Received: 18 June 2021 | Revised: 12 October 2021 | Accepted: 28 November 2021
DOI: 10.1002/pip.3523

APPLICATION

PROGRESS IN PHOTOVOLTAICS WILEY

Automated detection of photovoltaic cleaning events: A performance comparison of techniques as applied to a broad set of labeled photovoltaic data sets

Matthew Muller[1] | Kirsten Perry[1] | Leonardo Micheli[2] | Florencia Almonacid[2] | Eduardo F. Fernández[2]

[1]PV Performance and Reliability Group, National Renewable Energy Laboratory, Golden, CO, USA
[2]Advances in PV technology Research Group (AdPVTech), University of Jaen, Jaen, Spain

Correspondence
Matthew Muller, PV Performance and Reliability Group, National Renewable Energy Laboratory, Golden, CO, USA.
Email: matthew.muller@nrel.gov

Funding information
Solar Energy Technologies Office, Grant/Award Numbers: 35893, 34348; Spanish Ministry of Science, Innovation and Universities, Grant/Award Number: RYC-2017-21910; European Union's Horizon 2020, Grant/Award Number: 793120

Abstract

Extracting accurate soiling loss information from photovoltaic (PV) production data first requires segmenting the time series data per natural or manually occurring cleaning events. Maintenance logs are often incomplete, rain data are often unavailable, and the debate on rain thresholds for cleaning and dew or wind cleanings is still ongoing. The present work aims to overtake these issues by improving automated methods to detect these cleaning events and therefore improve extraction of soiling loss information. Time series power production data from 22 PV inverters were labeled for natural or manually occurring cleaning events. The data sets were carefully selected to include varying degrees of soiling, cleaning events, and noise. Several algorithms, including filtering logic and change point detection, were examined for efficacy at detecting the labeled cleanings. All the methods introduced except for changepoint detection showed significant improvement at detecting the labeled cleaning events per the mean F_1 score. Furthermore, the highest performing cleaning detection algorithm achieved an absolute increase in the mean F_1 score of 43% over the default version of the RdTools stochastic rate and recovery (SRR) algorithm. The highest performing algorithm included irradiance filtering and a cleaning detection threshold, adjusted based on the 40-day centered rolling median of the absolute day-to-day deviations in the daily performance index (PI). These improvements are promising as cleaning detection is an essential step in the automated analysis of PV soiling.

KEYWORDS
photovoltaics, soiling, soiling extraction, time series data

1 | INTRODUCTION

Photovoltaic (PV) soiling is the loss of PV output power due to the deposition of dust, dirt, pollen, or other airborne particulates on the surface of PV modules. The deposited soil prevents incoming solar irradiance from being transmitted through the front surface glass into the active PV semiconductor material, resulting in output power loss. While the concept of PV soiling is simple (dirty air leads to dirty glass and reduced transmission of light), both modeling and measuring PV soiling has proven to be a complex topic resulting in much academic work over the last decade.[1,2] Modeling efforts have focused on relating environmental parameters such as airborne particulate levels, pollution sources, soil types, land use categories, climate zones, windspeed, relative humidity, and other variables to dust deposition rates.[3–12] Research efforts in the area of measuring PV soiling losses have focused on techniques and novel equipment,[13–23] as well as developing algorithms to extract soiling losses directly from PV system data.[24–31] Each of these efforts is intimately tied together. Independent engineers, project planners, and financial institutions must forecast the performance of new PV systems, and soiling loss models are needed to reduce uncertainty and de-risk projects. Logically, any academic effort to develop soiling loss models based on environmental parameters requires validation against soiling loss measurements or losses extracted from PV time series data. It is paramount for PV system owners to quantify soiling losses at the system level, as this directly accounts for lost energy and drives project operations and maintenance (O&M) and subsequent financial decisions.[32] System level soiling losses can be complicated by the practice of overbuilding PV DC capacity in relation to an inverters AC capacity (resulting in what is called inverter clipping). For example, if dust on the PV surface is reducing light reaching the solar cells by 10% but at the same time, due to DC overbuild, the inverter is clipping more than 10% of the available power, no soiling losses will be directly observable in the AC signal. Soiling losses are further mixed in with other loss factors that are dependent on parameters such as irradiance, shading, temperature, and other variables; hence, the effort to develop direct soiling measurement equipment. All soiling measurement equipment proposed thus far is of small-scale geometry (typically less than 1 m^2), while solar fields can be on the order of kilometers. For this reason, even accurate soiling measurements must be translated to losses that are equivalent for the PV system. Consequently, one must have parallel high-quality soiling measurements and PV time series data (for extracting soiling losses) in order build and validate all the necessary models.

