Photovoltaic System Fault Detection and Diagnostics using Laterally Primed Adaptive Resonance Theory Neural Network

C. Birk Jones, Joshua S. Stein, Sigifredo Gonzalez, and Bruce H. King

Sandia National Laboratories, Albuquerque, NM, 87185, U.S.A

***Abstract*—Cost effective integration of solar photovoltaic (PV) systems requires increased reliability. This can be achieved with a robust fault detection and diagnostic (FDD) tool that automatically discovers faults. This paper introduces the Laterally Primed Adaptive Resonance Theory (LAPART) artificial neural network to perform this task. The present work tested the algorithm on actual and synthetic data to assess its potential for wide spread implementation. The tests were conducted on a PV system located in Albuquerque, New Mexico. The system was composed of 14 modules arranged in a configuration that produced a maximum power of 3.7kW. The LAPART algorithm learned system behavior quickly, and detected module level faults with minimal error.**

***Index Terms*—photovoltaic faults, fault detection, fault diagnostics, artificial neural network, laterally primed adaptive resonance theory**

I. Introduction

Solar energy provides 0.3% of the total energy consumed in the U.S. However, the total on-grid photovoltaic (PV) capacity nearly doubled in 2011 [1]. In 2014 alone, PV systems contributed about 0.43 quads of energy [2]. These systems are typically equipped with fault protection and isolation devices. However, faults such as ground fault, line-to-line faults, arc faults, shading, and hot spot formation can occur undetected [3]. In many cases these faults create hazardous, damaging, or inefficient conditions. Implementing reliable and automatic fault detection and diagnostics (FDD) tools will not only mitigate safety concerns, but also improve the operations and maintenance costs associated with PV systems.

Related literature has tested various FDD tools, including rule-based expressions [4], decision trees [5], and feed-forward neural networks [6]. Past research efforts have also investigated automatic monitoring and FDD of systems through remote communications [7]. The present work introduces a Laterally Primed Adaptive Resonance Theory (LAPART) neural network algorithm that is designed to detect and diagnose PV faults automatically. The LAPART neural network is a unique learning algorithm that can learn system behavior quickly and effectively [8].

The LAPART neural network can act as a FDD tool by first learning from defined set of data to develop knowledge of system behavior. When the learning is complete, the knowledge is used to evaluate previously unseen data. During this testing phase, the algorithm determines whether the behavior is normal or if there is a fault condition. The intent of this paper was to expose the algorithm to data from an actual PV system and define its ability to detect faults throughout the month of May 2015.

The PV system used in the present work is composed of two strings arranged in a parallel configuration. Each module can produce a maximum power of 200W and the entire system provides about 3.7 kilowatts (kW) of electricity. Similar to other PV arrays, the modules used in the present work are susceptible to failures that can be caused by shading, cell damage, diode failures, etc. Module failures cause a mismatch in the string that reduces the voltage and current thus reducing the overall power output. These faults can also lead to expedited module degradation. For example, hot spots, which degrade the integrity of module performance, can be created in a partial shading situation [9]. Therefore, quick and accurate identification of mismatch issues can help to maintain PV system performance.

The present work used actual data that was collected over a 4 day period to train the LAPART algorithm. This data was categorized as normal system behavior and the algorithm learned how to detect fault conditions. After training was complete the algorithm was presented with a single day of normal data and 3 fault data points. This initial test was successful, however the amount of data used was not statistically significant to prove that the algorithm could provide accurate results over an entire month. Therefore, synthetic data, produced by a PV simulation package called PV_LIB (available at the pvmc.sandia.gov website), was calibrated and used to produce 30 days of normal and fault data points. This data was presented to the LAPART algorithm to evaluate its performance over a statistically significant data set that represented the entire month of May.

II. Methodology

PV systems can experience faults that often go unnoticed. These faults decrease electrical power output as well as degrade module properties. Real-time identification of faults that is accurate and reliable can improve overall operations. The present work investigated the potential detection of a module failure in an array using the LAPART neural network algorithm. The algorithm does not require any knowledge of the system's physical properties. Instead it requires historical or past data to learn system performance characteristics. Therefore, two experiments were conducted that used (1) actual data and (2) component-based modeled data from a 3.7kW system that is currently operational in Albuquerque, New Mexico to train and test the algorithm.

