This paper appears in: IEEE Transactions on Sustainable Energy Issue Date: October 2015 Volume: 6 Issue: 4 pages: 1200-1207 DOI: 10.1109/TSTE.2015.2421447

Online Fault Detection in PV Systems

Radu Platon, Jacques Martel, Norris Woodruff, and Tak Y. Chau

***Abstract*—This paper presents the development of a practical fault detection approach in photovoltaic (PV) systems, intended for online implementation. The approach was developed and validated using field measurements from a Canadian PV system. It has a fairly low degree of complexity, but achieves a high fault detection rate and is able to successfully cope with abnormalities present in real-life measurements. The fault detection is based on the comparison between the measured and model prediction results of the ac power production. The model estimates the ac power production using solar irradiance and PV panel temperature measurements. Prior to model development, a data analysis procedure was used to identify values not representative of a normal PV system operation. The original 10-min measurements were averaged over 1 h, and both datasets were used for modeling. In order to better represent the PV system performance at different sunlight levels, models for different irradiance ranges were developed. The results reveal that the models based on hourly averages are more accurate than the models using 10-min measurements, and the models for different irradiance intervals lead to a fault detection rate greater than 90%. The PV system performance ratio (PR) was used to keep track of the system's long-term performance.**

***Index Terms*—Data analysis, fault detection, lagging, normal operation, online implementation, performance ratio (PR), predictive model, photovoltaic (PV) systems.**

I. Introduction

THE photovoltaic (PV) market and technology have shown a rapid growth over the past years, representing today a mature technology for power production from renewable energy sources and a common on-site electricity generation strategy. In Canada only, the installed capacity for solar PV power grew at an annual rate of almost 150% annually over the 2008–2011 period, reaching 495 MW in 2011 [1]. However, the number of monitored PV systems has not followed the same growing trend, as many PV plants, especially smaller ones, are operated without a proper supervision system [2]. A fault detection algorithm for PV systems can provide an accurate estimate of the electricity production under normal operating conditions and detect PV system faults—periods of abnormally low power production, when the system produces significantly less power than it should at the given operating conditions. This would enable operators to take timely corrective actions, in order to prevent the PV system to under-perform for prolonged periods of time, minimize the power losses caused by these faults, and improve the PV system performance. Different factors can be responsible for the production losses of a PV system, such as maximum power point tracking error [3], electrical disconnection [4], wiring losses and ageing [5], shading effects [6], dust or snow accumulation on the surface of the solar panels [7], and faulty power conditioning equipment such as dc–ac converters.

Some of the PV fault detection algorithms reported in the literature are based on electrical-circuit simulations of a PV panel [8]–[10], and some use statistical analysis of different PV system measurements, as well as system efficiency values [11]–[13]. Electrical signal analysis methods, such as time-domain reflectometry, were also used to detect faulty PV strings [14]. Some of these approaches are developed using field data, while a few use meteorological and satellite data [15].

A number of fault detection algorithms are based on the comparison between measured and modeled PV system outputs to identify faults [2], [16]–[18]. Different methods are used to develop predictive models for PV system power production. In some approaches, the electricity production was predicted with parametric models that use PV system and weather variables along with adjustable parameters [19], [20]; other approaches use artificial intelligence techniques, such as neural networks, fuzzy logic, and expert systems [21]–[23]; PV system modeling using commercial simulation packages was also reported [24].

In this work, we present the development of a fault detection procedure that is intended for online implementation. As such, the main objective was the development of a practical and fairly simple (low-complexity) approach that has a high fault detection rate and is robust enough to successfully cope with abnormal values inherently present in real-life data, such as erroneous measurements and lagging between measurements with different sampling rates. The fault detection approach is based on the comparison between the measured and model prediction results of the ac power production: a significant difference between those two values was considered a fault. The predictive model used in this approach was selected such that its development does not require knowledge of advanced modeling methods, artificial intelligence techniques, detailed comprehension of the PV process and electrical circuits necessary to carry out a circuit-based simulation, or familiarity with commercial energy simulation software packages. The model used to predict the PV system electricity production consists of one equation and was selected not only based on its accuracy but also on its low complexity. Prior to model development, a simple but effective data analysis procedure was used to process the measurements, in order to minimize the effect of lagging and

Manuscript received November 17, 2014; revised February 10, 2015; accepted April 07, 2015. This work was supported in part by Natural Resources Canada's Program of Energy Research and Development (PERD). Paper no. TSTE-00619-2014.

R. Platon and J. Martel are with Natural Resources Canada, CanmetENERGY, Varennes J3X 1S6, QC, Canada (e-mail: rplaton@nrcan.gc.ca; jmartel@nrcan.gc.ca).

N. Woodruff and T. Y. Chau are with Cachelan, Markham L3R 0H6, ON, Canada (e-mail: nwoodruff@cachelan.com; tchau@cachelan.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2015.2421447

identify measurements not representative of a normal (fault-free) PV system operation. To enable the generalization of this, the fault detection algorithm to other PV systems, the predictive model development and the calculation of the normal operation limits can be automated.

