Online module-level fault detection of PV arrays using an improved two-stage Hampel identifier

F C Han\textsuperscript{1}, Z C Chen\textsuperscript{1,3}, L J Wu\textsuperscript{1}, J L Yu\textsuperscript{1,3}, C Long\textsuperscript{2}, P J Lin\textsuperscript{1}, S Y Cheng\textsuperscript{1} and F P Su\textsuperscript{1}

\textsuperscript{1}College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China
\textsuperscript{2}Institute of Energy, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

E-mail: zhicong.chen@fzu.edu.cn/jlyu@semi.ac.cn

Abstract. This paper presents a novel fault detection method for photovoltaic (PV) array using an improved two-stage Hampel identifier, which can quickly detect the occurrence of faults and locate them at PV module level. The fault detection method is implemented in a two-level wireless sensor network based distributed on-line monitoring system which is developed to monitor the voltage of each PV module and current of all parallel PV strings in real-time. The 1\textsuperscript{st} stage Hampel identifier detects and locates the faulty PV strings by comparing the instantaneous currents of different PV strings. The 2\textsuperscript{nd} stage Hampel identifier detects and locates the faulty PV modules by comparing the instantaneous voltages of different PV modules in the faulty strings. Experimental results show that the proposed method can successfully detect and locate faulty strings and modules in several faulty cases, including the line-line faults, the degradation faults, the partial shading faults and the open-circuit faults.

1. Introduction

Due to the rapid depletion of fossil fuel sources worldwide, the global installed capacity of photovoltaic (PV) power plants has been obviously growing because solar energy is friendly and inexhaustible [1]. However, PV systems are prone to suffer many kinds of faults, which can affect electricity generation efficiency and even threaten the safety of user [2-4]. Therefore, developing efficient fault detection approaches is essential to detect and locate the faults for PV system maintenance. Among the existing detection methods, the higher detection resolution usually requires the larger external assistance.

The time-domain reflectometry (TDR) [5] is proposed to locate the fault at the PV-module level. However, it operates completely off-line and needs an expensive signal generator. The earth capacitance measurement (ECM) [6] is proposed to detect the open-circuit location in a string through analyzing the module capacitance value to earth. However, it is only suitable for series PV arrays and operates offline with low efficiency. The terminal characteristic analysis methods in [7] can locate the faulty modules based on the difference between various attributes, but the thresholds for different voltage readings need to be established by trial and error. Analytical method proposed in [8] can determine the number of open-circuit and short-circuit PV modules. However, it can locate only open-circuit and short-circuit faults at PV-string level. Chen et al proposed a fault location algorithm to adaptively localize different faults up to the string even module level while it requires the prior

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
knowledge about the training of autoregressive (AR) model and application of the generalized local likelihood ratio (GLLR) approach [1]. The machine learning approaches require prior knowledge about the training of model and usually have high computational complexity, which are not always practical for real time applications. In this paper, we propose an online modular level fault detection approach for PV arrays based on two-stage Hampel identifier with real-time operation, low complexity and high detection performance. The proposed fault detection method could be integrated into the PV system with a reconfiguration algorithm to maximize the energy generation of PV array during its lifetime [9], also could be embedded into the combiner box to realize the design of the intelligent combiner box, which shows the great potential for solving PV fault problems.

2. System architecture
A distributed on-line monitoring system based on a two-level wireless sensor network (WSN) is developed for real-time status monitoring and online modular level fault location, as illustrated in figure 1. Note that the PV array consists of three parallel PV strings and each PV string is configured as six PV modules in series, as illustrated in figure 2. The first level WSN monitors the PV module working voltage and achieves single hop communication between the sensing nodes installed on PV modules (PVM) and sensing nodes installed on combiner boxes of PV sub-arrays (PVA) via the low-cost wireless transceiver nRF24L01, as illustrated in figure 3. The PVA nodes receives the module voltage data from PVM and monitors each PV string current, and then send the acquired and received data to the PC via the CC2530 based ZigBee network, which makes up the second level WSN. Finally, the received PV string currents and PV module voltages are analyzed and calculated to build the outlier detection rules in Matlab, as illustrated in figure 4.

