SAND2016-1468C

Automatic Fault Classification of Photovoltaic Strings Based on an In-Situ IV Characterization System and a Gaussian Process Algorithm.

C. Birk Jones*, Ryan Smith†, Craig K. Carmignani*, Olga Lavrova*, Bruce H. King*, and Joshua S. Stein*

*Sandia National Laboratories Solar PV & Grid Integration, Albuquerque, NM, U.S.A
†Pordis LLC, Austin, TX, U.S.A

***Abstract*—Current-voltage (I-V) curve traces of photovoltaic systems can provide detailed information for diagnosing fault conditions. The present work implemented a in-situ, automatic I-V curve tracer system coupled with a Gaussian Process classification learning algorithm to diagnose normal and abnormal behavior. The approach successfully identified normal, mismatch conditions, and other faults. In addition, the Gaussian Process regression algorithm was used to estimate normal behavior given irradiance and temperature conditions. The estimation results were then used to calculate the lost power due to the fault condition.**

***Index Terms*—IV characterization, gaussian process algorithm, pv fault classification**

I. Introduction

Reliable operations of photovoltaic (PV) plants requires advanced monitoring of string level performance. Many PV arrays now include the monitoring of DC voltage and current at the combiner box level. This level of monitoring can increase the chances of detecting faults and has been discussed in past literature [1], [2], [3]. However, a complete understanding of string level characteristics can be achieved through current-voltage (I-V) curve traces [4]. For example, in-situ module level I-V traces were performed by Quiroz *et al.* to test the impact of partial shading and increased series resistance effects [5]. However, the approach did not integrate an automatic evaluation tool.

The present work implemented the model 140A automatic, string level I-V curve tracer produced by Pordis LLC. The I-V curve results were presented automatically to a Gaussian Process (GP) learning algorithm that provided two services: (1) classification of string behavior, and (2) a estimation of lost power due to degraded performance.

II. Methodology

Automatic fault diagnostics of in-situ I-V curves was performed using a GP algorithm to first classify the existing condition, and then estimate normal behavior. The estimate of normal behavior was performed so that a potential loss of electrical power caused by the fault condition could be calculated. The process, described in Figure 1, began with the presentation of curve data to the GP classification machine. The classifier determined if the particular I-V curve was a fault or not. If a fault was not detected the curve was determined to be normal and the process ended. However, if the fault was found then the GP regression algorithm estimated the potential I-V curve under normal operating conditions. Based on this estimate the lost power was calculated by comparing it to the actual I-V curve data for the particular instance.

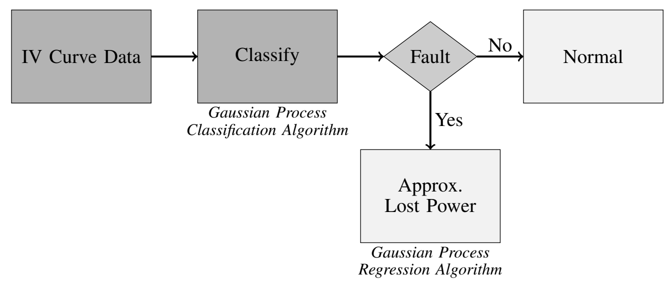

Fig. 1: The I-V curve data was evaluated in a multistep process. First, the GP was used to classify the I-V curve as either a fault or normal condition. If a fault was discovered then the lost power production was computed. If not, then operations continued as normal.

The proposed approach used a GP classification and regression algorithm. The algorithms were presented with a training data set, $D = (x_i, y)|i = 1, ...n)$. The data set included the input feature vectors x and the expected value(s) y. The testing data set, which included the same x input features from training, but different vector values (x_*). The testing outcome was the expected value y_*. The classification of the I-V curve data as normal or fault condition was performed by a GP algorithm that considered the data set where x = ([irradiance$_i$,temp$_i$,voltage vector$_i$,current vector$_i$]) and y = (fault label$_i$). The approximation of the lost power performed by the GP regression algorithm used the data set where x = ([irradiance$_i$,module temp$_i$] and y = ([voltage vector$_i$, current vector$_i$]) to determine the most likely curve without a fault present. Once the most likely curve was estimated the difference between the actual and estimated was calculated to determine the lost power.

A. In-Situ IV Characterization

The present work used a Regional Test Center (https://rtc.sandia.gov/) PV array as a test-bed for the in-situ I-V characterization and fault classification experiments. The array was constructed facing due south and had five strings as shown in Figure 2. Each of the strings connected to the Pordis LLC I-V tracer system. This system, known as the Pordis LLC 140A I-V curve tracer, was designed to be inserted into an array between the strings and the combiner box. The system has the capability to accommodate eight strings of up to 15A and 1000V per string [6]. It was designed as an in-situ tracer, which means that it may remain connected to the array at all times without impacting normal operations.