The two experiments evaluated the accuracy of the LAPART algorithm to detect faults within intermittent and smooth PV behavior. The accuracy of the LAPART algorithm depends on its ability to provide a high probability of detection while maintaining a low rate of false alarm. The probability of detection considers the number of true positive produced by the FDD process in relation to the total number of actual positive values as described by Eqn. 1.

$$\text{Prob. of Detection} = \frac{\text{True Positive}}{\text{True Positive + False Normal}} \quad (1)$$

$$\text{Prob. of False Alarm} = \frac{\text{False Positive}}{\text{False Positive + True Normal}} \quad (2)$$

The probability of false alarm (Eqn. 2) is based on the number of false positives compared to the total number of normal data points.

The intent of the present work is to describe the probability of detection and false alarm for the LAPART algorithm based on weather conditions in the month of May. The significance of the probability computations is based on the sample size of the test data. To produce FDD results that had a confidence interval of 95% and a margin of error less than 5%, a sample size of 465 fault points would be required. The first experiment, which used measured data from the actual system, had a total of 3 faults. A total of 3 data points was considered a statistically insignificant sample size. Experiment 2, however, had over 10,000 fault data points, and was considered statistically significant.

A. System Layout & Data Collection

The system was composed of two strings, each with seven modules, arranged in a parallel configuration as shown in Fig. 1. The modules were mounted facing south at tilt angle of 35°. Each of the modules had a maximum power of 200 watts, a short circuit current of 3.83 amps, an open circuit voltage of 67.7 volts, maximum power current of 3.59 amps, and a maximum power voltage of 55.8 volts. The entire system produced around 3.7kW with a maximum power current and voltage of 7.6 amps and 480.9 volts respectively.

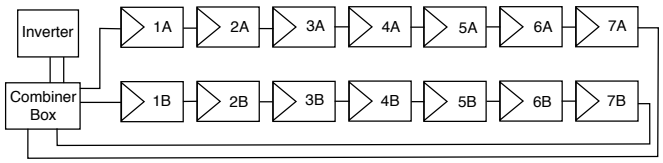

Fig. 1: PV array composed of two strings in a parallel configuration. Each of the strings has seven modules. The maximum power output is about 3.7kW.

The amount of data that accumulated throughout a single day was overwhelming for a person to manage and process on one's own. Therefore, monitoring and flagging anomalies in an accurate and reliable manner required standard data logger software capable of collecting data points in one minute intervals and applying the LAPART algorithm.

B. Laterally Primed Adaptive Resonance Theory

Artificial neural networks (ANN) are a form of machine learning that function like a simplified version of an animal's nervous system to acquire and store knowledge. ANNs can learn system behavior during a training process. Then the algorithms can evaluate new data and provide system performance predictions. The predictions can be generalized, which means that the ANN can provide reasonable outputs for inputs not encountered during training. The ANN can solve complex problems such as linear and nonlinear systems. One such algorithm is the LAPART neural network.

The LAPART algorithm was introduced by Healy and Caudell for logical inference and supervised learning [8]. The algorithm can be used as a prediction tool and has been shown to provide accurate weather forecasts [10]. It has also been applied successfully to solar micro-forecasting to predict solar irradiance [11]. The LAPART algorithm can converge rapidly towards a clear solution because it does not depend on the gradient descent method that is used in many popular algorithms such as the multi-layer perceptron. The gradient descent approach is susceptible to issues that include slow and/or incorrect convergence to the optimal solution [12].

The LAPART architecture couples two Fuzzy Adaptive Resonance Theory (ART) algorithms to create a mechanism for making predictions based on learned associations. The underlying equations for the single Fuzzy ART algorithm include category choice (Eqn. 3), match criterion (Eqn. 4), and learning (Eqn. 5) [13]:

$$C_j = \frac{|\boldsymbol{I} \wedge \boldsymbol{w}|}{\alpha + |\boldsymbol{w}|} \quad (3)$$

$$\frac{|I \wedge w|}{|I|} \geq \rho \quad (4)$$

$$w^{new} = I \wedge w^{old} \quad (5)$$

The intent is to develop the best template ($\boldsymbol{w}$) matrix that represents the input data set. The algorithm uses Eqn. 3 to find the existing w that the given input (I) best matches. Also, the free parameter α is often set to 10^{-7} for fast learning applications. Then Eqn. 4 checks to see if the I and w being compared meet the given vigilance parameter (ρ) criterion that is defined by the user. The vigilance free parameter can vary from 0 to 1 depending on the degree of complexity desired. For instance, a high vigilance parameter of 0.9 provides high complexity and low generality, while a low parameter of 0.5 provides the opposite. Finally, if it passes, then the template, w, is updated based on Eqn. 5.