3. Proposed location algorithm for PV array faults
3.1. Hampel identifier for outlier detection
A method with real-time operation, low complexity and high detection performance is necessary to identify the fault location in PV array. It is reliable to employ some statistical approaches to remove
anomalous observations from data. 3-Sigma rule, Boxplot outlier rule and Hampel identifier are three popular outlier detection approaches. However, 3-Sigma has poor performance in practice and may break down when the contamination level of data exceeds 10% because it takes the sample mean as the reference value, which is greatly vulnerable to outliers. Boxplot outlier rule is based on the quartile and the interquartile range, which has stronger robustness regardless of the probability distribution of the data. Unfortunately, it can only identify certain configurations of outlier data and the performance for detecting the outliers obviously drops when the sample size increases, and it breaks down when the contamination level of data is higher than 25%. The Hampel identifier possesses the highest tolerance for outliers among aforementioned outlier detection approaches, which takes the sample median as the reference value, described as:

$$|x_i - \bar{x}| > \alpha S$$

(1)

$$S = 1.4826 \text{median} \left\{ |x_i - \bar{x}| \right\}$$

(2)

where $\alpha$ is the detection threshold factor (default value is 3, the value can be changed according to the actual situation), and $S$ is the median absolute deviation (MAD). The sample median in Hampel identifier, unlike the sample mean, measures the exact middle of dataset and is relative outlier-resistant. Although one or more elements in data sets are replaced by other value, it does not lead to significant changes in the sample median. It is worth noting that breakdown threshold value of the median is as high as 50%, which enables it identify any outlier of no more than 50% contamination level [10]. To detect the fault more accurately, an outlier detection algorithm should have a high degree of tolerance for outliers. Given that the Hampel identifier can resist up to 50% outliers, it is adopted in this study for outlier detection.

3.2. Two-stage Hampel identifier method and location procedures

![Figure 5. The flow chart of proposed two-stage Hampel identifier.](image)

To quickly respond to the fault happening and identify the fault location at the PV-module level, a two-stage Hampel identifier is proposed in this paper to detect various faults in PV array, as illustrated in figure 5. In the first stage, the Hampel identifier algorithm is used to realize the fault location at PV-string level by comparing the instantaneous currents of different PV strings. If one PV string is
suffering the faults, the ability to generate electricity from sunlight is greatly reduced and inefficient, and this inefficient state does not change until the failure is cleared. In most cases, the outlier in PV string means a decreased string current [11]. Therefore, the PV string current, as a measurement parameter for string level fault detection feature in this study, might be optimal to carry out the outlier detection rules because each PV string provides the same PV-array voltage whereas their string current can have some differences. Note that the string current at normal conditions may be out of the upper and lower bounds occasionally, leading to false alarms [11]. Therefore, a reasonable threshold is added in this study to avoid false detection, described in (3).

$$\frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}}} \leq 5\%$$

(3)

After localizing the faulty PV strings based on the each PV-string current measurement, a further step is indispensable to determine the fault location among all modules in the faulty strings. Then in the second stage, Hampel identifier algorithm is adopted again to identify the fault location at PV-module level by comparing the instantaneous voltage of different PV modules in faulty string. If one PV module is suffering the faults, it may behave differently from intra-string others, leading to abnormal electrical parameters. Therefore, the PV module voltage, as a measurement parameter for modular level fault detection feature in this study, might be optimal to carry out the outlier detection rules because each PV module shares the same string current whereas their voltage can be different.

4. Experimental setup and results

4.1. Simulation of typical faults

Four classical faults in PV arrays are considered to conduct fault experiments, as shown in figure 6. The line-line (LL) fault refers to an accidental short-circuit connection between two different potential points in PV arrays. There are two major fault scenarios: intra-string line-line (I-LL) fault and cross-string line-line (C-LL) fault. Particularly, C-LL fault will influence two or more neighboring strings simultaneously, which is a great challenge to correctly localize all short-circuit points in different strings [1]. The degradation fault means the gradual deterioration of PV component characteristics, which usually causes changes in parasitic resistances of PV modules and then reduce the power output. Shading fault mainly refers to the non-uniform irradiation during the operation of the PV array, which happens when some PV cells/modules of PV array are blocked by some obstructions. Open-circuit (OPEN) fault is mainly caused by the accidental disconnection in PV cells /modules/strings.

![Figure 6. The setup schematic of typical faults.](image)
the electrical connection wirings among the PV modules.

4.2. Experimental results
Four typical fault experiments on the real PV array are carried out on sunny days, and the experimental results are shown as in figures 7 to 16. By applying two-stage Hampel identifier method, the normal range for PV string currents and PV module voltages can be simply found, and the normal range is calculated and updated continuously based on instantaneous measurement of all PV string currents and PV module voltages. It can be observed that under normal working conditions, all the PV string currents and PV module voltages are always in the normal range of Hampel identifier. Once a fault occurs in the PV array, the faulted string currents would drop significantly below the lower bound of Hampel identifier, and then quickly locate faulty string. We can observe that the first stage Hampel identifier can successfully locate faulty string in various cases, including all the LL faults, the degradation faults, the partial shading faults and the OPEN faults. Note that the string current at normal conditions may be out of the upper and lower bounds occasionally, leading to false alarms, as shown in figure 12. Therefore, the threshold is calculated to eliminate these misjudgments.

