contamination level greater than 25%. In addition, Boxplot rule has better computational complexity than Hampel identifier. According to (6)

and (8), Hampel identifier needs two sorting steps to find medians, but Boxplot only needs one. This could explain why Boxplot rule has a smoother normal range than Hampel identifier in all fault cases.

- The comparison of outlier detection rules between 12.5% and 33% contamination levels has shown that more strings in measurement could lead to more reliable fault detection methods and less false alarms.

V. CONCLUSIONS

This paper demonstrates that common types of solar PV faults may not be cleared by traditional OCPD (e.g., fuses), including line-line faults under low irradiance, line-line faults with high fault impedance, open-circuit faults, degradation faults, and moving partial shadings. These faults can remain hidden in the PV system, keeping all the possible dangers associated with it (fire hazard, sub-optimal performance, arcing, etc.).

To detect these types of hidden faults, a real-time PV-string monitoring and fault detection method is proposed. Different from previous works, this paper developed and compared the fault detection methods based on three different outlier detection rules: 3-Sigma, Hampel identifier, and Boxplot rule. The experimental results show that Boxplot rule has the best accuracy and robustness under PV faults at relative high contamination level (e.g., 12.5% in our case). However, at extremely high contamination level (e.g. 33.3% in this paper), Hampel identifier performs best since it has the lowest outlier-sensitivity compared to other rules. In general, the paper shows that the outlier detection rules tend to become more reliable if more PV strings are taken into consideration.

Although the outlier detection rules are based on PV-string level measurement in this paper, the proposed methods should be easily implemented on any PV installation level with minor modification. For example, the outlier rules can be applied on PV-module level or sub-array level, if we assume the solar irradiance is identical on the same PV level. Besides using PV string current, the measurement may also be broadened to include some readily available PV parameters, such as PV insulation impedance, output power or energy yield. This could bring more flexibility into the fault detection methods.

ACKNOWLEDGMENT

The authors would like to thank Ling Yang of Department of Electrical and Computer Engineering at Northeastern University for her invaluable assistance in experimental set up and fault tests on PV systems.

REFERENCES

[1] Y. Zhao, J. F. de Palma, J. Mosesian, R. Lyons, and B. Lehman, "Line-Line Fault Analysis and Protection Challenges in Solar Photovoltaic Arrays," *IEEE Trans. Ind. Electron.*, to appear.

[2] B. Brooks. (2011) The Bakersfield Fire - A Lesson in Ground-Fault Protection. *SolarPro Mag.* pp. 62-70.

[3] J. Rocabert, G. M. S. Azevedo, A. Luna, J. M. Guerrero, J. I. Candela, P. Rodriguez, "Intelligent Connection Agent for Three-Phase Grid-Connected Microgrids," *IEEE Trans. Power Electron.*, vol. 26, pp. 2993-3005, 2011.

[4] H. Berriri, M. W. Naouar, and I. Slama-Belkhdja, "Easy and Fast Sensor Fault Detection and Isolation Algorithm for Electrical Drives," *IEEE Trans. Power Electron.*, vol. 27, pp. 490-499, 2012.

[5] J. Urresty, J. Riba, and L. Romeral, "Diagnosis of Interturn Faults in PMSMs Operating Under Nonstationary Conditions by Applying Order Tracking Filtering," *IEEE Trans. Power Electron.*, vol. 28, pp. 507-515, 2013.

[6] S. Yantao and W. Bingsen, "Survey on Reliability of Power Electronic Systems," *IEEE Trans. Power Electron.*, vol. 28, pp. 591-604, 2013.

[7] P. Xuejun, N. Songsong, C. Yu, and K. Yong, "Open-Circuit Fault Diagnosis and Fault-Tolerant Strategies for Full-Bridge DC-DC Converters," *IEEE Trans. Power Electron.*, vol. 27, pp. 2550-2565, 2012.

[8] H. Haeblerlin and C. Beutler, "Normalized Representation of Energy and Power for Analysis of Performance and on-line Error Detection in PV-Systems," in *13th EU PV Conference on Photovoltaic Solar Energy Conversion*, Nice, France, 1995.

[9] A. Drews, A. C. de Keizer, H. G. Beyer, E. Lorenz, J. Betcke, W. G. J. H. M. van Sark, W. Heydenreich, E. Wiemken, S. Stettler, P. Toggweiler, S. Bofinger, M. Schneider, G. Heilscher, and D. Heinemann, "Monitoring and remote failure detection of grid-connected PV systems based on satellite observations," *Solar Energy*, vol. 81, pp. 548-564, 2007.

[10] S. K. Firth, K. J. Lomas, and S. J. Rees, "A simple model of PV system performance and its use in fault detection," *Solar Energy*, vol. 84, pp. 624-635, 2010.

[11] S. Vergura, G. Acciani, V. Amoroso, G. E. Patrono, and F. Vacca, "Descriptive and Inferential Statistics for Supervising and Monitoring the Operation of PV Plants," *IEEE Trans. Ind. Electron.*, vol. 56, pp. 4456-4464, 2009.

[12] A. Chouder and S. Silvestre, "Automatic supervision and fault detection of PV systems based on power losses analysis," *Energy Conversion and Management*, vol. 51, pp. 1929 - 1937, 2010.

[13] B. K. Kang, S. T. Kim, S. H. Bae, and J. W. Park, "Diagnosis of Output Power Lowering in a PV Array by Using the Kalman-Filter Algorithm," *IEEE Trans. Energy Convers.*, vol. PP, pp. 1-10, 2012.

[14] Y. Zhao, Y. Yang, B. Lehman, J. F. de Palma, J. Mosesian, and R. Lyons, "Decision tree-based fault detection and classification in solar photovoltaic arrays," in *Proc. 27th IEEE APEC*, 2012, pp. 93-99.

[15] NFPA70, "National Electrical Code," in *Article 690 - Solar Photovoltaic Systems*, ed, 2011.

[16] E. L. Meyer and E. E. van Dyk, "Assessing the reliability and degradation of photovoltaic module performance parameters," *IEEE Trans. Rel.*, vol. 53, pp. 83-92, 2004.

[17] UL, "Outline of investigation for low-voltage fuses - fuses for photovoltaic systems," in *UL Subject 2579*, ed: Underwriters Laboratories Inc., 2010.

[18] R. K. Pearson, *Mining imperfect data: Dealing with contamination and incomplete records*: Society for Industrial and Applied Mathematics, 2005.

[19] IEA, "Trends in photovoltaic applications - Survey report of selected IEA countries between 1992 and 2009," IEA, Rep. IEA-PVPS T1-19:2010, 2010.

[20] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley, 2001.

[21] Y. Zhao, B. Lehman, J. F. de Palma, J. Mosesian, and R. Lyons, "Challenges to Overcurrent Protection Devices under Line-line Faults in Solar Photovoltaic Arrays," in *Proc. IEEE ECCE*, Phoenix, AZ, 2011, pp. 20-27.