The coupling of the two Fuzzy ARTs to create the LAPART algorithm is described graphically in Fig. 2. The A and B Fuzzy ARTs are connected through the L matrix that associates the A and B templates. Each Fuzzy ART has its respective vigilance parameters ρA and ρB, and during the learning process inputs are presented to the A and B side simultaneously. The A and B side create and update templates while at the same time producing links between one another.

After training is complete, testing inputs are only applied to the A side and allowed to resonate with the previously learned templates. Then the associations in the L matrix are used to connect with the B side and provide the prediction outputs.

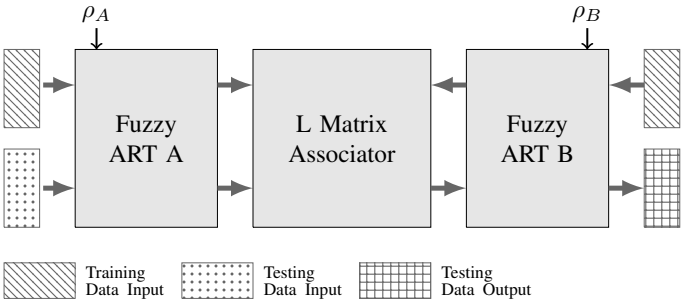

Fig. 2: LAPART algorithm training uses two Fuzzy ART (A&B) algorithms connected by an associator matrix (L). During training inputs are applied to both the A and B sides. The algorithm produces A and B templates. It also produces an L matrix that link the templates in the A and B side to one another. During testing the B side learning is turned off and only A side inputs are applied. The inputs resonate with the stored weights in the A, and through an association in the L propagate to the B side template that provides the prediction output.

C. Experiment 1: Actual Data

The actual data experiment used sensor data from the field as inputs into the LAPART algorithm. Training of the LAPART algorithm involved the presentation of four days of one minute data from April 29, 2015 to May 2, 2015 as shown in Fig. 3. The LAPART algorithm used the training data to gain knowledge of the system. The knowledge was then stored as memory and accessed during the testing process. The testing

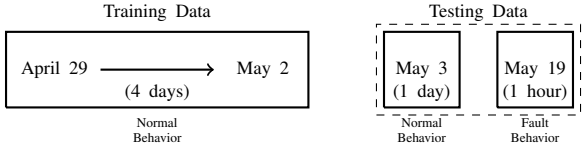

Fig. 3: Actual data experiment includes four days, at one minute intervals, of training data from April 29, 2015 to May 2, 2015. The testing data included normal and fault data points collected on May 3, 2015 and May 19, 2015 respectively. Data collected on May 3, 2015 contained all normal data points, and May 19 data had normal data as well as 3 fault conditions.

process involved the presentation of previously unseen data. In the present experiment, normal behavior data on May 3, 2015, and fault behavior on May 19, 2015 were presented to the trained algorithm. The LAPART algorithm then determined if the particular data instance was normal or a fault condition.

The testing data that was measured on May 19, 2015 included a fault condition. In this case, PV module 2A was completely covered with an opaque material. This caused a mismatch in the modules connected in series. A mismatch of PV modules in series is often due to a non-uniform distribution of irradiance or temperature [4]. In this case, module 2A was completely covered, reducing the overall DC power, current, and voltage. This fault condition was duplicated in the component-based model used in Experiment 2.

D. Experiment 2: Component-based Model Data

A component-based model of the PV array was created to represent system components and provide 30 days of one minute data for both normal and fault conditions. The model outputs allowed the experiment to perform a statistically significant review of the LAPART algorithm's abilities to detect anomalies during the month of May. The model accepted