The work presented here is focused specifically on improving the algorithms and models for extracting soiling loss information from time series PV power data. Each day in the time series is first converted to a daily performance index (PI) by dividing the cumulative daily energy produced by the expected or modeled daily energy produced,[27] where a PI value of 1 is expected when no soiling is present. Methods to extract soiling losses from PV PI data generally are based on early work by Kimber et al.[24] that assume that the embedded soiling signal follows a sawtooth pattern, that is, a rainfall or other cleaning causes and abrupt upward shift in the daily PI, followed by a linear decline in the PI due to the solar panels getting dirty each day until the next cleaning event and the start of another sawtooth. Figure 1 provides an example soiling data that generally follows a sawtooth pattern (shown in red) with the exception of a multimonth period in the winter of 2018–2019 where frequent rainfall (shown with green dashed lines) prevents substantial soiling from occurring therefore the red fit remains flat at 1.

Mejia et al.[25] tested slight variation in fit methods as compared with Kimber but concluded that using a linear assumption for soiling periods was acceptable. Both Mejia and Kimber divided the data into soiling intervals based on rain events and made subjective decisions on the rain threshold used to divide the data set; therefore, the extraction process was not automated. In Figure 1 there are two upward shifts or cleaning events in the PI (July 2018 and August 2019) that do not coincide with a recorded rainfall. These cleanings could be due to unrecorded rain or other types of cleanings such as sufficient dew runoff, wind cleaning, or a manual cleaning by the O&M staff. Deceglie et al.[27] introduced an algorithm named the stochastic rate and recovery (SRR) method that avoided use of rainfall data and instead detected cleaning events based on daily upward steps in the rolling median that were greater than $Q3 + \alpha(Q3–Q1)$, where $Q3$ and $Q1$ are the third and first quartiles ($Q3–Q1$ is known as the interquartile range [IQR]), respectively, and α is a user-adjustable multiplier. Furthermore, Deceglie fit the soiling intervals using the Theil Sen estimator, rather than the least squares method,[27] to reduce sensitivity to outliers. Deceglie also applied a Monte Carlo approach to report statistical fitting uncertainty of the soiling intervals. This algorithm is available via the National Renewable Energy Laboratory's (NREL's) RdTools open-source repository, so others in the PV industry can test the method on their own PV data sets.[33] Skomedal et al.[28] applied a version of SRR to three PV systems in the Middle East where manual cleanings were known from O&M reports and rainfall data were available to visually identify natural cleanings. The SRR method was reported as valuable for soiling trend extraction with the following caveats: (1) the threshold for detecting cleanings should be adjusted based on the amount of noise in the data set, (2) the SRR's accuracy has significant dependence on the noise in the underlying PV time series data, and (3) implementing filtering to reduce PI noise

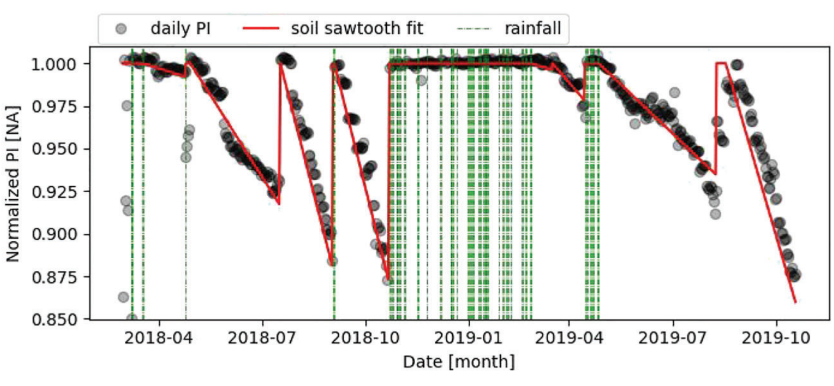

FIGURE 1 PI data that demonstrates the assumption that soiling can be modeled per a sawtooth pattern (given by the solid red lines). The winter of 2018–2019 has frequent rainfall (dashed green lines) and therefore little to no soiling occurs in this time period. Upward shifts in the PI (vertical red lines) are aligned with rainfall in all but 2 days (July 2018 and August 2019) where these days are assumed to have manual cleaning of the system [Colour figure can be viewed at wileyonlinelibrary.com]

improved SRR results. In a later article, Skomedal and Deceglie[29] presented the combined degradation and soiling (CODS) algorithm, with the intent of extracting degradation and soiling trends from a PI time series using a single algorithm. CODS assumes that the PI signal can be decomposed into distinct soiling, degradation, and seasonal signals, with noise converging through an iterative process by minimizing the root-mean square error (RMSE) between the resulting model and the daily PI. The soiling signal derivation is based on the assumption that cleaning events can be detected, and serve as a boundaries for soiling periods, consistent with the other above approaches. Skomedal and Deceglie[29] examined various methods to detect cleaning events using 1000 variations of synthetic data, including change point detection, total variation filtering, shift detection with the Kalman filter, and shifts in the rolling median per the SRR method. The approach per SRR was reported to have the highest mean F_1 score (described in Section 2.2) for detecting cleanings. Furthermore, CODS included applying a Kalman filter to derive the soiling signal between cleanings, which allows for nonlinear or slowly changing soiling rates over each period.