The 140A I-V curve tracer system could efficiently perform I-V traces in-situ with the array because of a unique hybrid switch circuitry. The circuitry provided a low resistance path though the device during periods of normal energy production. Trigger events, defined in the user interface, commanded the I-V characterization sweeps for each string at predefined instances throughout the day. When triggered, the tracer redirected the selected string to the load portion of the device,

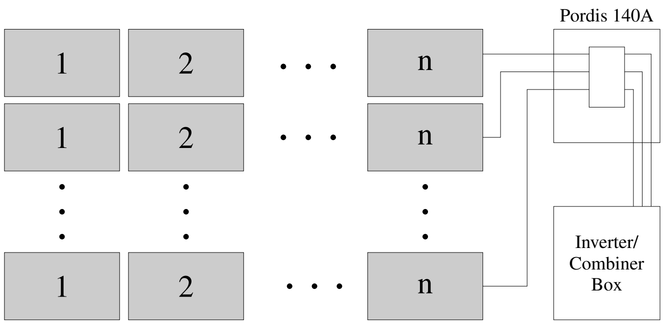

Fig. 2: The test array connected each string to the Pordis IV Characterization System and then to the combiner box before connecting to the inverter.

a I-V trace was performed, and then the string was switched back into the array; the duration of the entire tracing cycle was less than 100ms. Additionally, the hybrid switch circuitry incorporated in the tracer did not trip the high-frequency arc fault detection of the inverter used in the experiments. The results from each of the string I-V traces were stored in a database located in the tracer system and could be viewed through a web-interface. The GP algorithms accessed the I-V data automatically and provided feedback to the user immediately after the I-V trace tests were performed.

B. Gaussian Process Algorithm

The observations of inputs x_i and outputs y_i were presented to the GP supervised learning algorithm. Typical learning algorithms assume that $y_i = f(x_i)$ for some unknown function f. For example, if the expected underlying function was linear then a least-squares method to fit a straight line could be applied. On the other hand, $f(x)$ may be quadratic or cubic, and in this case other model types can be used. In this case, the GP algorithm provides a unique approach that does not relate $f(x)$ to a specific model. Instead it represents $f(x)$ by inferring a distribution over functions given a set of training data and then uses it to make predictions given new inputs [7].

GP can be defined as a set of random variables where any finite number of the set have a joint Gaussian distribution. GP applies a distribution over functions that are specified by a mean function and a covariance function as shown in Equation 1.

$$f(\mathbf{x}) \sim GP(\mu(x), k(x, x')) \tag{1}$$

The mean function, $\mu(x)$, is usually defined to be zero and the covariance $k(x, x')$ defines the prior properties of the functions considered for inference [8]. The k in the covariance represents the kernel function which projects the data into a higher dimensional feature space to increase the computational power of the algorithm [9]. The present work applied the GP regression and classification methods to the I-V characteristic data set to perform fault diagnostics and estimate lost power caused by the fault.

III. Results

The GP classification and regression algorithm performed well for the given I-V curve data set produced by the Pordis I-V tracer system. The classification results, described in Section III-A, indicated that the GP algorithm could identify normal and mismatch conditions accurately. The GP regression was able to predict the normal I-V curve well, and therefore determine the potential lost electrical power due to a fault condition. These results are outlined in Section III-B below.

A. Classification

The GP classification algorithm was able to differentiate between normal and mismatch behavior as shown in Figure 3. The I-V characterization data, represented on the left side of Figure 3, for each instance was presented to the GP classification algorithm. The algorithm was able to classify

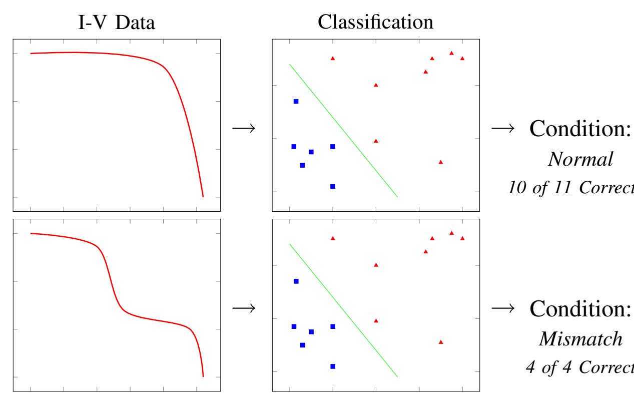

Fig. 3: The classification of the I-V characterization data was performed by a Gaussian Process algorithm. The algorithm evaluated the I-V curve data at a particular instances and classified the data as normal or a fault condition.

normal and fault behavior by seperating the data into classes as depicted in the middle of Figure 3.

The experiment used 120 data points that contained normal, mismatch, and complete shading scenarios. The training data samples were labeled with a 0 for normal, 1 for mismatch, and 2 for complete shading. The I-V data and its respetive labels were presented to the classification algorithm for training. Then, 15 data points that were previously unseen were presented to the algorithm for testing. The algorithm was able to classify 10 out of 11 normal conditions correctly and 4 out of 4 mismatch conditions as shown in Figure 3. In conjunction with the classification process a GP regression algorithm estimated the I-V curve to calculate the lost power due to the fault.

