Robust and Fast Detection of Small Power Losses in Large-Scale PV Systems

Åsmund F. Skomedal, Mari B. Øgaard, Halvard Haug, and Erik Stensrud Marstein

***Abstract*—Due to the fast growth in global installed photovoltaic (PV) capacity, performance monitoring for large-scale PV systems is an increasingly relevant and important topic. A large volume of research exists in this field, but there is a need for comparison of different methods and their performance toward relevant metrics, a broad discussion of the different choices involved, and subsequent consolidation. In this article, we focus on the detection of small power losses on string-level. We discuss the different choices involved in building a robust string performance monitoring scheme. We suggest the following approach: 1) identify bad data quality and do data-filtering; 2) calculate the daily specific yield on string level; 3) calculate the relative difference in specific yield between the strings (relative yield); 4) identify historical faults; 5) correct for seasonal variations; and 6) apply control charts to detect performance losses in new data and issue alarms/report to the system operators. Based on data from a utility scale PV power plant we compare different control charts in terms of detection time and sensitivity. We show that the cumulative sum (CUSUM) median and the Tukey-CUSUM charts are the most promising fault detection methods of the ones we have tested. We can robustly detect faults causing a performance loss of about 1% within 35 days of the drop in performance.**

***Index Terms*—Data analysis, monitoring, photovoltaics (PV), PV systems, solar power generation, time series analysis.**

I. Introduction

THE motivation for automatic fault detection in photovoltaic (PV) systems is twofold: The first goal is to avoid safety hazards and system breakdown, the second is to avoid revenue loss due to reduced production. Although large-scale PV systems are reliable compared to other sources of power, having the lowest outage rate of all power sources in the US [1], robust and fast detection of faults can avoid large revenue losses, and is important if we want to increase the competitiveness of PV energy.

Severe faults, such as blown fuses, open circuits, and inverter outages are easily detected by considering the production data, and many PV systems today apply manual or automatic monitoring schemes to handle these issues. Less severe faults, such as cell cracks, potential induced degradation, broken front glass, cell and busbar corrosion, shunts, and other problems at module level, cause small power losses, and are most commonly detected through aerial infrared thermography. However, a thermographic image is only a snapshot of the situation, and the thermal signatures depend on the ambient conditions during imaging. There is little research on how to estimate energy losses based on thermal images, making it difficult to make mitigation actions that are proportional to the severity of the faults. In a recently published paper, we try to mitigate this lack of knowledge by combining aerial thermographic imaging with power loss analysis on string level [2].

Here, we focus on fast detection of small and medium sized power losses (faults) in large commercial and utility scale systems through production data monitoring on string-level. By "large" we mean systems that have at least 20 subarrays that are individually monitored in terms of power or current. We do not consider fault diagnosis.

In the literature, there is no shortage of suggestions for different PV fault detection routines (FDRs) (see, for instance, [3]–[6]). Many publications propose novel techniques/algorithms for fault detection, and many of these techniques are undoubtedly very useful. However, the performance of any FDR depends heavily on the choices made in the preparation stages, and few publications focus on these choices. A couple of recent papers propose data quality routines for performance monitoring [7], [8], but this is an area that needs more research. Furthermore, there is no standardized way of benchmarking different FDRs. Hence, there is little comparison and consolidation of different FDRs in terms of performance, making it difficult for PV system operators to choose the right option.

In this study, we contribute toward filling these gaps by discussing what kind of PV performance metric should ideally be used in an FDR, and the choices involved in calculating this metric. Of the countless FDRs that exists in the PV fault detection literature, we have chosen to consider a certain subset of methods that have been used extensively in other industries for almost a century, namely *control charts.*

II. Introduction to Control Charts

The *control chart* was developed by Walter A. Shewhart in the 1920s [9], as he was addressing need for quality control for the quickly growing manufacturing industry. In its original version, the control chart is a graphical tool made to enable detection of

Manuscript received December 16, 2020; revised January 21, 2021; accepted February 16, 2021. Date of publication March 9, 2021; date of current version April 21, 2021. The work was performed within ENERGIX Project 282404, supported by the Research Council of Norway and Industry Partners. *(Corresponding author: Åsmund F. Skomedal.)*

Mari B. Øgaard, Halvard Haug, and Erik Stensrud Marstein are with the Department of Technology Systems, University of Oslo, 2007 Kjeller, Norway, and also with the Institute for Energy Technology, 2007 Kjeller, Norway (e-mail: mari.ogaard@ife.no; halvard.haug@ife.no; erik.marstein@ife.no).

Åsmund F. Skomedal is with the Institute for Energy Technology, 2007 Kjeller, Norway (e-mail: asmunds@ife.no).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JPHOTOV.2021.3060732.

Digital Object Identifier 10.1109/JPHOTOV.2021.3060732

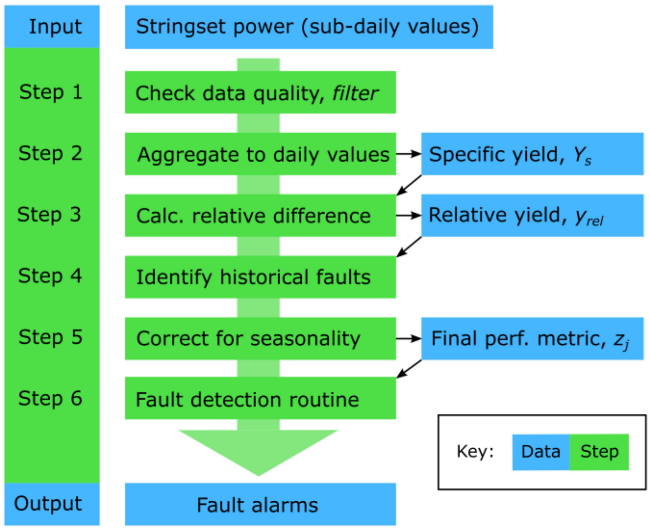

Fig. 1. Flowchart over the suggested methodology. Blue boxes represent different kinds of time series data, while green boxes represent the steps involved in the method.

unwanted changes in a process [10]. One plots a control metric, representing the quality of the process, on a graph, and for each sample checks whether the quality falls outside either the *upper* or *lower control limit*s (UCL and LCL). If it does, a corrective measure is made. In our case, "quality" is represented by the performance of the individual strings in the PV system.