This approach was developed and validated using historical measurements from a 120-kW PV plant located in Toronto, Canada. The fault detection algorithm was validated with data that include measurements taken during an actual faulty operation of the PV system-a period of time when the inverter was malfunctioning. The model consists of one equation that uses solar irradiance and PV panel temperature measurements to predict the ac power production. In order to determine if this simple model is sufficiently accurate, its predictive performance was compared to that of a neural network model. Two approaches were tested for developing the models: an approach involving separate models for different irradiance ranges, in order to better represent the PV system performance at different sunlight levels, and a global approach over all irradiance values. The measurement sampling rate was 10 min; these measurements were also averaged over 1 h, to determine if this will lead to an improved model accuracy and fault detection rate. The performance ratio (PR) was used to complement the model-based fault detection approach, by keeping track of the long-term performance of the PV system.

Based on the results, recommendations for online implementation were formulated.

II. Data Analysis

The ac power production model should represent the expected output of the PV system under normal operation conditions; therefore, prior to model development, the data should be analyzed to identify and remove from the modeling dataset the measurements representative of a faulty PV system operation. The presence of faulty data can be caused by different factors, such as instrument or equipment malfunction. Lags can also be present in the dataset, due to improper synchronization between sampling times of different measurements.

A. Data Collection

Historical data were collected from a PV system located in Toronto, Canada. The system is mounted on the roof of an institutional building, and it has a dc nominal capacity of 120 kW, generated by 400 Heliene 300W 72-cell panels connected to a KACO XP100 inverter with a rated power output of 100-kW ac. Measurements recorded every 10 min, covering the period of March 15–July 31, 2014, of the solar irradiance in the PV array plane, panel temperature, and ac power output were used to develop the fault detection approach. The irradiance value is obtained by averaging 120 measurements taken every 5 s, and the panel temperature value is obtained by taking one instantaneous measurement every 10 min. The ac power value is obtained using the lifetime energy measurement generated by the inverter—the total energy produced by the PV system since its installation—by calculating the difference between the energy produced after two consecutive 10-min time intervals.

TABLE I
Basic Statistics for the 10-Min Measurements Dataset

Operation mode	Measurement	Min. value	Max. value	Number obs.
Normal	AC power (kW)	0.6	107.4	8 842
	Irradiance (W/m^2)	50.0	1 311.0	
	Module temp. ($^\circ$C)	–14.6	50.6	
Faulty	AC power (kW)	0.6	105.6	776
	Irradiance (W/m^2)	53.0	1 491.0	
	Module temp. ($^\circ$C)	–13.2	46.3	

B. Data Averaging

The 10-min measurements were averaged over 1 h, in order to determine if the hourly average data would reduce the measurement variability and ultimately lead to a model accuracy increase. The same analysis and model development approach was applied to both 10-min and hourly averages datasets.

C. Data Analysis

The first step of data analysis was the elimination of observations corresponding to very low sunlight levels—less than $50\ \mathrm{W/m^2}$—for which the measurement accuracy is significantly reduced; observations corresponding to zero power output were also removed.

Under normal conditions, the solar irradiance and power production measurements follow a strong linear relationship. In order to detect observations not representative of a normal PV system operation, the plot of the ac power as a function of the irradiance was examined. The coefficient of determination R^2, indicating the strength of the linear relationship between the irradiance and power, was calculated-an R^2 value of 1 would indicate that the data perfectly fit a straight regression line. The R^2 values for the 10-min data and hourly averages were 0.78 and 0.91, respectively.

Observations close to the irradiance–power linear relationship line are considered as normal operation measurements, while observations far from this line are considered as faulty operation data. The faulty data correspond to abnormal power production, when the system produces significantly less or more power than it should at the given irradiance level. The normal and faulty operation observations were identified through visual inspection. A mathematical approach that can be automated is currently under development; however, for developing and validating the approach, the visual inspection was deemed sufficient.

The irradiance–power linear relationship is greatly improved when the faulty data are removed: the R^2 coefficient increases from 0.78 to 0.93 and from 0.9 to 0.97 for the 10-min data and hourly averages, respectively. Basic statistics of the 10-min and hourly averages datasets, for both normal and faulty operation, are shown in Tables I and II, respectively. The normal and faulty operation data for the 10-min measurements and hourly averages are shown in Figs. 1 and 2, respectively—the R^2 values indicated on the plots correspond to the normal operation data only.

For the 10-min measurements, the faulty points located above the irradiance–power straight line occur mostly at

TABLE II
BASIC STATISTICS FOR THE HOURLY AVERAGES DATASET

Operation mode	Measurement	Min. value	Max. value	Number obs.
Normal	AC power (kW)	2.1	102.8	1 587
	Irradiance (W/m^2)	50.0	1 120.7	
	Module temp. ($^\circ$C)	−12.9	49.8	
Faulty	AC power (kW)	0.4	84.9	53
	Irradiance (W/m^2)	52.7	1 181.0	
	Module temp. ($^\circ$C)	−12.7	38.4	