While algorithms for soiling trend extraction have improved or offer new methodologies, there are still several limitations in the existing body of work: (1) validations have been generally limited to synthetic data; (2) field data validation has been observational, and methodologies have had limited comparison against each other; (3) synthetic data sets have not captured many of the complexities of field data (e.g., the field data in Figure 2 shows PI day-to-day variation of 20% [considered noise per soiling fitting] while Skomedal and Deceglie[29] only included 1% to 2% in noise within synthetic data); and (4) each model requires tuning to handle the varying noise and complexity that comes with real PV time series data. In this work, we introduce a dataset based on inverter power from 22 PV systems, labeled for cleaning events based on a predefined set of criteria for soiling and cleaning expectations. The data set is intended to test the accuracy of several cleaning detection algorithms within this work; however, it will also be made open source (https://doi.org/10.21948/1813454) so the larger research community can validate additional algorithms and continue to expand the data set. The paper is arranged as follows: (1) Introduction, (2) Methods: Labeled Data, Accuracy Metrics, and Algorithm Variation, (3) Results, (4) Discussion, and (5) Conclusions.

2 | METHODS: LABELED DATA, ACCURACY METRICS, AND ALGORITHM VARIATION

2.1 | Labeled data

The first focus of this work is to provide a labeled data set of normalized daily PIs for multiple PV systems, where a Boolean label indicates if a cleaning occurs or not on each day. The data set is intended for testing within this work but is also made publicly available for model validation by the broader PV community. Data streams were selected from the NREL's PV Fleets project,[34] where fleet owners across the United States share time series data for the purpose of studying broader fleet-level performance characteristics such as degradation rates and regional variation in soiling losses. The base data consists of 15-min PV power data at the inverter level and corresponding 15-min irradiance data. The irradiance is identified as plane-of-array (POA) measurements for the corresponding inverter, but as the systems are not curated by NREL, there is the expectation that the systems can have varying planar misalignments between the inverter and the POA sensor. All data streams are first passed through a set of quality assurance algorithms as part of the PV Fleets project.[34] Existing algorithms, documented and available per RdTools,[33] are used to generate a daily PI for each inverter time series per the following steps: (1) filter the data to remove unphysical and erroneous data, as well as conditions where inverter power clipping is occurring; (2) normalize each data point by dividing measured power by modeled power, where the model accounts for both irradiance and temperature variation; and (3) aggregate the 15-min normalized data into a single daily PI. There are countless PV models of varying complexity that can be applied when generating a daily PI. Generally, the more that is known about a system and included in the underlying model, the less noise there will be in the time series of a daily PI. While this approach is ideal, often the details of a system are unavailable; therefore, subsequent algorithms such as soiling extraction are best designed to be robust to noise or error that can result from simple or incorrect models.

PV systems were selected based on the following criteria: (1) More than 2 years of data were available; (2) if seasonal variation was present in the data (sinusoidal behavior with a period of 365 days), the range of variation (two times the amplitude) must be

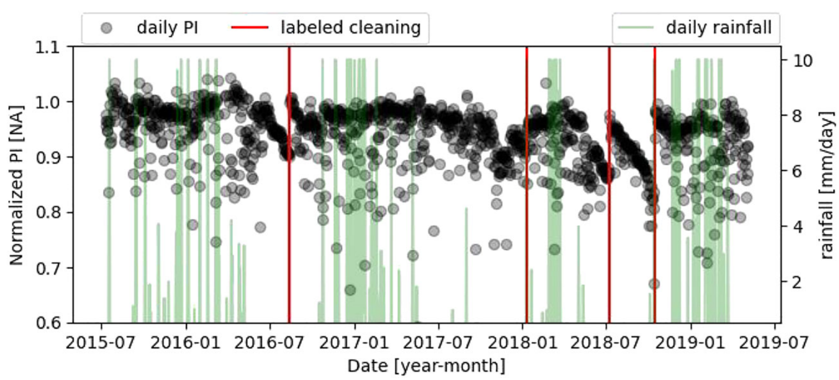

FIGURE 2 A labeled data set where few distinct soiling trends are identifiable. This data set was chosen for both low soiling and increased noise (the range of scatter in the daily PI is approximately 20%, especially in rainy periods). Note that daily rainfall (green data) is capped at 10 mm/day to simplify graph readability [Colour figure can be viewed at wileyonlinelibrary.com]

less than or equal to 15%; (3) there is sufficient remaining data after filtering inverter clipping; and (4) there are visually detectable cleaning instances in the time series. The final 22 systems include a range of extremes in soiling and represent climates with both regular and irregular rainfall. Once the set of PV systems was selected, each time series was evaluated independently by three analysts against the following rules to identify cleaning events. Each analyst completed independent labeling. Discrepancies between the labeled data sets were evaluated jointly by the analysts, and a final determination was made for all 22 data streams.

The overall goal of the data set labeling, where a True Boolean indicates a cleaning event, is to identify upward shifts that are suspected system cleaning events (natural rain or manual cleaning). All other dates where a suspected cleaning event does not take place are labeled as False, including upward shifts that are inconsistent with the expectations for a cleaning event.