With the (original) Schewart chart, the UCL and LCL are defined by the mean +/– 3 standard deviations (3σ) of the distribution [10]. Each sample is considered individually and deemed either in-control (if the quality value falls within UCL or LCL) or out-of-control. Later variants, such as the Cumulative Sum (CUSUM) [11] and Exponentially Weighted Moving Average (EWMA) [12] control charts, consider aggregates of samples over time, thus taking into account the evolution of quality. These aggregating charts are typically able to detect smaller changes and are less sensitive to noise.

Over the years, countless variations of control charts have been proposed, each with their own advantages. To name a few, there are control charts for autocorrelated time series data [13], multivariate data [14], nonparametric control charts [15], and robust control charts [16]. For a more detailed introduction to control charts, we refer to [10], [17], [18], as well as the above cited papers and the references therein.

A. Control Charts in the PV System Literature

The advantage of control charts, as compared to alternative fault detection methods for PV, is that they are simple to use and interpret. Of the countless publications on FDRs for PV systems, there are only a few that consider the use of control charts. In these papers, we may differentiate between two main approaches: those that require modeling PV output data and those that rely on comparison of the output of strings.

1) Approaches Based on Modeling: Platon *et al.* [19] model the ac power with an empirical model, and applies a 3σ Schewart chart to the *ratio* between the measured and modeled power. Livera *et al.* [20], [21] apply Schewart charts to the *difference* between the measured and modeled DC power, using the same parametric model as [19]. While Livera *et al.* in [20] use a 3σ detection rule, Platon *et al.* [21] use a 1σ rule, without justifying this choice.

In [22]–[27], Harrou *et al.* and Garoudja *et al.* parametrize the PV system through fitting the single diode model to a set of outdoor *I–V* measurements, and simulate the dc operating current, voltage, and power of the system. They then apply different control chart variants to monitor the difference between measured and modeled values. Platon *et al.* [19] and Livera *et al.* [20] apply CUSUM and multivariate CUSUM charts, respectively, Ramirez and Ramíez [18] and Livera *et al.* [21] apply EWMA and multivariate EWMA charts, respectively, and Harrou *et al.* in [22] and [23] apply a combination of wavelet multiscale representation of the data and EWMA charts (dubbed MW-EWMA). The multivariate variants simultaneously monitor the current, voltage, and power residuals, thus accounting for correlations in these values, reportedly giving better detection sensitivity. Unfortunately, no paper summarizing and comparing these approaches has been published. However, these papers demonstrate the advantage of using aggregating control charts, as opposed to basic Schewart charts.

2) Approaches Based on String Comparison: A string is an array of PV modules connected in series. Typically, the current of each string, or group of strings (hereafter referred to as a *stringset*), is measured, making it the smallest sub-array with electrical measurements in most PV systems.

Zhao *et al.* [28] makes the point that the different strings in a PV system is expected to have the same current, as long as they have the same installation conditions. Thus, they consider differences in current between strings, detecting faults through three different outlier rules: the 3σ rule, the Hampel identifier, and the Tukey (boxplot outlier) rule. The latter two are deemed more useful since they are robust to contaminated data. Although Zhao *et al.* do not mention control charts explicitly, their methods may be seen as variations of the basic Schewart chart.

Mallor *et al.* [29], too, compare different detection rules, although they use the cumulative range of power of the strings in the PV system as their control metric, where the range is the difference between the maximum and minimum power of the strings at each moment. In addition, they use functional principal component analysis to remove noise. This method is unique in the PV monitoring literature and seems to be effective in detecting faults. However, its weakness is that the PV system is monitored as a single unit, instead of monitoring each string separately. In addition, neither of the two methods mentioned in this section account for preexisting differences in performance between the strings. This means that a performance loss in the best performing string is less likely to be discovered than in the worst performing string.

III. Method

In this section, we will discuss the choices involved in setting up an FDR. One of the main contributions of this article is the proposed six step method shown in Fig. 1, which follows from the discussion. The steps and the associated choices will be

considered in an order that follows logically from the discussion, and this is not in their chronological order. Hence, in Subsection III-A we argue for the use of the relative difference in specific yield between the strings (Step 3), in Section III-B we consider data quality and filtering (Step 1) and different levels of time aggregation (Step 2), in Section III-C we discuss the correction of seasonal effects (Step 5), in Section III-D we introduce the different control chart variants (Step 6), and in Section III-E we consider the effect of historical faults in the data (Step 4).

One of the biggest challenges of PV performance monitoring is bad data quality. Typical data quality issues are missing data, values outside physical limits (corrupted data), sensor drift, nonconforming sensor installation, lacking maintenance of sensors, and wrong metadata. In addition, phenomena such as inverter and plant downtime, clipping, curtailment, irregular shading, irregular orientation of PV arrays, irregular ambient temperature patterns, and soiling cause features in the data that, unless adequately accounted for, may be treated as data quality issues.

One of the most important aspect of a robust FDR is how it deals with these data quality issues. Here, we make the use of *data filtering*, discussed in Section III-B2, *time aggregation* (Section III-B3, and *robust statistics*, i.e., statistics that work on data that are not normally distributed and that contain outliers (Section III-D).

In Section III-F we consider how to set the parameters of the control charts, and in Section III-G we describe how we generate a validation dataset by introducing synthetic performance losses in data from the field.

A. String Comparison Versus Modeling

Following [28] and [29], we argue that, in the case of string performance monitoring in large-scale PV systems, using string comparison instead of modeling the output of each string is preferred for the following reasons: If the strings are mounted similarly (both electrically and physically), the response to ambient conditions is virtually identical. Thus, by considering relative differences between the strings, we implicitly account for ambient conditions. The effect of curtailment and inverter clipping (which would have to be accounted for in a model) is also implicitly accounted for, assuming it affects all fault-free strings in each comparison group similarly.