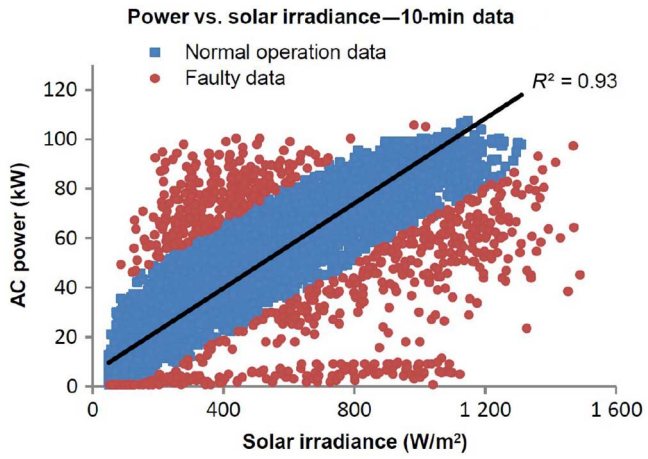

Fig. 1. Power and irradiance values for the 10-min measurements.

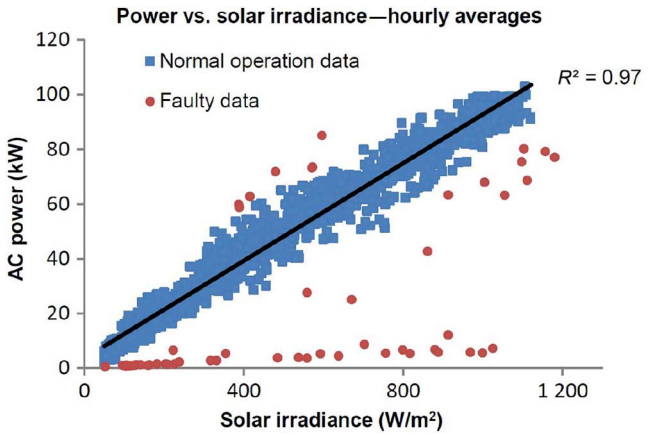

Fig. 2. Power and irradiance values for the hourly averages.

irradiance values under 700 W/m^2. Most of them correspond to abnormally high power production at relatively low sunlight levels, e.g., a power production of 95 kW at an irradiance of 215 W/m^2. These values are evidently erroneous; their number is reduced when the 10-min measurements are averaged hourly, suggesting that they may be caused by lagging between the irradiance and power measurements. If this lagging does not last for a long period, its effect is reduced when averaging 6 consecutive 10-min measurements into one hourly value.

The faulty points located below the line occur at almost all irradiance levels, and most of them are still present when the 10-min measurements are hourly averaged. This suggests that they represent actual PV system faults, when the power production is abnormally low despite relatively high sunlight levels. An inverter problem did affect the PV system output during the month of April, causing most of the points corresponding to power production levels below 10 kW despite high irradiance levels.

The lagging between the irradiance and power values might be caused by the manner in which the ac power measurement is obtained every 10 min from the inverter, as the actual measurement time and the time that the inverter generates the lifetime energy value might not be perfectly synchronized. Also, the measurements are not taken at the same moment and they might be sent to the cloud server storing the data at different times, where they are recorded with the same timestamp; this would falsely indicate that they were obtained simultaneously.

The data analysis also revealed that the resolution of the ac power measurements is 600 W; this relatively high data granularity might represent a modeling challenge, as for different irradiance values, the same power output is measured-since the power is measured in increments of 600 W.

III. MODEL DEVELOPMENT

A. AC Power Model

Models predicting the ac power production of the PV system using solar irradiance and PV panel temperature measurements were developed using both the 10-min and hourly averaged data. The model is based on a parametric approach to model the PV system efficiency. It was used to develop an automated failure detection routine [19]; modified versions of this model also used the develop PV system output forecasts [25]. In the study presented in this paper, a variation in this model was used, as the power production is calculated as follows:

$$P_{\mathrm{ac}} = G\ (a_1 + a_2 G + a_3 \log(G))\ (1 + a_4\,(T_m - 25)) \quad (1)$$

where P_{ac} is the ac power production (W), G is solar irradiance in the PV module plane (W/m^2), T_m is the module temperature ($^\circ$C) and a_1, a_2, a_3, and a_4 are coefficients calculated, so that the model result is as close as possible to the measured data.

B. Training and Validation Errors

Prior to computing the predictive models, the data are separated into training and validation sets. The training dataset is used for developing the model, while the validation dataset is used to validate the model performance. If only the training data are used to determine the prediction performance, the accuracy can be overestimated, since the model is specifically tuned to fit the training data. This is called model over-fitting, and it occurs when a model performs very well on the training dataset but is not able to generalize from the data trend and performs poorly on unseen data-the validation data. The validation dataset is obtained by setting aside a portion of the original dataset that will not be used during the training process. After the model is fitted on the training data, its performance is tested on the validation data.

From the available data, 30% of observations are randomly selected, in a uniformly distributed manner, as validation data. The remaining 70% of observations represent the training dataset and are used to develop the model.

The models were scored in terms of the validation error, using the Coefficient of Variation of the Root Mean Square Error-CV(RMSE); it is calculated as the square root of the average of the squares of the error for each validation observation, normalized to the mean of the measured ac power values from the validation dataset.