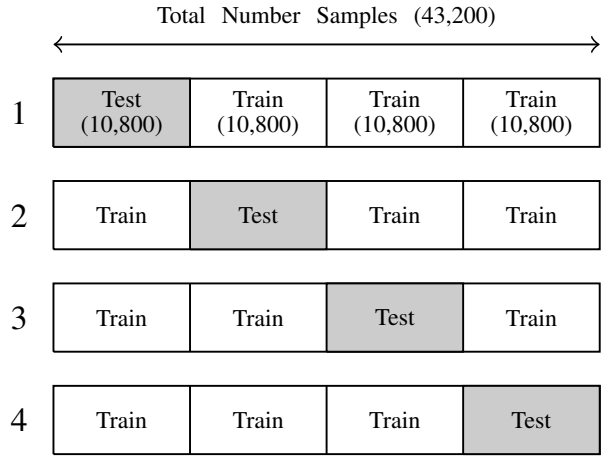

Fig. 4: K-Fold method where K = 4. The method splits the modeled normal and fault data into training and testing sections or folds. Throughout the process the method loops through the different folds so that each data point is eventually used for both training and testing.

actual weather data collected throughout the month of May and included solar irradiance, ambient temperature, and wind speed. The fault, simulated in the model data was the same condition evaluated in Experiment 1. It represented the case where a single module failed to produce a desired voltage or current. The development of the component-based model involved a calibration that considered the three outputs: DC voltage, DC current, and DC power. The output values where compared with actual over the five day period from April 29 to May 3, 2015. Similar to Riley and Venayagamoorthy, the performance of the model was based on the coefficients of determination (R^2) and the evaluation of the intercept and slope of the linear fit equation [14].

The results for the entire 30 day period produced by the component-based model were used as inputs to train and test the LAPART algorithm. In this case a review of the algorithm over multiple vigilance scenarios was conducted to define the probability of detection and false alarm for different parameter scenarios. The process was performed using the K-Folds method. The K-Folds method is a common form of parameter tuning and was used successfully by Duan *et. al* to implement a support vector machine algorithm [15].

The K-Folds process began with randomly distributing 10,000 fault condition data points throughout the data set; the

data was then split into K equal parts or folds. This division of the data for K = 4 is shown in Fig. 4. For each fold k $\in$ {1,2,..,K} the model was trained on the data that was located in all of the folds except for the kth. Then the algorithm used the data in the kth fold for testing [16]. This process was conducted in a round-robin manner until each of the folds was used for training and testing. The probability of detection and false alarm was computed for each free parameters scenario. Results from this experiment and Experiment 1 describe the effectiveness of the LAPART algorithm to perform FDD of a single module failure within an array.

III. Results

The present work performed two experiments that evaluated the ability of the LAPART algorithm to detect a single module failure in a PV system. The first experiment used five days of actual data and the second used a month of modeled data. The 30 days of modeled data was produced from a calibrated component based model. The actual data was collected over a five day period from April 28 to May 3, 2015.

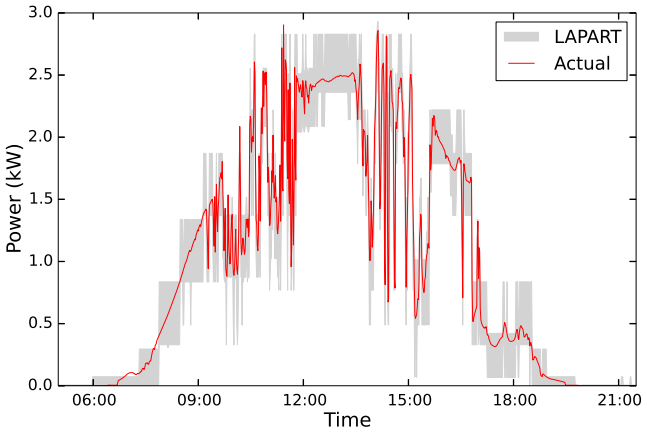

Fig. 5: May 3, 2015 normal behavior testing results show that the LAPART algorithm was able to represent the PV system power output accurately during intermittent activity. The accurate prediction provides for a false detection rate of zero.

A. Experiment 1: Fault Detection Results

The results form the actual data experiment were broken out into normal and fault behavior tests. Accurate identification of normal behavior required that the LAPART algorithm understand smooth and intermittent behavior caused by cloud cover. Fortunately, the LAPART algorithm was able to predict normal behavior well and did not flag any false alarms.

The measured data test began with the training of the LAPART algorithm using 5,760 data points for the four day period. Then, the algorithm was applied to 1,440 normal behavior data points collected on May 3, 2015. The output from the algorithm described a predicted range that is shown in Fig. 5. The actual power output consistently fell inside the predicted range. The data point that fell outside of the range resonated with the stored memory and thus were still considered normal behavior.