Similar strings imply that they are mounted at the same inclination angle, or with the same tracker configuration, they have the same module and inverter type, and the same shading conditions. In a typical large-scale PV power plant, these conditions are fulfilled for all the strings except at the edges.

Furthermore, in the context of fault detection, the main advantage of a modeling based approach, in our view, is that it can be applied to systems with only one, or a few, stringsets. The main disadvantage is that it relies on sensor measurements, which themselves are prone to error. For instance, a pyranometer needs to be cleaned weekly and calibrated every year (in the case of class A monitoring systems [30]) to mitigate sensor drifting, and this is not always done. In addition, it can be time consuming (and hence expensive) to model a PV system, especially if *I–V* measurements have to be made or sensor equipment has to be installed. Finally, we know of no evidence that modeling-based approaches have any advantage in terms of accuracy as compared to approaches based on string comparison. For these reasons, we recommend an approach based on string comparison.

We propose using the *relative yield*, defined by (1), as the basis for measuring the performance (or quality) of individual strings. For a given period, the relative yield is defined by

$$y_{\mathrm{rel},j} = \frac{Y_{s,j} - \widetilde{Y_s}}{\tilde{Y}_s} \cdot 100\% \tag{1}$$

where $Y_{s,j} = \frac{E_j}{P_{STC,j}}$ is the specific yield of string(set) number *j*, E_j is its energy output in the period, $P_{\mathrm{STC},j}$ is the STC power rating of the string, and $\widetilde{Y_s}$ is the median specific yield of the group of strings associated with string *j*. The use of specific yield means it is possible to compare the output of strings with different power rating. In the case of central inverters, we recommend grouping strings connected to the same inverter, as these strings are likely to be co-located and operating under the same maximum power point tracker, and hence have similar operating conditions. In the case of string inverters, we propose creating groups of about 50 co-located strings with equal mounting configuration. In Fig. 2(a) and (b), we show unfiltered and filtered variants of y_{rel}.

B. Sources of Variation

Any variation in y_{rel} between the strings in a PV system is due to either 1) string-level performance deviations, 2) intrasystem variations in ambient conditions, or 3) measurement noise.

1) Performance Deviations: It is useful to differentiate between *pre-existing* and *currently occurring* performance deviations. The former is present already from the start-up of the PV system. Such deviations are caused primarily by variations in the capacity of PV modules, installation angle, and (in the case of central inverters) distance from inverter (causing different series resistance due to different cable lengths). This also necessarily includes any deviations due to damage during transport and installation. Pre-existing performance deviations can be dealt with by subtracting the (fault-free) historical performance (represented by x_0 in Section III-D) from the day-to-day performance. We will come back to this in Section III-D.

Currently occurring performance deviations, on the other hand, are caused by electrical faults, avoidable shading (e.g. vegetation and soiling), or any of the module faults summarized in, e.g., [31] that develop during operation. For simplicity, we refer to occurring performance deviations simply as *faults*. Of course, it is the detection of faults we are addressing in this study. But to do this effectively, we need to account for the other sources of variation in y_{rel} in a satisfactory manner.

2) Variations in Ambient Conditions: Intrasystem variations in ambient conditions are primarily caused by partially cloudy conditions (covering parts of the system), shading, or variations in ambient temperature. Our preferred strategy to deal with this is filtering, ensuring the comparison is made at times with as uniform conditions as possible. By using the clearsky detection algorithm provided in pvlib Python [32], we can filter away times

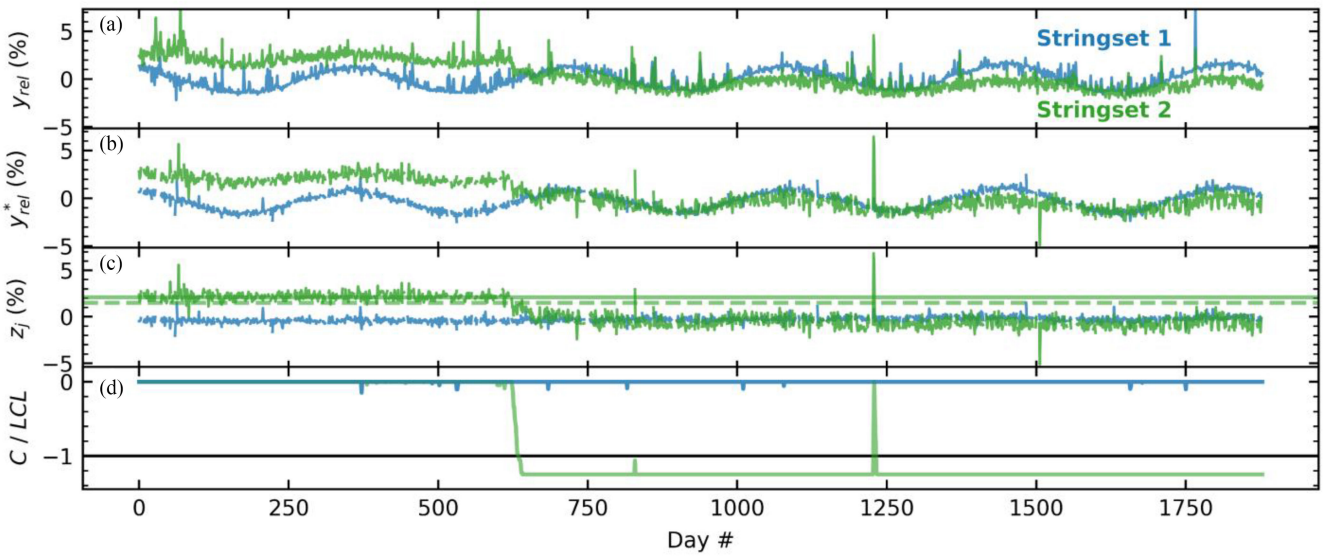

Fig. 2. Field data from two stringsets in the PV system used in this study (see Section III-G), showing the progression of the method presented in this article. (a) Unfiltered relative yield. (b) Filtered relative yield. (c) Seasonally corrected relative yield. (d) Normalized *CUSUM median* control statistics of the two stringsets. In (c), the solid horizontal line represents the expected value $x_{0,j}$ of z_j, and the dashed line represents the reference value $(x_0 - K)$ of stringset 2. Stringset 2 has a fault approximately at day 625 which is detected on day 635 (when $C/LCL < -1$). Note that C has been limited to a minimum value of $-1.25 \cdot LCL$.