C. Normal Operation AC Power Models

Both the 10-min measurements and hourly averages identified previously as being representative of a normal PV system operation were used to develop the models of the ac power production.

Due to the technological quality of the solar modules, the efficiency of a PV system is dependent on the light intensity levels. At low irradiance levels-generally below $300\ \mathrm{W/m^2}$-the efficiency is low; the efficiency increases with the sunlight levels, and remains relatively stable until the irradiance reaches higher values-generally over $900\ \mathrm{W/m^2}$-when the efficiency slightly drops [26], [27]. In order to capture the behavior of the PV system according to sunlight levels, models for different irradiance intervals were developed. This would determine if multiple models corresponding to different sunlight levels lead to a better accuracy than a single global model that covers the whole irradiance range. For both the 10-min and hourly averaged datasets, models for the following solar irradiance intervals were developed: complete irradiance range measurements ($50-1311\ \mathrm{W/m^2}$ for the 10-min data and $50-1120.7\ \mathrm{W/m^2}$ for the hourly averaged data); intervals of $100\ \mathrm{W/m^2}$; intervals of $200\ \mathrm{W/m^2}$; and intervals of 50–250, 250–500, and 500–max. irradiance.

For all irradiance intervals, the models developed using hourly averages outperform the models developed with 10-min measurements, especially at lower irradiance levels. For the global model, covering the complete irradiance range, the error of the model using the hourly averaged data is 33.9% lower than that of the model using the 10-min data-errors of 10.15% and 15.34%, respectively.

The models developed using $100\ \mathrm{W/m^2}$ and $200\ \mathrm{W/m^2}$ irradiance intervals do not significantly outperform the models developed using the intervals of 50–250, 250–500, and 500–max. irradiance; therefore, it was decided that only the global model that covers the complete irradiance range, and the models corresponding to 50–250, 250–500, and 500–max. irradiance will be used for fault detection. The model errors are shown in Table III.

D. Model Performance Benchmark

The one-equation model presented in this section was selected for predicting the ac power production mainly because, in terms of online deployment, its implementation would be less effort-intensive than programming a more complex model. However, in order to determine if this one-equation model reaches sufficient accuracy levels, its predictive performance was measured against the performance of a more complex model.

TABLE III
VALIDATION ERRORS—CV(RMSE)

Irradiance interval ($\mathrm{W/m^2}$)	10-min. measurements (%)	Hourly averages (%)
Complete range—global model	15.34	10.15
50–250	35.87	17.63
250–500	21.47	12.06
500–max. irradiance	9.48	7.25

TABLE IV
VALIDATION ERRORS OF THE ONE-EQUATION AND ANN MODELS

Irradiance interval ($\mathrm{W/m^2}$)	Validation errors		Hidden layer size (neurons)
	One-eq. model (%)	Neural network (%)	
Complete range	10.15	9.17	27
50–250	17.63	16.66	29
250–500	12.06	11.59	9
500–max. irradiance	7.25	6.81	19

Feedfoward neural network models were also developed. Neural networks represent an artificial intelligence-based modeling method that mimics the reasoning of the human brain; they are known for their ability to model complex and highly nonlinear processes. The feedforward configuration is widely used for predictive applications. A description of this method is outside the scope of this report; however, references are abundant in published literature and on the internet.

The predictive accuracy of the neural network model was used as a benchmark for evaluating the performance of the one-equation model.

1) If its predictive performance is comparable to that of the neural network model, its accuracy is deemed satisfactory.
2) If its predictive performance is significantly much poorer than that of the neural network, a more complex model should be considered to better represent the ac power production.

The hourly averages of the solar irradiance and PV panel temperature were used as inputs to the neural network models. The number of neurons in the hidden layer was determined on a trial and error basis: it was varied from 1 to 50 to determine the size of the hidden layer leading to the smallest validation error. The results revealed that setting the maximum number of neurons at 50 was appropriate, as the evolution of the validation error indicates that model overfitting starts before the size of the hidden layer reaches 50 neurons. The predictive errors of both the one-equation and feedforward neural network models developed using hourly averages are shown in Table IV; the number of neurons corresponding to each neural network model is also indicated. It can be seen that although the neural networks are slightly better, the improvement in accuracy is not significant; therefore, it is considered that the accuracy of the one-equation model is satisfactory.

IV. Fault Detection

A. Approach

The ac power models developed using hourly averaged data were more accurate than those using the 10-min measurements. The fault detection is therefore carried out using the hourly average models based on the fault-free operation data described previously. The normal operation limits are calculated using the historical ac power measurements used for model development and their corresponding model prediction results. In an online implementation, the real-time irradiance and panel temperature measurements are used as inputs to the model to calculate a power output; the measured ac power is compared to this calculated value to determine if it lies inside or outside the normal operation limits. Points outside the normal operation limits are considered faults.

The points located far from the irradiance–power straight line, previously identified as representative of a faulty operation, were used to validate this approach: the measured ac power values are compared to their corresponding model results to determine if they are outside the normal operation limits.