1. Shifts must be greater than 1%. This threshold must also be true when considering the scatter or noise in the PI for at least a week before and after the potential shift. See Figure 3a for an example of 1% shift that is labeled as a cleaning while the 1% in Figure 3b is not labeled as a cleaning because the day-to-day scatter outside the red vertical bars is greater than 1%.
2. Extreme outliers are excluded when visually examining the data for trends. Statistically, this typically corresponds to data in excess of ±3 standard deviations from the mean. Prior to identifying a cleaning event, the data must be examined over multiple time intervals: (1) the entire time series, (2) 1 year, (3) 3–4 months, and (4) 1 month or less.
3. No cleaning event is labeled if the previous two or more weeks of data are missing.
4. If multiple upward shifts occur in sequence over several days, but less than a week, and the trends both before and after this week indicate a multiday cleaning event, all days in the group are labeled as cleaning. For test purposes, this is considered one cleaning event, and unique cleaning events must have at least 1 week of separation. Additionally, a low data point between 2 days that each meet cleaning criteria should also be labeled as True so that the "single" cleaning event would be grouped together as a consecutive period. October of 2018 in Figure 4 provides an example a grouped cleaning event while Figure 5 provides a zoomed image of this event.
5. If the analyst examining a potential shift has reasonable doubt that the shift is not due to cleaning, then that date is not labeled as a cleaning event.
6. The analysts were instructed to be cautious (default to a False Boolean) when labeling days during periods of upward seasonality in the data set. For example, there are multiple cases where the PI shows significant linear recovery from January to July, and due to this type of trend, several days of high noisy data can look like a cleaning if taken out of context of the multimonth trend.
7. The graphic time series examination includes both the PI and daily rainfall above 0.75 mm/day per the PRISM database.[35] The analysts were instructed to only use rain as additional information in confirming or denying shifts in marginal cases. For example, if an upward shift was observed to be slightly above the day-to-day scatter in the PI and rain also occurred in conjunction with the shift, the rain was considered extra confidence that the shift was real and the data point was labeled as a cleaning event.

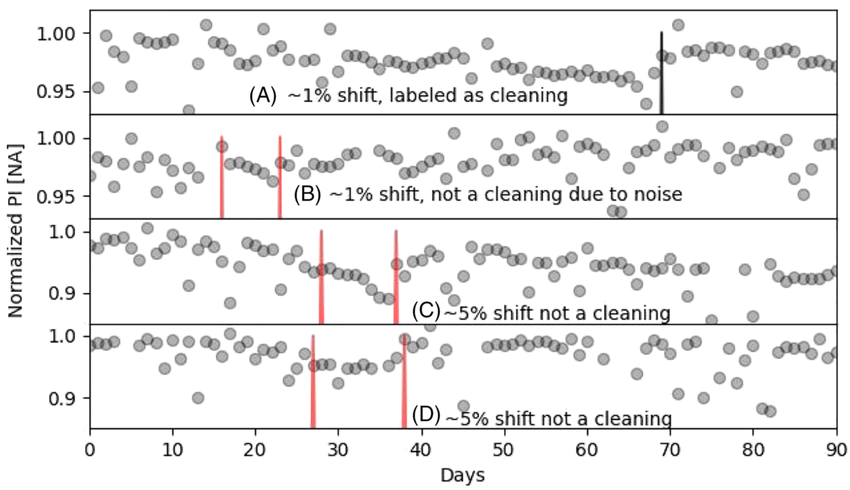

FIGURE 3 Examples of decisions in the data labeling process: (a) 1% shift near Day 70 is labeled a cleaning because the multiple weeks both before and after the shift have minimal scatter and support a real performance improvement. (b) Red bars near Days 17 and 24 show a downward trend that could be a soiling period followed by a 1% recovery. The data before Day 17 and after Day 24 show %5 daily scatter, and therefore, Day 24 is not labeled as a cleaning. (c) Red bars near Days 28 and 37 show a downward trend that could be soiling period followed by a 5% recovery. No rain was reported near Day 37 and the entire 90-day period shows a steady soiling trend if this short period is excluded. Therefore, day 37 is not labeled as a cleaning. (d) The entire 90 days except for the period between the red vertical bars shows PI scatter just below 1. No rain was reported near Day 38; therefore, Day 38 was not labeled as a cleaning [Colour figure can be viewed at wileyonlinelibrary.com]

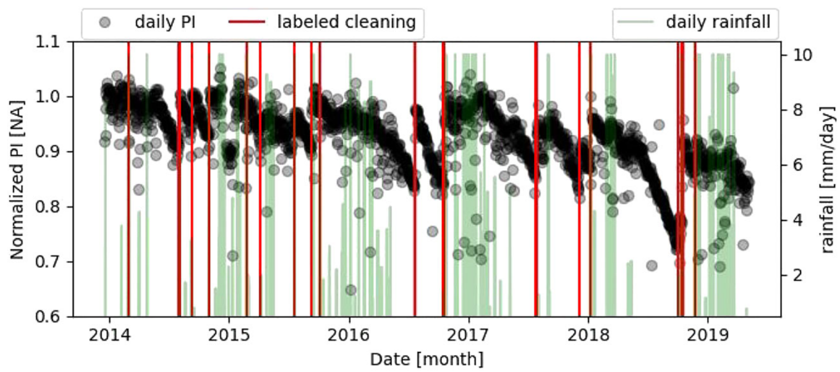

FIGURE 4 A labeled data set with regular soiling trends that in most cases result in 10% to 15% losses. Note that daily rainfall (green data) is capped at 10 mm/day to simplify graph readability [Colour figure can be viewed at wileyonlinelibrary.com]