of cloudy conditions. This algorithm works by comparing the measured global horizontal irradiance signal to the theoretical irradiance given clear skies. Furthermore, we suggest following the framework proposed in [33], where we filter by minimum solar elevation angle and maximum angle-of-incidence. This removes morning and evening data, ensuring a minimal amount of inter-row shading and shading from the surroundings, and minimal angle-of-incidence effects. Following [33], we find optimal filtering thresholds by minimizing the ratio of the standard deviation of y^*_{rel} to the fraction of remaining data.

We denote the filtered relative yield by an asterisk: y^*_{rel}. In Fig. 2(a) and (b), we show the difference between unfiltered and filtered y_{rel}.

Sometimes, intrasystem variations in ambient conditions can have a seasonal pattern, causing seasonal effects in y_{rel}. We will come back to how these seasonal effects can be accounted for in Section III-C.

3) Measurement Noise: Measurement noise is an unavoidable source of variation in y^*_{rel}. In the following, we will treat all unexplained and unfiltered variations as noise. Thus, after filtering, any remaining variation caused by differences in ambient conditions (such as temperature) is also treated as noise. The magnitude of the noise will determine how small performance deviations we will be able to detect. Hence, it is desirable to minimize the noise. In addition to filtering, there is another strategy for minimizing noise, namely time aggregation.

Y_s (and y_{rel}) is defined for a given period, and this period might be anything from a minute to a year. Even though the string currents are typically logged with a periodicity between 1 and 15 min, it is advantageous to aggregate the data into hourly or daily values, because this reduces the random noise in the data. By comparing the median absolute deviation of y_{rel} on a 10 min basis to hourly and daily aggregates of y_{rel} and y^*_{rel}, we can get a sense of what choice gives the smallest amount of noise. Obviously, the noise levels decrease with increasing aggregation level. As an example, the median absolute deviation (MAD) of the unfiltered 10 min data is 1.42%, while the MAD of the filtered daily data is 0.27%.

Of course, by aggregating to daily values we lose the opportunity to detect and mitigate faults on a minute-to-minute or hour-to-hour basis. However, in the case of PV monitoring we would argue that, unless there is a large outage disabling many arrays, the revenue loss of a faulty string is negligible on an intraday basis. Thus, daily values constitute a good balance between minimizing noise (maximizing detection sensitivity) and minimizing detection times. We also assume that an outage causing a non-negligible revenue loss on an intraday basis (e.g., inverter outage) will be discovered independently of our fault detection routine.

C. Correcting for Seasonal Variations

In Fig. 2(b), there is a systematic seasonal pattern in the relative yield. These seasonal variations are present in the y^*_{rel} of many PV systems, and they likely stem primarily from small differences in installation angle. Other effects that might give rise to seasonal variations are systematic temperature differences between the strings and inter-row shading that has not been filtered out. In any case, this introduces a seasonal component in y^*_{rel} which we need to account for.

The most straightforward approach to dealing with the seasonal component is to adjust the detection thresholds to account for the extra variation [34]. A more advanced approach that does not reduce the detection sensitivity is to model the data with a time series model, such as an autoregressive integrated moving average (ARIMA) model, and to run the fault detection routine

on the residuals of the model [35]. The seasonal ARIMA is not designed for periodicities as large as 365, as is the case here. Instead we choose to model the seasonality with *Seasonal and Trend Decomposition using Loess* [36]. This is a convenient analysis tool for decomposing time series into a trend and a seasonal component. The model is simple to instantiate, only requiring the specification of the periodicity, and it is robust to outliers [37]. We make use of the implementation in version 0.11.0 of the statsmodels package in Python [38]. Of course, we are only interested in the seasonal component, but the trend component allows seasonal components to be detected even in the presence of small shifts in the trend, as is the case in stringset 2 in Fig. 2.

After finding the seasonal component SC_j of each stringset *j*, we can correct for seasonality by subtracting the seasonal component from the filtered relative yield, giving us the *seasonally corrected relative yield,* which we denote by $z_{i,j}$ for each day *i* and each string *j*

$$z_{i,j} = y^*_{\mathrm{rel},i,j} - SC_{i,j}. \tag{2}$$

$z_{i,j}$ is the performance metric we will use for the fault detection routine, i.e., the measure of performance (or quality) that we will apply the control charts to.

In Fig. 2(c) we show the seasonally corrected relative yield of two stringsets from the PV system we have studied. By filtering and correcting for seasonality, we magnify the effect of performance losses relative to the variation in the data.

D. Control Charts

We will now introduce six different control charts: the Schewart control chart, three variants of the CUSUM chart, one variant of the EWMA chart, and one variant employing the moving median. As we want to monitor the performance of each individual string, we apply one univariate control chart to each string (the subgroup size is one). The control charts all have a lower control limit (*LCL*), which is defined by

$$\mathrm{LCL}_j = h \cdot \xi_j \tag{3}$$

where ξ_j represents the variation in $z_{i,j}$, and h is a multiplier that can be adjusted according to the desired detection sensitivity of the control chart. The statistic used to estimate ξ_j depends on which control chart is used. An alarm is issued on day *i* if the *control statistics* is smaller (more negative than) $-\mathrm{LCL}_j$. In the following, we will present the control statistics of each control chart.

1) Schewart Chart: The Schewart chart is defined by the control statistics

$$S_{i,j} = z_{i,j} - x_{0,j} \tag{4}$$

where $x_{0,j}$ represents the expected (in-control) value of z_j, estimated by the mean. In this sense, the control statistics is simply $z_{i,j}$ corrected to its mean value. The use of $x_{0,j}$ ensures that differences in $z_{i,j}$ caused by pre-existing performance deviations is accounted for. ξ_j is estimated by the standard deviation of fault-free (in-control) values of $z_{i,j}$. We will come back to how fault-free samples are identified in Section III-E.