B. Normal Operation Limits

The normal operation limits were calculated using the ratio between the measured and modeled ac power for the normal operation data used to develop the model. The limits are calculated as

$$\text{Lower limit} = \mu - 3\sigma \tag{2}$$

$$\text{Upper limit} = \mu + 3\sigma \tag{3}$$

where μ and σ are the average and standard deviation, respectively, of the values of the ratio *measured ac power/modeled ac power* over the training dataset. This ratio is used to determine how close the measurements are to their expected values, according to the model: the closer this ratio is to 1, the closer the measured power is to the modeled value.

The plus and minus 3 standard deviations ($\pm 3\sigma$) interval was chosen to calculate the normal operation limits. If the data are normally distributed, about 99.7% of the points will lie within $\pm 3\sigma$, and it is considered that values outside this interval do not follow the statistical distribution of the bulk of the data. Even if the data are not normally distributed, when using three standard deviations, at least 88.9% of the observations fall in the $\pm 3\sigma$ interval.

Normal operation limits were calculated for the global model, covering the complete irradiance range, and for the models corresponding to $50-250, 250-500,$ and $500-1120.7\ \mathrm{W/m^2}$ intervals; their values are shown in Table V.

C. Fault Detection

The irradiance and PV panel temperature measurements corresponding to faulty data were used as inputs to the model, and the ratio between the modeled and measured power was calculated. The values of this ratio were then compared to the normal operation limits to determine if faults are detected. There were 53 faulty points; the global model successfully identified 43 faults, while the models for the solar irradiance intervals of $50-250, 250-500,$ and $500-1120.7\ \mathrm{W/m^2}$ successfully identified 51 faults. The fault detection rates were 81.13% and 96.23%, respectively; the fault detection rate obtained using models developed for different solar irradiance intervals is almost 16% superior to that of the global model that covers the complete irradiance range. At higher irradiances, those models have narrower normal operating limits, and are more effective in detecting faults that are relatively close to the normal operating limits. This is shown in Figs. 3 and 4, where a group of faults occurring at irradiance values higher than $900\ \mathrm{W/m^2}$ is outside the normal operation limits corresponding to the models developed for different solar

TABLE V
Normal Operation Limits

Irradiance interval ($\mathrm{W/m^2}$)	Lower limit	Upper limit
Complete range—global model	0.5716	1.4132
50–250	0.4351	1.5616
250–500	0.6234	1.3656
500–max. irradiance	0.7644	1.2327

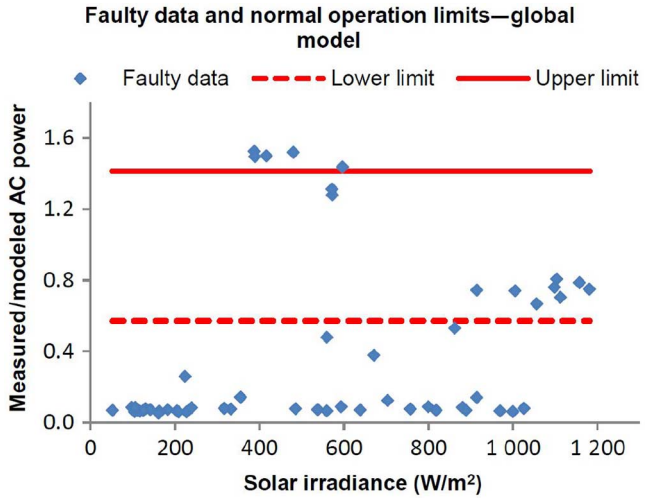

Fig. 3. Faults not detected by the global model.

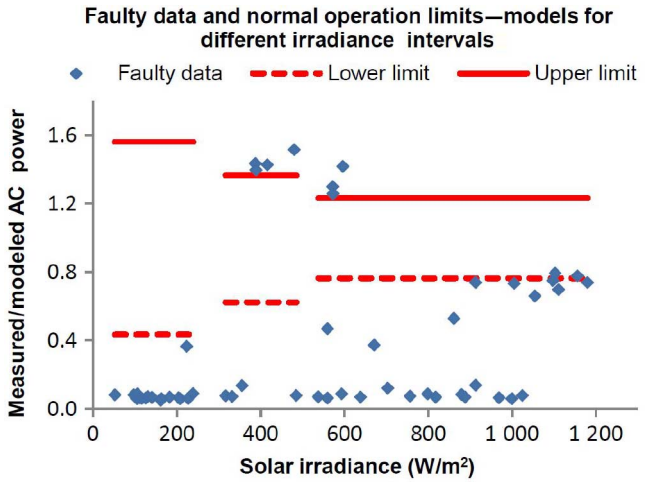

Fig. 4. Faults detected by the models for different irradiances.