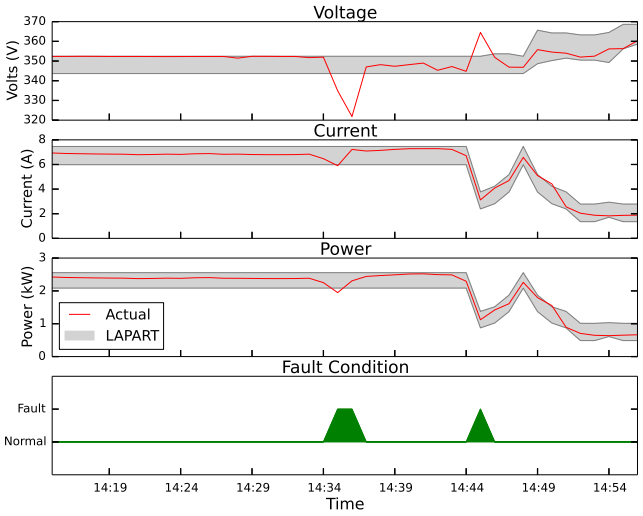

Fig. 6: May 19, 2015 normal and fault behavior. The fault conditions where caused by the introduction of an opaque material placed over module 2A. The normal LAPART algorithm provided a prediction of normal behavior and deviations from the range define the fault conditions.

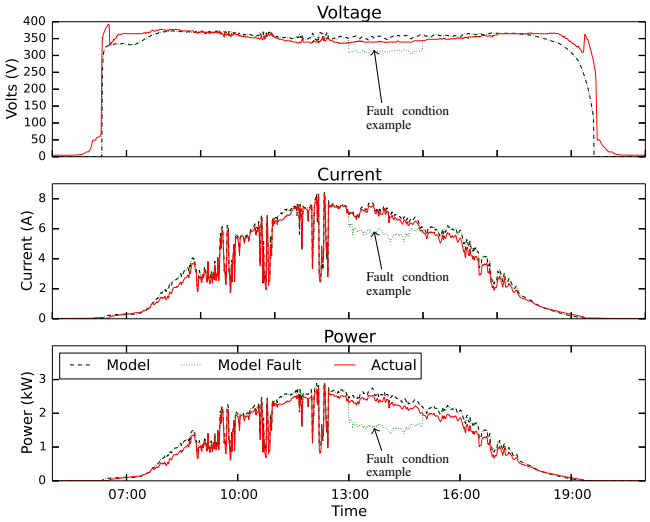

Fig. 7: Component-based model in comparison to the actual output of the PV array for May 3, 2015. The actual current and power outputs match, however the voltage prediction does not represent actual as well.

The fault conditions, created within PV module 2A, can be observed in the data as shown in Fig. 6. The actual voltage, current, and power all dropped at time 14:35. The voltage also spiked at time 14:45. The trained algorithm was presented with this data and it correctly defined the normal and fault behavior as shown in the bottom graph of Fig. 6. The current and power

dropped significantly due to cloud cover at time 14:44, and was correctly recognized by the LAPART algorithm as normal behavior. Each of the tests conducted on May 3 and May 19, 2015 produced no false positives or false negatives. Therefore, the probability of false alarm and detection were equal to 0% and 100% respectively.

B. Experiment 2: Model & Fault Detection Results

The second experiment used data produced by a component-based model. The model was able to represent actual operations well as described in Fig. 7. Fig. 7 compares the voltage, current, and power outputs for the model and actual system on May 3, 2015. The model was able to predict current and power with a high accuracy, but was not able to represent voltage as well. Additionally, Fig. 7 plots the fault condition simulated in the model. Similar to the actual sensor data, the model experienced a drop in voltage, current, and power when the fault occurred. The accuracy of the model is defined in more detail within the scatter plots shown in Fig. 8, 9, and 10 for voltage, current, and power respectively. The distribution of the voltage scatter plot did not match well with the ideal y=x line and produced a low R^2 of 0.47 as shown in Fig. 8. The current and power scatter plots (Fig. 9 and 10) described a sufficient match each with an R^2 equal to 0.99. In addition the linear fit line for each matched very close to the ideal y=x line. The current and power results had an intercept close to zero at or below 0.19 and a slope that was close to 1 at 0.98 and 1.03 respectively.