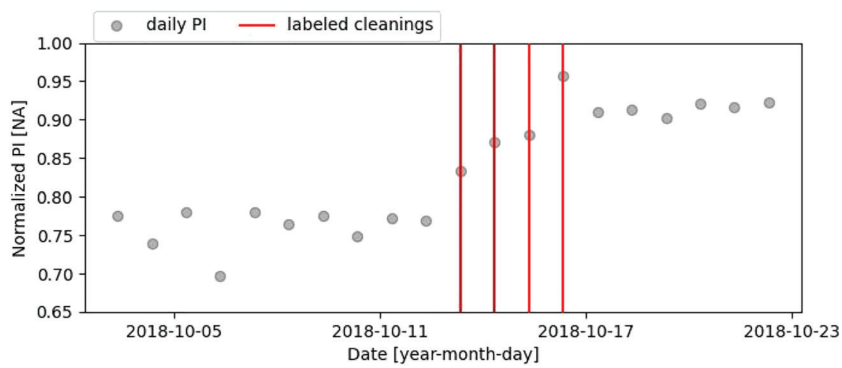

FIGURE 5 The PI improves from 0.77 to 0.92 over the 4-day period beginning on October 13 and meets all criteria to be a cleaning event. Because it is not clear that cleaning occurs on just 1 day, all days identified per the red vertical bars are labeled as cleaning and are grouped as a single cleaning event per F_1 score purposes [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 3c,d provides further examples of questionable cases where the analysts applied caution per Rule 6 (no cleaning). In Figure 3c, the PI is gradually declining from 1 to 0.92 over 90 days while there is a sharper decline for 1 week between the two red vertical lines. While there is an upshift of 5% at the end of this week, there was no recorded rainfall on this date, and this particular week appears unique from the remaining data. In Figure 3d, the PI is scattered just below 1 for the 90 days with the exception of the 11 days between the two vertical red lines. Because there is no distinguishable soiling rate both before and after this period, a cleaning event is considered unlikely compared with alternatives such as a change in a calibration value, a string repair, or removal of weeds or an obstruction causing shading; all of which can cause a sudden upward shift in the data which is otherwise shows flat performance before and after the shift.

We emphasize that the labeling process was conducted visually and independently by soiling experts per the above rules; however, there is room for interpretation, and we have the expectation that future public discourse will be used to help refine the labeled dataset.

2.2 | Accuracy metrics

The primary focus of this work is to test accuracy of different algorithms or variations of algorithms at successfully detecting the labeled cleaning events (true positives [TPs]) while not detecting cleanings on additional days (false positives [FPs]) within the 22 systems describing in Section 2.1. In this context, this is a classification problem, where each individual data point in a time series is labeled as either a cleaning event or not. Performance metrics associated with classification problems include the TP rate or recall, the precision, and the harmonic mean of precision and recall, or the F score. The F score has a maximum value of 1, which occurs when all data labels match their associated ground truth values, and a minimum value of 0 (i.e., all classifications are wrong). The equations for recall, precision, and F score are provided below:

$$\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}), \tag{1}$$

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}), \tag{2}$$

$$F_1 = \mathrm{TP}/(\mathrm{TP} + 0.5(\mathrm{FP} + \mathrm{FN})), \tag{3}$$

where FN is the false negative rate or labeled cleanings that are not detected. We will benchmark model performance using the F score, consistent with Skomedal and Deceglie,[29] as it measures the model's ability to correctly identify cleaning events without falsely classifying additional days. While it is critical to correctly classify true cleanings

(TP), it is unclear if the penalty should be equally weighted for falsely identifying additional cleanings (FP), as the impacts of these additional subdivisions in the data set only become clear in the context of a full soiling extraction model. Considering that TPs may be more important than FPs, we also examine the more general F score, with $\gamma = 2$ to more heavily weight the TP or recall rate ($\gamma = 1$ results in F_1):

$$F_\gamma = \frac{(1+\gamma^2)\mathrm{TP}}{(1+\gamma^2)\mathrm{TP}+\gamma^2\ \mathrm{FN}+\mathrm{FP}}. \quad (4)$$

As noted previously, in some cases, there are multiple cleaning days labeled that fall in a sequence but are considered one cleaning event (see Figure 5). In the calculation of the F score, one TP would be counted if the test algorithm identified a cleaning event within ±1 day of the given group of days. Similarly, if the algorithm identified more than 1 day that aligned with a labeled cleaning group, this was only labeled as a single cleaning event. The same methodology is applied for FPs, meaning if there were adjacent days identified as cleaning by the test algorithm but there was no labeled cleaning, this results in a single FP. In other words, cleaning events consisting of multiple adjacent days were coded to only be counted as a single event, representing a TP or a FP depending on the scenario.