2) CUSUM Charts: The cumulative sum chart is defined by the control statistics

$$C_{i,j} = \min\begin{pmatrix} 0, \\ C_{i-1,\mathrm{j}} + z_{i,j} - (\mathrm{x}_{0,\mathrm{j}} - K_j) \end{pmatrix} \tag{5}$$

where $K_j = k \cdot \xi_j$ is the *reference value*, and $C_{0,j} = 0$. Note that $C_{i,j}$ represents the cumulative sum of $z_{i,j} - (\mathrm{x}_{0,\mathrm{j}} - K_j)$ limited so that it never goes above 0. Note also that if $C_{i-1,j} = 0$, $z_{i,j}$ needs to be smaller (more negative) than $x_{0,j} - K_j$ to affect $C_{i,j}$. Hence, *K* determines the distance between $z_{i,j}$ and $x_{0,j}$ needed for $C_{i,j}$ to start aggregating. Thus, the reference value multiplier *k* is decisive for the sensitivity of the chart, i.e., how small faults the method will be able to detect.

Robust CUSUM charts, based on metrics such as the median and the median absolute deviation have been proposed in previous research [16], [39]–[41]. Inspired by [9] and [29], we propose three different estimators for x_0 and ξ: In the *standard CUSUM* chart, x_0 and ξ are estimated respectively by the mean and the standard deviation of historic in-control values of $z_{i,j}$. In the *CUSUM median* chart [40], x_0 and ξ are estimated by the median and the median absolute deviation (MAD) respectively, where $\mathrm{MAD}(z_{i,j}) = |z'_{i,j} - \widetilde{z'_{i,j}}|$. Here, $z'_{i,j}$ is the set of historic in-control values of $z_{i,j}$,, and $\widetilde{z'_{i,j}}$ is the median of $z'_{i,j}$. In the *Tukey-CUSUM* chart [41], x_0 and ξ are estimated by the first quartile and the interquartile range of $z'_{i,j}$ respectively.

Note that a computationally effective way of calculating C_i is the cumulative *sum* of $D_i = z_i - (x_0 - K)$ (over all *i*), minus the cumulative *max* of the cumulative sum of D_i (where $D_0 = 0$).

3) EWMA Chart: The exponentially weighted moving average chart is defined by the control statistics:

$$E_{i,j} = E_{i-1,\mathrm{j}}\ (1-\lambda) + (z_{i,j} - x_{0,j})\lambda \tag{6}$$

where $\lambda \in (0, 1]$ is the *smoothing parameter* and $E_{0,j} = 0$. The EWMA statistics is simply a weighted mean of previous observations, where the weights decrease exponentially from the latest observation and backwards in time. If λ is large, more weight is given to recent observations, giving a control chart with a fast response, but a higher sensitivity to outliers. If λ is small, the opposite is true. Typical values for λ is between 0.2 and 0.3 [10].

4) Moving Median Chart: We propose the moving median chart, which is defined by the control statistics:

$$M_{i,j} = \widetilde{(\mathrm{z}_{\mathrm{i,j}} - x_{0,j})}_{SM,d} \tag{7}$$

which denotes the simple moving median of $z_{i,j} - x_{0,j}$ over the d days before day i, and where $M_{0,j} = 0$. $x_{0,j}$ is estimated by the median, and ξ_j is estimated by the MAD of historic in-control samples. Note that the moving median chart is similar to the fault detection scheme that was proposed in our previous publication, except that we proposed using a moving max [33]. We do not think a moving max is the best option, however, as the max is more sensitive to outliers (less robust) than the median.

5) Control Parameters: All the four control chart categories mentioned above contain *control parameters* that need to be set. All four categories have *h* in common, which decides the value of the lower control limit, *LCL*. In addition, we need to

set the reference value multiplier k of the CUSUM charts, the smoothing parameter λ of the EWMA chart, and the number of aggregation days d in the moving median chart. These should be set in order to minimize the detection time, without creating false positives (false alarms). We will come back to how optimal control parameters are found in Section III.F

E. Identifying Fault-Free Samples

Usually, two phases are involved in any FDR. If phase II is the online fault detection phase, phase I is an offline analysis where the control parameters are set. This involves identifying fault-free samples in historical data and testing the control charts at different settings. We will discuss how phase I can be done in the following three sections.

In a small dataset, the fault-free samples can be found manually. In our case, however, since there may be thousands of strings in the PV system, and each of these is to be treated as a separate unit, we need a robust way of identifying fault-free samples automatically. It is important that we know that the data is actually fault-free, so we would rather brand too much data as faulty ("out-of-control") than too little. Hence, we propose a relatively "strict" offline fault detection routine:

We propose using a combination of the moving median and the Hampel identifier [42], identifying both individual outliers and periods of lower performance. To be specific, we suggest first identifying individual outliers by

$$z_{i,j} < \widetilde{z}_j - 3 \cdot \xi' \tag{8}$$

where $\widetilde{z}_j$ is the median of $z_{i,j}$ and ξ' is the median absolute deviation of $z_{i,j} - \widetilde{(z_{i,j})}_{SM,31}$. Next, outliers are removed from $z_{i,j}$, giving us $z^*_{i,j}$. Longer periods of performance loss are identified by

$$\widetilde{\left(z^*_{i,j}\right)}_{SM,31} < P_{75}\left(z^*_{i,j}\right) - 2 \cdot \xi' \tag{9}$$

where $P_{75}(z^*_{i,j})$ is the 75th percentile of $z^*_{i,j}$. Note that the combination of subtracting the 75th percentile and the small detection threshold $-2 \cdot \xi'$ makes for a relatively strict detection scheme, which is what we aim at. Also note that, by using a 31 day rolling median, the 15 days at the beginning and end of the time series will be excluded from the analysis. This is acceptable, assuming that the time series is longer than a year, meaning we can afford to brand this data as out-of-control for the purpose of phase I.