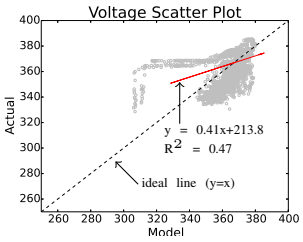

Fig. 8: Actual versus modeled voltage where the linear fit was had a calculated R^2 equal to 0.47 and a fit that did not match well with the ideal y=x line.

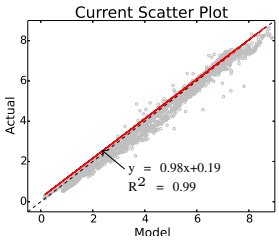

Fig. 9: Actual versus modeled current where the linear fit was had a calculated R^2 equal to 0.99 and a fit that did match well with the ideal y=x line.

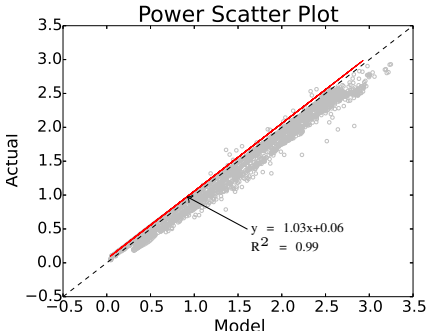

Fig. 10: Actual versus modeled power where the linear fit was had a calculated R^2 equal to 0.99 and a fit that did match well with the ideal y=x line.

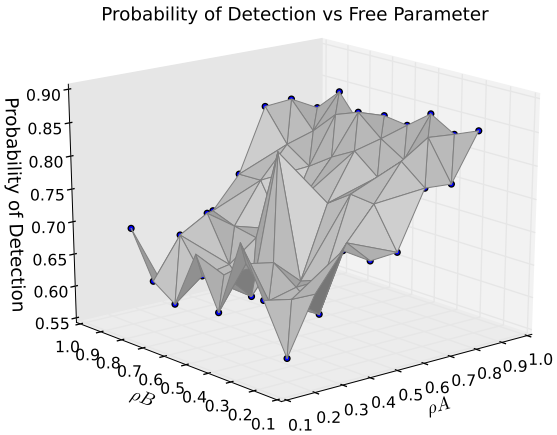

Fig. 11: The probability of detection was greater than 80% for vigilance scenarios where ρA was greater than 0.8. Additionally, the probability did not decrease for the different ρB values.

The voltage output produced by the model was not used as an input into the LAPART algorithm because of the lack of correlation between actual and modeled data. Therefore, the current and power normal and fault condition outputs were evaluated by the LAPART algorithm only. This meant that the solar irradiance, wind speed, ambient temperature, current, and power were inputs on the A-side of the LAPART algorithm. The B-Side inputs and outputs were the PV array condition which was either normal or a fault.

The K-Folds method was used to train and test on a total of 43,200 data points. The data set contained randomly intermixed normal and fault condition data. The LAPART algorithm trained on this data set to learn normal and fault behavior. Then, during testing the LAPART algorithm classified the new data as either normal or as a fault. The probability of detection and false alarm results are described in Fig. 11 and Fig. 12 respectively. Fig. 11 shows that the probability of detection went up as the A side vigilance increased. The probability of detection reached a very high 85% for ρA values greater than 0.8. Additionally, the probability of detection was maintained across the various ρB values. Fig. 12 shows a decrease in the probability of false alarm as the A side vigilance values increase. The probability of false alarm reached a rate that was less than 10% for ρA vigilance parameter greater than 0.8 at any ρB value. The lowest probability of false alarm was found to be 7%, and the highest probability of detection was a very respectable 86%.

The ρB vigilance parameter does not impact the probability of false alarm or detection. This can be attributed to the fact that the experiment only included one feature with two

potential classifications on the B-side. The feature was the

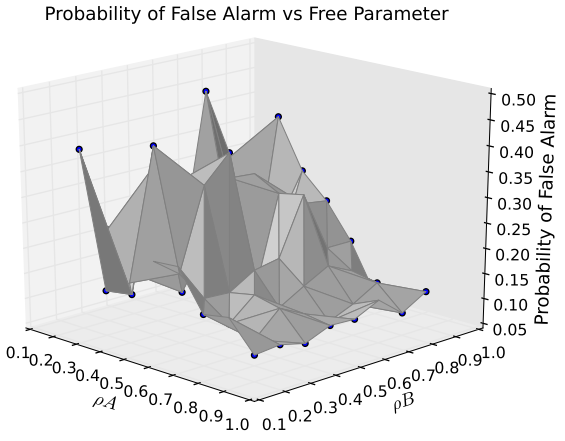

Fig. 12: The probability of false alarm was calculated to be below 10% for ρA vigilance parameters above 0.8. Similar to the probability of detection results the ρB values did not impact probability results.

status of the array which could have a classification of either a fault or normal. The B side vigilance could have a greater effect on experiments that involve more than two potential outputs or with multiple features.