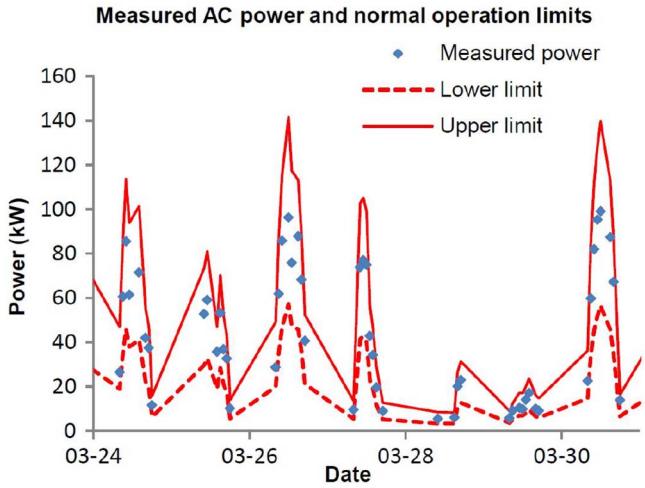

Fig. 5. Normal operation limits expressed in kW (data source: www.gcc.solarvu.net).

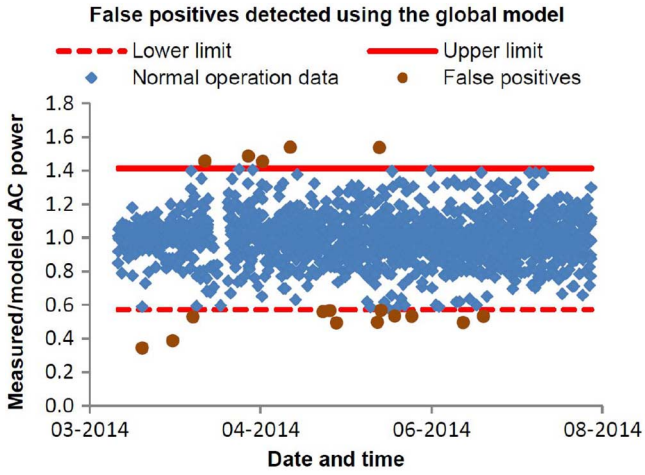

Fig. 6. False positives detected by the global model.

irradiance intervals, and inside the limits calculated using the global model. Therefore, these faults are detected by the models developed for different solar irradiance intervals, but they are not detected by the global model.

For every ac power output measurement, the normal operation limits can also be expressed in terms of kW by taking their values expressed in terms of the ratio *measured ac power/modeled ac power*, and multiplying them with the power output calculated by the model. An example of the measured ac powers and their corresponding limits calculated using the global model and expressed in terms of kW is shown in Fig. 5.

D. False Positives

A false positive represents a false alarm, i.e., a fault is detected when, in reality, the system is operating normally. There are 1585 normal operation points in the dataset; the global model detected 17 false positives, while the models for the irradiance intervals of $50-250$, $250-500$, and $500-1120.7\ \mathrm{W/m^2}$ detected 12. The false positives detected by the global model are shown in Fig. 6.

The occurrence of false positives cannot be completely eliminated, as the model is not 100% accurate. By considering that multiple consecutive points outside the limits represent a fault, the number of false positives can be reduced. This strategy increases the robustness of the fault detection system, as brief low power production periods will not be falsely considered as faults. This will impact the fault detection rate as well; however, since a real PV system fault is likely to last for at least a few hours, this should not have a significant negative impact on the fault detection rate.

An analysis of the false positives from the normal operation data revealed that no two consecutive hourly values are outside the normal operation limits; however, this value might change in time as more data become available.

V. Long-Term PV System Monitoring Using the PR

The PR evaluates the efficiency of the PV system by comparing its real performance to the ideal performance at standard test condition (STC) of $1000\ \mathrm{W/m^2}$ solar irradiance and 25°C PV module temperature. The ideal performance is represented by the rated dc power output of the system, while the real performance is calculated using the system's output over a specified period of time and the amount of sunlight received over the same time period. The PR is calculated as follows:

$$\mathrm{PR} = \frac{\text{total ac (kWh)}}{\text{rated dc (kW)} \times \text{total insolation (Wh/m}^2\text{)}/1000\ \mathrm{W/m^2}} \tag{4}$$

where $\text{total ac} = \text{ac energy (kWh)}$ produced over the time interval for which the PR is calculated, $\text{rated dc} = \text{dc power at STC (kW)}$ and $\text{total insolation} = \text{total solar energy}$ received by the PV system over the interval for which the PR is calculated ($\mathrm{Wh/m^2}$). The PR value is calculated over a specific period of time, and it is used to track the PV system performance over time. The closer the PR value for a PV system approaches 1, the closer the system production is to its ideal, rated power production.

Monthly PR values were calculated, and it was seen that the PR value for the April is lower than those for the other months; this is caused by an inverter malfunction that occurred in April. By calculating weekly PRs for the month of April, we can further identify the time period when the fault occurred: the second week of April, which has a PR value almost 35% lower than the average of the other 3 weeks. The PR values are shown in Table VI.

The PR can be a useful tool for monitoring system performance; it does not require a model, only measured data. Although it can be used for fault detection, faults can only be detected after the time period for which the PR value is calculated; for example, in the case of monthly PR values, a month will pass before detecting that there is a problem with the PV system. The PR value can complement the model-based method for detecting PV system faults, as the model detects faults in real time, and the PR value keeps track of the PV system performance, as it degrades over time.