The detection of cleaning events is being tested as an isolated classification problem in this work because it has only been given minor consideration in past works.[24,27–29] While a critical end goal is to evaluate cleaning detection within the overall extraction of accurate soiling metrics, knowledge of cleanings can also be used to better understand if and when dew or wind events correlate with cleanings. The scope was intentionally limited here so that a complete statistical process of cleaning detection could be considered without being impacted by other processes within soiling extraction, for example, soiling rate estimators or consideration of nonlinear soiling events. The intent is to present future work that references the cleaning detection methods and data in this work within the context of the full extraction of soiling metrics.

2.3 | Algorithm variation

2.3.1 | Baseline SRR-IQR

The baseline algorithm examined for cleaning detection is a portion of the SRR algorithm as described in Deceglie et al.[27] The default case of SRR within RdTools uses the following steps to identify cleanings:

1. Handle missing or invalid PI data by forward filling the last valid PI up to a preset number of days (day_scale) into the future.
2. Calculate the centered "day_scale" rolling median of the PI.
3. Calculate the daily change in the rolling median (termed "deltas").
4. Classify cleanings as positive delta > $Q3 + \alpha(Q3\text{–}Q1)$, where the quantiles are calculated for the entire time series of "delta" and ($Q3\text{–}Q1$) is known as the IQR.

RdTools defaults are day_scale = 13 days, and $\alpha = 1.5$. We also examine a slight modification to the baseline SRR; rather than forward filling missing or invalid data up to 13 days, we remove these indices from the data set for step detection. For example, if July 3 and 4 have an invalid PI, these days are removed and the 13-day rolling median is calculated using the surrounding 13 days. If more than 13 days are missing, we set the rolling median to NAN (not a number) for all days in the missing interval.

2.3.2 | Irradiance filtering

Previous work experience by Micheli et al.,[6,17] Muller et al.[18] and Skomedal et al.[28] have documented that the accuracy of calculating daily PI or extracting soiling losses is decreased in low irradiance conditions. Per this previous work, we attempted irradiance filtering of the 15-min time series data prior to calculating the daily PI for each of the 22 systems. Some of the systems had potential timestamp discrepancies between the power and irradiance data, and therefore, filtering at this level did not have the expected results. We therefore filtered at the daily level, removing low irradiance days by filtering out days where the daily sum of irradiance values was less than the 15th percentile of all of the associated daily irradiance sums. Figure 6 shows one data set where the red data are removed per the described irradiance filter. All algorithms were tested with and without irradiance filtering.

2.3.3 | Rolling window outlier filtering

Skomedal et al.[28] suggested benefit from removing outliers from the PI time series by comparing the step change in the PI (both forward and backward) to the standard deviation of the noise, where noise was defined as the difference between the PI and the centered rolling median of the PI. In a similar but more localized approach, a rolling window outlier filter is tested as an alternative to the irradiance filter previously described. The general intent is to remove outliers in the daily PI time series if a data point is substantially shifted from both

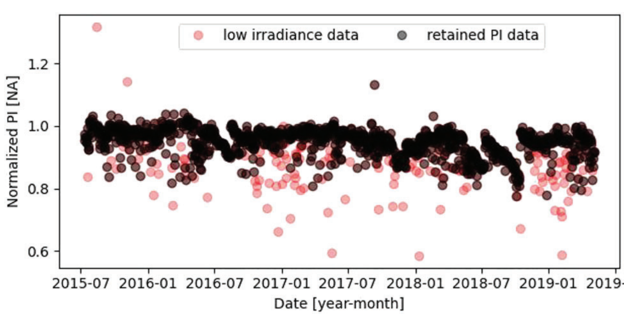

FIGURE 6 An example of removing days (red data) where the irradiance total was less than the 15th percentile of the daily irradiance totals for the entire data set [Colour figure can be viewed at wileyonlinelibrary.com]

the 7 days before and 7 days after. This is mathematically achieved by (1) calculating both the backward and forward 7-day rolling median (with a minimum window size of 5 days data); (2) calculating the percent difference between the current data point and the 7-day forward rolling median and 7-day backward rolling median, respectively; and (3) removing the data point if it is not within 3% of the backward or forward rolling median. An example of the rolling window filter is given in Figure 7 for the same data set that was used as an example for the irradiance filter in Figure 6. Similar to the irradiance filter, all algorithms are tested both with and without applying the PI rolling window filter prior to cleaning detection.