Figure 7. I-LL with one module difference.
Figure 8. I-LL with two modules difference.
Figure 9. Shading with one module difference.
Figure 10. Shading with two modules difference.
Figure 11. Module degradation with one module difference.

Figure 12. Threshold analysis.

Figure 13. C-LL with one module difference.

Figure 14. C-LL with two modules difference.

Figure 15. Open-circuit fault.

Figure 16. String degradation fault.
A further step is to determine the fault location among all modules in the faulty strings. However, some faults cannot be accurately located to the modular level due to almost identical electrical parameters for each module, including the string degradation faults and OPEN faults. Additionally, C-LL fault may be misclassified by Hampel identifier, since the fault may cause dataset more than 50% contamination level. The decision on C-LL fault localization is recommended to make by the terminal voltage difference between the module in faulty string and the corresponding module in normal string, but it is not the aim of this work to calculate and compare those difference. Proposed location procedures can identify most fault locations successfully at PV-module level, such as the I-LL faults, the module degradation faults, the partial shading faults, as shown in figures 7-11.

5. Conclusions
A novel PV array fault detection method based on an improved two-stage Hampel identifier is proposed in this study, which only uses the current of all PV strings and voltage of each PV module. The algorithm is implemented in a developed two-level wireless monitoring system based on NRF24L01 and ZigBee. Different types and levels of faults have been created in a laboratory PV system to verify the proposed approach, and experimental results show that the proposed detection method can correctly detect and locate various faults at the strings and modules level, such as the line-line faults, the module degradation faults, the partial shading faults and open-circuit faults.

Acknowledgments
The authors would like to acknowledge the financial supports by the National Natural Science Foundation of China (Grant Nos. 61601127, 51508105 and 61574038), the Fujian Provincial Department of Science and Technology of China (Grant No. 2016H6012 and 2018J0106), the Fujian Provincial Economic and Information Technology Commission of China (Grant No. 830020, 83016006), and the Science Foundation of Fujian Education Department of China (Grant No. JAT160073).

References
[1] Chen L and Wang X 2017 Adaptive fault localization in photovoltaic systems IEEE T. Smart Grid Doi: 10.1109/TSG.2017.2722821
[2] Chen Z C, Wu L J, Cheng S Y, Lin P J, Wu Y and Lin W C 2017 Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and IV characteristics Appl. Energ. 204 912-31
[3] Chen Z C, Wu L J, Lin P J, Wu Y and Cheng S Y 2016 Parameters identification of photovoltaic models using hybrid adaptive Nelder-Mead simplex algorithm based on eagle strategy Appl. Energ. 182 47-57
[4] Alam M K, Khan F, Johnson J and Flicker J 2015 A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques IEEE J Photovolti 5 982-97
[5] Alam M K, Khan F, Johnson J and Flicker J 2013 PV ground-fault detection using spread spectrum time domain reflectometry (SSTDR) 2013 IEEE Energy Conversion Congress and Exposition (Denver, CO, USA) pp 1015-102
[6] Takashima T, Yamaguchi J, Otani K, Kato K and Ishida M 2006 Experimental studies of failure detection methods in PV module strings 2006 IEEE 4th World Conference on Photovoltaic Energy Conference pp 2227-30
[7] Madeti S R and Singh S N 2017 Online fault detection and the economic analysis of grid-connected photovoltaic systems Energy 134 121-35
[8] Gokmen N, Karatepe E, Celik B and Silvestre S 2012 Simple diagnostic approach for determining of faulted PV modules in string based PV arrays Sol. Energy 86 3364-77
[9] Balato M, Costanzo L and Vitelli M 2016 Reconfiguration of PV modules: A tool to get the best compromise between maximization of the extracted power and minimization of localized heating phenomena Sol. Energy 138 105-18
[10] Field M S 2011 Application of robust statistical methods to background tracer data characterized by outliers and left-censored data *Water Res.* **45** 3107-18

[11] Zhao Y, Lehman B, Ball R, Mosesian J and de Palma J F 2013 Outlier detection rules for fault detection in solar photovoltaic arrays *2013 Twenty-Eighth Annual IEEE Applied Power Electronics Conference and Exposition (APEC)* pp 2913-20