TABLE VI
PR VALUES

Time period	Performance ratio (%)
March 15–31	82.91
April	74.44
May	80.36
June	79.31
July	79.61
April first week	84.69
April second week	53.89
April third week	79.52
April four week	81.51

VI. CONCLUSION: RECOMMENDATIONS FOR ONLINE IMPLEMENTATION

This section contains the conclusion of the study presented in this paper. The approach that will be used to implement online the fault detection algorithm is based on this conclusion. The future work required to generalize the algorithm to other PV systems is also presented.

Hourly averages of the measurements should be used for an online implementation, since the ac power models developed using hourly averages are more accurate than the models developed using 10-min measurements. The models for irradiance intervals of $50-250, 250-500,$ and $500-1120.7\ \mathrm{W/m^2}$ lead to superior accuracy and fault detection rates than the global model that covers the whole irradiance range. Consequently, the models based on hourly averages and developed for the three different irradiance intervals will be used for the online implementation.

Developing a practical fault detection system that has a relatively low degree of complexity, is able to deal with erroneous measurements and abnormal values, and achieves a high level of accuracy was the main requirement of this study. Despite the low complexity of the ac power production model, the predictive accuracy is quite high—the model for irradiance values above $500\ \mathrm{W/m^2}$ has a validation error of 7.25%; the fault detection rate is better than 90%. The fault detection accuracy and robustness to abnormal measurements of this approach will be further validated once it is implemented online. The model accuracy can be improved by lowering the 600-W measuring resolution of the ac power production. Using a watt meter, instead of the inverter, to measure the power output, and the same sampling rate as for the solar irradiance, would probably eliminate most of the lagging between the power and irradiance measurements.

The predictive model and fault detection algorithm are site-specific, meaning that they are based on historical measurements of the PV system to be monitored. In order to enable the generalization of the approach to other PV systems, future work will include the automation of the data analysis procedure that identifies, prior to developing the predictive model, erroneous values, and measurements related to faulty operation. For example, limits for identifying observations not representative of a normal PV system operation can be automatically determined by calculating the distance between each observation and the irradiance–power straight line and iteratively eliminating observations with a high distance until a satisfactory linear relationship is achieved. Future work also includes the development of fault diagnosis rules, once the model is implemented online and its accuracy and robustness is validated. The specific causes of the faults would be identified, providing operators with valuable information regarding the required corrective actions.

ACKNOWLEDGMENT

The authors wish to thank S. Pelland, Y. Poissant, D. Turcotte, and V. Delisle from CanmetENERGY-Varennes for their guidance on PV system modeling.

REFERENCES

[1] Natural Resources Canada. *About Renewable Energy* (2014) [Online]. Available: http://www.nrcan.gc.ca/energy/renewable-electricity/7295#solar

[2] A. Chouder and S. Silvestre, "Automatic supervision and fault detection of PV systems based on power losses analysis," *Energy Conver. Manage.*, vol. 51, pp. 1929–1937, 2010.

[3] P. Lei, Y. Li, Q. Chen, and J. E. Seem, "Extremum seeking control based integration of MPPT and degradation detection for photovoltaic arrays," presented at the Amer. Control Conf., Baltimore, MD, USA, 2010.

[4] T. Takashima *et al.*, "Experimental studies of fault location in PV module strings," *Sol. Energy Mater. Sol. Cells*, vol. 93, pp. 1079–1082, 2009.

[5] T. Nordmann, U. Jahn, and W. Nasse, "Performance of PV systems under real conditions," presented at the Eur. Workshop Life Cycle Anal. Recycl. Sol. Modules Waste Challenge, Brussels, Belgium, 2004.

[6] D. Sera, R. Teodorescu, and P. Rodriguez, "Partial shadowing detection based on equivalent thermal voltage monitoring for PV module diagnostics," presented at the 35th Annu. IEEE Ind. Electron. Conf., Porto, Portugal, 2009.

[7] J. Zorrilla-Casanova *et al.*, "Analysis of dust losses in photovoltaic modules," presented at the World Renew. Energy Congr., Linkoping, Sweden, 2011.

[8] K.-H. Chao, S.-H. Ho, and M.-H. Wang, "Modeling and fault diagnosis of a photovoltaic system," *Elect. Power Res.*, vol. 78, pp. 97–105, 2007.

[9] D. Guasch, S. Silvestre, and R. Calatayud, "Automatic failure detection in photovoltaic systems," presented at the 3rd World Conf. Photovoltaic Energy Convers., Osaka, Japan, 2003.

[10] M. Hamdaoui, A. Rabhi, A. El Hajjaji, M. Rahmoun, and M. Azizi, "Monitoring and control of the performances for photovoltaic systems," presented at the Int. Renew. Energy Congr., Poznan, Poland, 2009.

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

[12] Y. Yagi *et al.*, "Diagnostic technology and an expert system for photovoltaic systems using the learning method," *Sol. Energy Mater. Sol. Cells*, vol. 75, pp. 655–663, 2003.