2.3.4 | Median absolute deviation (SSR-MAD) algorithm variation

PI time series data often showing varying noise levels and daily soiling rates from year to year, winter to summer, or overcast to sunny days. For this reason, an alternate cleaning detection threshold is considered that is based on localized variation in the PI time series. Missing data and the calculation of both the rolling median and the deltas are calculated as described in Section 2.3.1. Alternatively, in order to determine the threshold where a positive delta is determined to be a cleaning, the following approach is taken: (1) The absolute value of the day-to-day deviation in the PI is calculated (absolute value of the deltas), (2) the 40-day centered rolling median is calculated for the these absolute deviations (median absolute deviation or SRR-MAD), and (3) a data period is classified as a cleaning event when *deltas* > β *MAD*, where β is a preset multiplier that allows tuning of the cleaning detection.

2.3.5 | Change point detection algorithm

The Python Ruptures library, which provides multiple off-line change point detection algorithms,[36] was directly applied to the rolling median of the PI. Specifically, window segmentation was used, which applies two adjacent sliding windows along the data set and looks for a discrepancy between the windows based on the model applied and the associated cost function. Specifically, the radial basis function (RBF) was used as the model for sliding window segmentation. Within this framework, a penalty value can be set, where the penalty impacts the threshold required for a change point to be identified. When tuning algorithm parameters, both the window width (number of days in the sliding window) and the penalty were varied to optimize the F_1 score on the 22 PV systems while the default of "None" was applied for both "Epsilon" and the prescribed number of break points. Ruptures does not provide a direct way to only consider positive change points (cleanings), and therefore, the library was applied as described and negative change points were rejected in a postprocessing step.

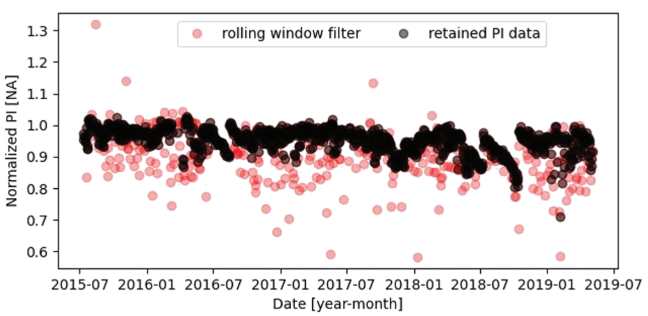

FIGURE 7 An example of removing days (red data) where the daily PI is not within 3% of the either the 7-day forward or the 7-day backward rolling median of PI [Colour figure can be viewed at wileyonlinelibrary.com]

2.3.6 | Algorithm parameter variation

To determine the algorithm with the best performance at detecting cleaning periods, as labeled in the data set, multiple parameters were varied, and the overall F_1 score was compared for each parameter combination. Varied parameters included:

- day_scale or the number days for the rolling median of PI;
- α, the IQR multiplier in the SRR shift threshold;
- β, the multiplier in the MAD shift threshold; and
- the window width and penalty value when using Ruptures-Window algorithm with the "RBF" model.

The complete list of algorithm variation and parameter testing is provided in Table 1. It should be made clear that all combinations of applied parameter and filter variations as given in Table 1 were tested. For example, the SSR-MAD cleaning detection algorithm was tested with each filter option (irradiance, rolling window, and no filter), and within each of these filter options, all combinations of day_scale and β were tested.

3 | RESULTS

3.1 | SRR baseline results and handling missing or filtered data

The SRR algorithm per RdTools[33] defaults to $\alpha = 1.5$ and day_scale $= 13$. Running this baseline case results in a mean F_1 of 0.345 for the 22 labeled data sets. Closer examination of the these results show that 100% of the labeled cleanings are identified for 15 of the systems, and on average, 95% of the labeled cleanings are identified when all 22 systems are considered. The low mean F_1 score of 0.345 is due to a high number of FPs or predicted cleanings that did not actually occur. Figure 8 shows the results of applying the baseline SRR to one system; all labeled cleanings were detected except one in late 2017 (missed by 3 days) while $F_1 = 0.47$ due to the high number of FPs in periods where there is particularly high scatter in the PI data.

TABLE 1 Algorithm and parameter variation

	Cleaning detection algorithm		
Parameter variation or filtering	SRR-IQR threshold	SRR-MAD threshold	Ruptures-Window-RBF
Irradiance filter (on/off)	Yes	Yes	Yes
Rolling window filter (on/off)	Yes	Yes	Yes
day_scale (7–17 odd days)	Yes	Yes	Yes
α (0.5–9 in 0.5 increments)	Yes	NA	NA
β (1.5–2.75 in 0.25 increments)	NA	Yes	NA
Penalty (1–10 in 1 increments)	NA	NA	Yes
Window width (2–5 weeks)	NA	NA	Yes