The in-control (fault-free) samples is then used to estimate $x_{0,j}$ and ξ_j, in addition to determining optimal control parameters for the control charts. If available, we recommend using more than a whole year of in-control $z_{i,j}$ when determining $x_{0,j}$ and ξ_j, as this will account for at least some of the remaining trend or seasonality in the data.

F. Evaluating Control Charts and Setting Control Parameters

The *detection time* (*DT*) (called *run length* in the control chart literature) is the time it takes from a performance loss occurs until it is detected. By running the control charts on data with known faults of size δ, we can determine the *average detection time* (ADT_δ) for each δ and each control chart. In this way we can also test the control charts with different control parameter settings, thus finding their optimal values.

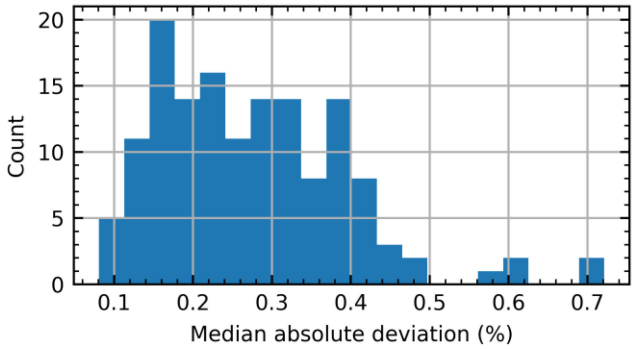

Fig. 3. Distribution of median absolute deviations (MADs) for fault-free stringsets in the field data in this study.

The choice of FDR will be based on a tradeoff between minimizing the number of *false positives* and the average detection time, ADT_δ. A false positive is an alarm that is raised in the absence of any fault. In our case, the smaller *h, k,* and *d* are, and the larger λ is, the smaller the ADT_δs are. However, smaller values of ADT_δ are correlated with a higher chance of false positives. For this reason, the control parameters are typically set by minimizing ADT_δ under the condition of a predetermined maximum false positive rate. Here, the requirement is that there are false alarms in less than 1% of the units over the course of 10 years. We will refer to this condition as the *maximum false positive rate*.

Furthermore, if a fault remains undetected for one year after it sets in, we deem this a *false negative*. We require the fraction of false negatives to be less than 10%. We refer to this condition as the *maximum false negative rate*.

G. Validation Dataset

To validate the methodology proposed in this article, and to compare the performance of the different control charts, we tested the procedure on a dataset from the field. The field data were from a utility-scale (>10 $\mathrm{MW_p}$) power plant in Sub-Sarahan Africa. In this plant, each inverter had about 150 strings connected to it, operating under one maximum power point tracker, and hence the same voltage. Power were logged on a one-minute basis for stringsets consisting of two strings (of 24 modules) in parallel. The comparison groups (defined in Section III-A) consisted of 75 parallel-connected stringsets. Due to business sensitivity, we cannot disclose the exact location of this power plant.

In order to make validation possible, the field data were modified by introducing synthetic faults. We first calculated the seasonally corrected relative yield, $z_{i,j}$, for each stringset. From this, we identified a set of 145 fault-free stringsets following the procedure in Section III.E. The faults were introduced at day 366 and onwards by subtracting $z_{i,j}$ by $\delta \cdot \mathrm{MAD}_j$, where MAD_j is the median absolute deviation of the fault-free $z_{i,j}$ of each time series. A histogram of the MADs of the stringsets is shown in Fig. 3. The control charts were tested with $\delta \in$ [1, 2, 2.5, 3, 4, 5, 10, 20, 30, 50].

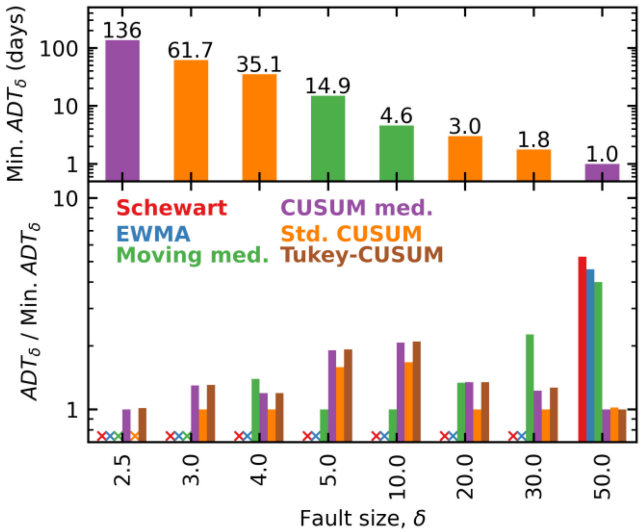

Fig. 4. Minimum average detection times (upper panel) and relative average detection times (lower panel) of the six control charts for different fault sizes δ (multiples of the median absolute deviation in the data) for *modified field-data*. Crosses signify a false negative rate above 10%. The colors in the upper panel correspond to which control chart has the minimum average detection time at each fault size.

IV. Results and Discussion

The optimal control parameters were found by brute force optimization. We iterated through a comprehensive set of combinations of control parameters, and the combinations enabling the detection of the smallest fault for each control chart was chosen. Note that, in addition to the conditions of maximum false positives and negatives (see Section III-F), the chosen combination is first the one that makes the control chart able to detect the smallest changes, and secondly gives the smallest ADT_δs. We make this choice because it is better that a fault is detected slowly than that it is never detected.

In Fig. 4, we compare the ADT_δs of the control charts. The upper subfigure shows the minimum ADT_δ of all the control charts, and the lower subfigure shows the ratio of the ADT_δ of each control chart to the minimum ADT_δ. Crosses (instead of bars) mean the given control chart was not able to detect the fault with a false negative rate below 10%.

Our first observation is that the Schewart and EWMA charts are not able to detect but the largest of the faults. The reason for this is that they are not able to fulfill the condition of the maximum false positive rate unless the detection threshold multiplier (h) is so large that smaller faults remain undetected. We take this to mean that the versions of the Schewart chart and the EWMA chart that we have tried are unsuited to do PV string performance monitoring.