[13] Y. Zhao, B. Lehman, R. Ball, J. Mosesian, and J.-F. de Palma, "Outlier detection rules for fault detection in solar photovoltaic arrays," presented at the 28th Annu. IEEE Appl. Power Electron. Conf. Expo. (APEC), Long Beach, CA, USA, 2013.

[14] T. Takashima, J. Yamagughi, and M. Ishida, "Fault detection by signal response in PV module strings," presented at the 33rd IEEE Photovoltaic Spec. Conf., San Diego, CA, USA, 2008.

[15] A. Drews *et al.*, "Intelligent performance check of PV system operation based on satellite data," presented at the Eurosun 5th ISES Eur. Sol. Conf., Freiburg, Germany, 2004.

[16] S. Silvestre, A. Chouder, and E. Karatepe, "Automatic fault detection in grid connected PV systems," *Sol. Energy*, vol. 94, pp. 119–127, 2013.

[17] N. Gokmen, E. Karatepe, B. Celik, and S. Silvestre, "Simple diagnostic approach for determining of faulted PV modules in string based PV arrays," *Sol. Energy*, vol. 86, pp. 3364–3377, 2012.

[18] T. Shimakage, K. Nishioka, and H. Yamane, "Development of fault detection system in PV system," presented at the 33rd Annu. IEE Telecomun. Energy Conf. (INTELEC), Amsterdam, The Netherlands, 2011.

[19] A. Drews *et al*., "Monitoring and remote failure detection of grid-connected PV systems based on satellite observations," *Sol. Energy*, vol. 81, pp. 548–564, 2007.
[20] A. Chouder, S. Silvestre, N. Sadaoui, and L. Rahmani, "Modeling and simulation of a grid connected PV system based on the evaluation of main PV module parameters," *Simul. Model. Pract. Theory*, vol. 20, pp. 46–58, 2012.
[21] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Prog. Energy Combust. Sci*., vol. 34, pp. 574–632, 2008.
[22] R. Platon, S. Pelland, and Y. Poissant, "Modelling the power production of a photovoltaic system: Comparison of sugeno-type fuzzy logic and PVSAT-2 models," presented at the EuroSun ISES-Eur. Sol. Conf., Rijeka, Croatia, 2012.
[23] A. Mellit and S. A. Kalogirou, "ANFIS-based modelling for photovoltaic power supply system: A case study," *Renew. Energy*, vol. 36, pp. 250–258, 2011.
[24] B. Quesada, C. Sanchez, J. Canada, R. Royo, and J. Payá, "Experimental results and simulation with TRNSYS of a 7.2 kWp grid-connected photovoltaic system," *Appl. Energy*, vol. 88, pp. 1772–1783, 2011.
[25] S. Pelland, G. Galanis, and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model," *Prog. Photovoltaics Res. Appl*., vol. 21, pp. 284–296, 2011.
[26] K. M. Kandil *et al*., "Investigation of the performance of CIS photovoltaic modules under different environmental conditions," *Smart Grid Renew. Energy*, vol. 2, pp. 375–387, 2011.
[27] G. Bunea, K. Wilson, Y. Meydbray, M. Campbell, and D. D. Ceuster, "Low light performance of mono-crystalline silicon solar cells," presented at the 4th World Conf. Photovoltaic Energy Conf., Waikoloa, HI, USA, 2006.

Radu Platon received the B.S. degree in mechanical engineering and the M.S. degree in mechanical engineering from Concordia University, Montréal, QC, Canada, in 1999 and 2009, respectively.

Since 2000, he has been a Research and Development Engineer with CanmetENERGY (Natural Resources Canada), Varennes, QC, Canada, where he is currently working on building energy load forecasting, fault detection and diagnosis, and predictive modeling. His research interests include artificial intelligence techniques and multivariate data analysis.

Jacques Martel received the B.S. degree in engineering physics from Ecole Polytechnique de Montréal, Montréal, QC, Canada, in 1968, and the Ph.D. degree in nuclear engineering from MIT, Cambridge, MA, USA, in 1971.

He started his career as a Professor with INRS-Energy in Montréal, QC, Canada, where he became a Director of the center. He went on to manage the Industrial Materials Institute, National Research Council Canada, Ottawa, ON, Canada, and the Hydro-Québec Research Institute. Since 2009, he has been with CanmetENERGY, Varennes, QC, Canada, where he is involved in the development of the Intelligent Buildings Program.

Norris Woodruff received the B.A. degree in electrical engineering from Cambridge University, Cambridge, U.K. and the M.B.A. degree from the University of Toronto, Toronto, ON, Canada.

He was a Professional Engineer licensed with the PEO since 1978. After working for General Electric developing protection and control products, he founded Cachelan in 2003 to apply smart grid technology using IoT devices with cloud servers for managing renewable energy assets.

Tak Y. Chau received the B.S. degree in statistics and computer science from the University of Toronto, Toronto, ON, Canada, in 2002, and the B.Com. degree in IT management from Ryerson University, Toronto, ON, Canada, in 2006.

As the SolarVu Energy Portal Project Manager with Cachelan, Markham, ON, Canada, he formulates analytics for cloud server databases and develops web software to monitor solar PV systems.