B. Lost Power Production

The lost power production estimate was based on the difference between the actual I-V curve and the GP regression algorithm estimate. The GP algorithm estimate, which considered module temperature and solar irradiance, was found to be accurate when compared to the I-V curve found during normal conditions. In the present work, 60 I-V traces, irradiance, and module temperatures were presented to the algorithm for training. Then, during testing the algorithm was able to accurately estimate normal I-V curve behavior as shown in Figure 4. In this example, the estimated curve had a maximum power of 1,382 watts, while the actual was calculated to be 1,363 watts. Initial GP estimation results indicated a significant fit to the actual value. However, the complete paper will review a large data set of estimates and actual curves to define the accuracy of the GP regression algorithm. For example, the final paper will report on the root mean square error and the mean bias error.

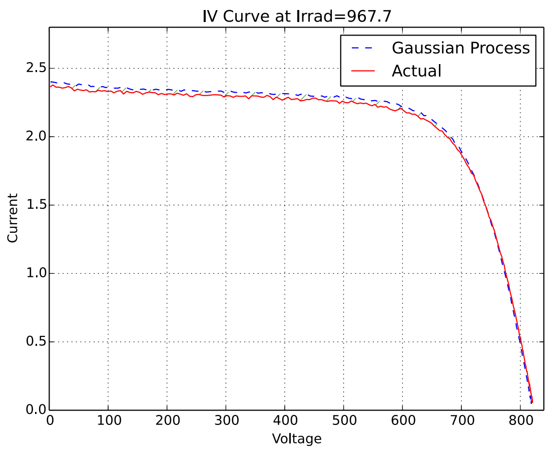

Fig. 4: The I-V produced by the Gaussian Process matched the actual. The max power for the estimated I-V was calculated to be 1,382 watts and the computed actual was 1,363 watts.

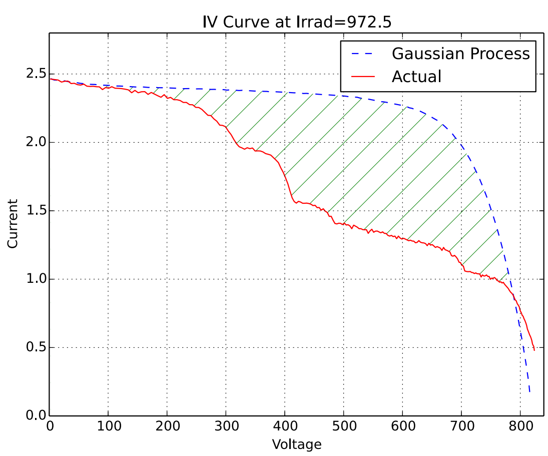

Fig. 5: The estimated I-V curve did not fit the actual because the system was experiencing a mismatch fault condition. The max power was estimated to be 1,423 watts, and the actual was 825 watts.

The estimated I-V curve was also compared with I-V curves produced under fault conditions as shown in Figure 5. The estimated curve was much different than the actual because of the fault condition. This produced a drastic difference in the power produced by the string. The estimated power was calculated to be 1,435 watts and the actual was 610 watts less at 825 watts. This power loss estimate can provide operators with an opportunity to define their maintenance activity in order to address system issues.

IV. Conclusion

The Pordis 140A in-situ I-V tracer system can provide valuable information to evaluate string level performance. However, operators may not have time to look through and review every I-V curve produced. Therefore, an automatic machine learning algorithm such as the GP classifier and regression can provide valuable information. The GP classifier could automatically alert operators of fault conditions. Then, the GP regression estimate can help operators prioritize maintenance activity by defining the lost power production. The proposed methodology can provide valuable information quickly and accurately to improve the overall reliability of a PV system.

Acknowledgment

This work was supported by the U.S. Department of Energy SunShot Initiative

Sandia National Laboratories is a multi-program managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000

References

[1] A. Coleman and J. Zalewski, "Intelligent fault detection and diagnostics in solar plants," in *2011 IEEE 6th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*, vol. 2, Sep. 2011, pp. 948–953.

[2] S. Silvestre, A. Chouder, and E. Karatepe, "Automatic fault detection in grid connected PV systems," *Solar Energy*, vol. 94, pp. 119–127, Aug. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X13001849

[3] C. B. Jones, J. S. Stein, S. Gonzalez, and B. H. King, "Photovoltaic system fault detection and diagnostics using laterally primed adaptive resonance theory neural network," New Orleans, LA, USA, Jun. 2015.

[4] *Photovoltaics: Fundamentals, Technology and Practice*, 1st ed. Chichester, West Sussex, United Kingdom: Wiley, Feb. 2014.

[5] J. E. Quiroz, J. S. Stein, C. K. Carmignani, and K. Gillispie, "In-situ module-level I-V tracers for novel PV monitoring," in *Photovoltaic Specialist Conference (PVSC), 2015 IEEE 42nd*, Jun. 2015, pp. 1–6.

[6] "Pordis LLC 140a String-Level I-V Curve Tracer Specification Sheet." Pordis LLC, Tech. Rep., Jan. 2016. [Online]. Available: http://www.pordis.com/

[7] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, Aug. 2012.

[8] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. University Press Group Limited, Jan. 2006.

[9] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, Mar. 2000.