Furthermore, faults of $\delta < 2.5$ could not be detected with any of the control chart variants. This is because there is a small, non-constant trend present in many of the $z_{i,j}$ time series. We do not know the exact source of this trend; it may be a remnant of the seasonal component, it may stem from a relative difference in module degradation between the stringsets, or it may stem from varying soiling conditions through the period. In any case, the trend forces the control parameters to be adjusted to decrease detection sensitivity and avoid false positives. This makes the control charts unable to detect small faults in this dataset.

The minimum ADT_δ for $\delta = 20$ is 3 days. Given a MAD of 0.27%, which is the average MAD($z_{i,j}$) in our data, this means that a loss in performance of about $20 \cdot 0.27\% = 5.4\%$ can be detected in about 3 days. Given a stringset with a nominal capacity of 16 kW, an average of 5 peak sun hours per day, and a tariff of 0.05 \$/kWh, this would mean a lost revenue of $5.4\% \cdot 3$ days $\cdot$ 16 kW $\cdot$ 5 h/day $\cdot$ 0.05 \$/kWh $\approx$ 0.65 \$ before the fault was discovered. If this fault was never discovered, it would lead to lost revenue of about 79 \$ per year. We can assume that any fault more costly than this would likely be detected in 3 days or less.

The moving median chart gives the smallest ADT_δ for the middle δs, while the standard CUSUM and the CUSUM median charts give the smallest ADT_δ for smallest and largest δs. The detection time of the moving median chart is determined by the length of the moving window. If the moving window is d days, it will on average take $d/2$ days to detect faults larger than *LCL*. Faults in the size range of *LCL* may take longer than $d/2$ days to detect, and faults much smaller than *LCL* will not be detected. With the CUSUM charts, however, larger faults will always be detected faster than smaller faults. This is why the CUSUM charts perform best at high δs. In other words: The moving median chart may detect some faults faster than the CUSUM charts, but at the price of detecting larger faults slowly and being unable to detect the smallest faults.

If we were to choose one control chart, we would choose one of the CUSUM charts. The main reason for this is that the moving median chart is unable to detect faults of $\delta < 4$ in our dataset. Furthermore, the difference in detection times between the moving median chart and the best CUSUM charts is small where the moving median chart is faster, but large where the CUSUM charts are faster. The three variants of the CUSUM charts perform very similarly. The biggest difference between them is that the standard CUSUM chart is unable to detect faults of $\delta = 2.5$. Other than that, it has slightly smaller detection times than the CUSUM median chart and the Tukey-CUSUM. All in all, the difference in ADTs between the CUSUM charts is small, but the median and Tukey variants are more robust to outliers and unknown data features. This is why they are able to detect faults of $\delta = 2.5$ better then the standard CUSUM chart. Therefore, of the control charts we have tested, we recommend using the CUSUM median or the Tukey-CUSUM charts.

V. Conclusion

In this study we discuss the most important choices involved in building a fault detection routine focused on fast detection of small performance losses in a robust manner. We discuss the use of modeling versus string comparison, data filtering, time aggregation, correcting for seasonality, correcting for pre-existing performance deviations, and the use of robust statistics.

Our proposed approach is summarized in a six-step method, where we calculate the daily aggregated, filtered, and seasonally corrected relative yield as a metric to represent the performance of PV strings (or stringsets). We compare six different control

charts, assessing them in terms of average detection time at different magnitudes of performance loss. Of these, we recommend using the CUSUM median or the Tukey-CUSUM control charts, as these are the robust methods that most consistently ensures the detection of the smallest faults. We are for instance able to detect sudden losses in performance of 4 and 20 times the median absolute deviation (about 1% and 5.4% power loss) in about 35 and 3 days, respectively.

APPENDIX

TABLE I
OPTIMAL CONTROL PARAMETER VALUES

Control chart	Control parameter value	
	h	k, λ, d
Schewart chart	20	-
Standard CUSUM	34	1
CUSUM median	82	1.8
Tukey-CUSUM	41	0.9
EWMA	17.5	0.9
Moving median	5	11

REFERENCES

[1] P. Hibbard, S. Tierney, and K. Franklin, "Electricity markets, reliability and the evolving us power system," White Paper, 2017.

[2] Å. F. Skomedal, B. L. Aarseth, H. Haug, J. Selj, and E. S. Marstein, "How much power is lost in a hot-spot? A case study quantifying the effect of thermal anomalies in two utility scale PV power plants," *Sol. Energy*, vol. 211, pp. 1255–1262, Nov. 2020.

[3] A. Mellit, G. M. Tina, and S. A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: A review," *Renewable Sustain. Energy Rev.*, vol. 91, pp. 1–17, 2018.

[4] D. S. Pillai, F. Blaabjerg, and N. Rajasekar, "A comparative evaluation of advanced fault detection approaches for PV systems," *IEEE J. Photovolt.*, vol. 9, no. 2, pp. 513–527, Mar. 2019.

[5] A. Livera, M. Theristis, G. Makrides, and G. E. Georghiou, "Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems," *Renewable Energy*, vol. 133, pp. 126–143, 2019.

[6] M. N. Akram and S. Lotfifard, "Modeling and health monitoring of DC side of photovoltaic array," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1245–1253, Oct. 2015.

[7] S. Lindig, A. Louwen, D. Moser, and M. Topic, "Outdoor PV system monitoring—Input data quality, data imputation and filtering approaches," *Energies*, vol. 13, no. 19, pp. 1–18, 2020.

[8] A. Livera *et al.*, "Data processing and quality verification for improved photovoltaic performance and reliability analytics," *Prog. Photovolt. Res. Appl.*, vol. 29, pp. 143–158, 2020.

[9] W. A. Shewhart, *Economic Control of Quality of Manufactured Product*. New York, NY, USA: D. Van Nostrand Company

[10] *Engineering Statistics Handbook*, NIST/SEMATECH, Gaithersburg, MD, USA, 2012.

[11] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1, pp. 100–115, 1954.