IV. Conclusion

The LAPART algorithm was able to accurately identify a module level fault within the data set produced by both the actual array and a component-based model. The LAPART algorithm was able to interpret both smooth and intermittent normal behavior caused by cloud cover and not signal unnecessary false alarms. In addition, each of the faults within the actual data set were identified by the LAPART algorithm. The second experiment applied the LAPART algorithm to synthetic data produced by the component-based model and produced very good probability of false alarm and detection results. The lowest false alarm rate and highest probability of detection were calculated to be 7% and 86% respectively.

The results showed that the LAPART algorithm can quickly learn PV performance data and provide accurate fault detection results. It only took four days of one minute data to train the algorithm to recognize intermittent behavior as normal. Further studies can expand on this baseline work to consider other fault scenarios at varying environmental conditions. The approach can also be compared with other machine learning techniques such as support vector machines.

Acknowledgment

Professor Thomas Caudell at the University of New Mexico Electrical Engineering Department guided me through the LAPART algorithm coding process.

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] "Utility-scale installation lead solar photovoltaic growth," 2015. [Online]. Available: http://www.eia.gov/todayinenergy/detail.cfm?id=8570

[2] "Renewable & Alternative Fuels," 2015. [Online]. Available: http://www.eia.gov/renewable/

[3] M. Alam, F. Khan, J. Johnson, and J. Flicker, "PV faults: Overview, modeling, prevention and detection techniques," in *2013 IEEE 14th Workshop on Control and Modeling for Power Electronics (COMPEL)*, Jun. 2013, pp. 1–7.

[4] A. Chouder and S. Silvestre, "Automatic supervision and fault detection of PV systems based on power losses analysis," *Energy Conversion and Management*, vol. 51, no. 10, pp. 1929–1937, Oct. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196890410000919

[5] Y. Zhao, L. Yang, B. Lehman, J.-F. de Palma, J. Mosesian, and R. Lyons, "Decision tree-based fault detection and classification in solar photovoltaic arrays," in *2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, Feb. 2012, pp. 93–99.

[6] D. Riley and J. Johnson, "Photovoltaic prognostics and heath management using learning algorithms," in *2012 38th IEEE Photovoltaic Specialists Conference (PVSC)*, Jun. 2012, pp. 001 535–001 539.

[7] 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, no. 4, pp. 548–564, Apr. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X06002040

[8] M. Healy, T. Caudell, and S. Smith, "A neural architecture for pattern sequence verification through inferencing," *IEEE Transactions on Neural Networks*, vol. 4, no. 1, pp. 9–20, Jan. 1993.

[9] E. Meyer and E. Ernest van Dyk, "Assessing the reliability and degradation of photovoltaic module performance parameters," *IEEE Transactions on Reliability*, vol. 53, no. 1, pp. 83–92, Mar. 2004.

[10] M. Healy and T. Caudell, "Acquiring rule sets as a product of learning in a logical neural architecture," *IEEE Transactions on Neural Networks*, vol. 8, no. 3, pp. 461–474, May 1997.

[11] A. Mammoli, A. Menicucci, T. Caudell, A. Ellis, S. Willard, and J. Simmins, "Low-cost solar micro-forecasts for PV smoothing," in *2013 1st IEEE Conference on Technologies for Sustainability (SusTech)*, Aug. 2013, pp. 238–243.

[12] S. Haykin, *Neural Networks A Comprehensive Foundation*, 2nd ed. Prentic-Hall, Inc, 1999.

[13] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, no. 6, pp. 759–771, 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/089360809190056B

[14] D. Riley and G. Venayagamoorthy, "Comparison of a recurrent neural network PV system model with a traditional component-based PV system model," in *2011 37th IEEE Photovoltaic Specialists Conference (PVSC)*, Jun. 2011, pp. 002 426–002 431.

[15] K. Duan, S. S. Keerthi, and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41–59, Apr. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092523120200601X

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