[12] S. W. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 42, no. 1, pp. 97–101, Feb. 1959.

[13] S. Knoth and W. Schmid, "Control charts for time series: A review," *Front. Statist. Qual. Control*, vol. 7, pp. 210–236, 2004.

[14] S. Bersimis, S. Psarakis, and J. Panaretos, "Multivariate statistical process control charts: An overview," *Qual. Reliab. Eng. Int.*, vol. 23, no. 5, pp. 517–543, Aug. 2007.

[15] S. Chakraborti and M. A. Graham, "Nonparametric (distribution-free) control charts: An updated overview and some results," *Qual. Eng.*, vol. 31, no. 4, pp. 523–544, 2019.

[16] H. Z. Nazir, M. Riaz, R. J. M. M. Does, and N. Abbas, "Robust CUSUM control charting," *Qual. Eng.*, vol. 25, no. 3, pp. 211–224, 2013.

[17] L. E. Mainstone and A. S. Levi, "Fundamentals of statistical process control," *J. Organ. Behav. Manage.*, vol. 9, no. 1, pp. 5–21, 1988.

[18] B. Ramirez and J. Ramiŕez, *Douglas Montgomery's Introduction to Statistical Quality Control : A JMP Companion*. Cary, NC, USA: SAS Institute, 2018.

[19] R. Platon, J. Martel, N. Woodruff, and T. Y. Chau, "Online Fault detection in PV systems," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1200–1207, Oct. 2015.

[20] A. Livera, G. Makrides, J. Sutterlueti, and G. E. Georghiou, "Advanced failure detection algorithms and performance decision classification for grid-connected PV systems," in *Proc. 33rd Eur. Photovolt. Sol. Energy Conf. Exhib.*, 2017, pp. 1–7.

[21] A. Livera, M. Florides, M. Theristis, G. Makrides, and G. E. Georghiou, "Failure diagnosis of short- and open-circuit fault conditions in PV systems," in *Proc. IEEE 7th World Conf. Photovolt. Energy Convers., Joint Conf. 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC*, 2018, pp. 739–744.

[22] F. Harrou, Y. Sun, and A. Saidi, "Online model-based fault detection for grid connected PV systems monitoring," in *Proc. 5th Int. Conf. Elect. Eng.*, 2017, pp. 1–5.

[23] F. Harrou, Y. Sun, and A. Saidi, "Model-based fault detection algorithm for photovoltaic system monitoring," in *Proc. IEEE Symp. Ser. Comput. Intell. SSCI Proc.*, 2018, pp. 1–5.

[24] E. Garoudja *et al.*, "Statistical fault detection in photovoltaic systems," *Sol. Energy*, vol. 150, pp. 485–499, 2017.

[25] F. Harrou, Y. Sun, B. Taghezouit, A. Saidi, and M. E. Hamlati, "Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches," *Renewable Energy*, vol. 116, pp. 22–37, 2018.

[26] F. Harrou, B. Taghezouit, and Y. Sun, "A robust monitoring technique for fault detection in grid-connected PV plants," in *Proc. 7th Int. IEEE Conf. Renew. Energy Res. Appl.*, 2018, pp. 594–598.

[27] F. Harrou, B. Taghezouit, and Y. Sun, "Robust and flexible strategy for fault detection in grid-connected photovoltaic systems," *Energy Convers. Manage.*, vol. 180, pp. 1153–1166, 2019.

[28] Y. Zhao, B. Lehman, R. Ball, J. Mosesian, and J. F. De Palma, "Outlier detection rules for fault detection in solar photovoltaic arrays," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2013, pp. 2913–2920.

[29] F. Mallor *et al.*, "A method for detecting malfunctions in PV solar panels based on electricity production monitoring," *Sol. Energy*, vol. 153, pp. 51–63, 2017.

[30] Photovoltaic System Performance - Part 1: Monitoring, IEC 61724-1, 2016.

[31] M. Köntges *et al.*, "Assessment of photovoltaic module failures in the field," IEA, Paris, France, IEA-PVPS T13-09: 2017, 2017.

[32] J. S. Stein, W. F. Holmgren, J. Forbess, and C. W. Hansen, "PVLIB: Open source photovoltaic performance modeling functions for Matlab and Python," in *Proc. IEEE 44th Photovolt. Spec. Conf.*, 2017, pp. 1–6.

[33] Å. Skomedal, M. B. Øgaard, J. Selj, H. Haug, and E. S. Marstein, "General, robust, and scalable methods for string level monitoring in utility scale PV systems," in *Proc. 36th Eur. Photovoltaic Sol. Energy Conf. Exhib.*, 2019, pp. 1283–1287.

[34] W. Jiang, T. E. Murphy, and K.-L. Tsui, "Statistical methods for quality and prdouctivity improvement," in *Handbook of Engineering Statistics*. New York, NY, USA: Springer, 2006, pp. 173–192.

[35] H. B. Nembhard and P. Changpetch, "Directed monitoring using cuscore charts for seasonal time series," *Qual. Reliab. Eng. Int.*, vol. 23, pp. 219–232, 2006.

[36] R. Cleveland, W. Cleveland, J. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on Loess (with discussion)," *J. Official Statist.*, vol. 6, no. 1, pp. 3–73, 1990.

[37] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, Vic, Australia: OTexts, 2018.

[38] Statsmodels version 0.11.0, 2020.

[39] P. Castagliola, F. O. Figueiredo, and P. E. Maravelakis, "The cusum median chart for known and estimated parameters," *Revstat Statist. J.*, vol. 17, no. 3, pp. 345–370, 2019.

[40] L. Yang, S. Pai, and Y. R. Wang, "A novel CUSUM median control chart," in *Proc. Int. MultiConf. Eng. Comput. Sci.*, 2010, pp. 1707–1710.

[41] Q. U. A. Khaliq and M. Riaz, "Robust Tukey-CUSUM Control Chart for Process Monitoring," *Qual. Reliab. Eng. Int.*, vol. 32, no. 3, pp. 933–948, 2016.

[42] R. K. Pearson, *Mining Imperfect Data*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 2005.