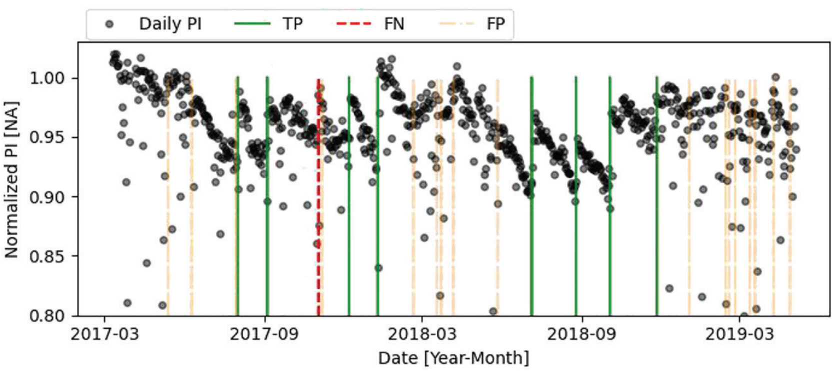

FIGURE 8 The results of detecting labeled cleaning events per the baseline SRR algorithm for one of the 22 systems. Solid green lines are successfully detected cleanings, the dashed red line is a missed cleaning, and in light orange are implied cleanings that are not included in the labeled set [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Optimized parameters and F_1 scores for each of the algorithm variations

	Algorithm variation					
	SRR-IQR		SRR-MAD		Ruptures-Window	
Filter type	Parameters	F_1	parameters	F_1	parameters	F_1
None	DS = 17, α = 7.5	0.63	DS = 13, β = 1.5	0.72	DS = 9, width = 4, penalty = 2	0.55
Rolling window	DS = 13, α = 7.0	0.75	DS = 13, β = 1.75	0.77	DS = 13, width = 2, penalty = 1	0.56
Irradiance	DS = 13, α = 7.0	0.76	DS = 13, β = 1.75	0.79	DS = 9, width = 5, penalty = 2	0.50

Abbreviation: DS, day_scale.

As mentioned in Section 2.3.1, the SRR algorithm handles missing data by forward filling the last valid PI up to day_scale days before calculating the rolling median. The baseline case was run with the alternate option of removing indexes up to 13 days when PI data were missing (as described in Section 2.3.1). The resulting F_1 for this scenario was 0.355. The two filter methods logically result in more data being removed or missing. For this reason, all 378 combinations that are described in Table 1 for both the SRR-IQR and SRR-MAD were run with the baseline forward fill logic, instead of removing gaps in the data up to 13 days. The mean F_1 value for 378 combinations increased by 0.12 for gap removal as compared with forward filling the last valid PI. For this reason, all further results presented apply gap removal, and no data are forward filled. In the event that more than 13 subsequent days are missing, no cleaning or shift detection is allowed immediately following the missing data, consistent with the rules followed when labeling cleanings for the 22 systems.

3.2 | Optimized F_1 results for all variations

Table 2 provides the optimized parameters and the resulting mean F_1 score for the 22 PV labeled data sets for each of the algorithms described. We do not present the complete set of F_1 scores (roughly 1000 combinations considered) because there is little value in comparing each and every combination.

In reviewing the complete set of mean F_1 scores, the following generalized results were found:

1. The mean F_1 score for SRR-IQR improved from 0.36 (day_scale = 13 and $\alpha = 1.5$) to 0.63 (day_scale = 17 and $\alpha = 7.5$). If day_scale was retained at the default of 13 and α was tuned to 6.5, the score improved to 0.59.
2. A day_scale of 13 achieved the optimized results for five of six variations of SRR, only SRR-IQR with no filtering had an increased day_scale of 17.
3. Filtering improved the SSR results for both the IQR or MAD thresholds.
4. Ruptures' sliding window change point detection significantly underperformed compared with all versions of SRR, and no clear set of parameter choices emerged from all combinations tested. In other words, a wide range of values for day_scale, width, and penalty was found to achieve F_1 scores in the 0.5–0.55 range.
5. If day_scale was fixed at 13, the irradiance filter was applied to SRR-IQR, and the best α was chosen for each of the 22 systems, the mean $F_1 = 0.82$ (compared with 0.76 for α fixed at 7.0).
6. If day_scale was fixed at 13, the irradiance filter was applied to SRR-MAD, and the best β was chosen for each of the 22 systems. The resulting mean F_1 was 0.82 (compared with 0.79 for β fixed at 1.75).
7. The range and median F_1 results for the best two algorithms per Table 2 are as follows: SRR-MAD-Irradiance range = (0.57–1), median = 0.77, SRR-IQR-Irradiance range = (0.44–0.94), median = 0.76.
8. The generalized F score as described in Section 2.2 (F_γ with $\gamma = 2$) was calculated for all combinations of SRR, as given in Table 1. The optimized F_γ occurs at reduced multipliers of IQR and MAD, as well as a day_scale value of 9 or 11 days. Considering that FPs are weighted less than TPs, these results are statistically consistent.

Figure 9 provides the results for the best performing algorithm, SRR-MAD-Irradiance, for the same system shown in Figure 8. In this case $F_1 = 0.94$, with no FPs and only one missed labeled cleaning.

4 | DISCUSSION

Significant improvement over baseline was achieved at detecting cleanings in 22 labeled data sets, where a 43% absolute improvement was achieved in the mean F_1 score (0.36 to 0.79) when going from the default SRR approach in RdTools to the optimized SRR-MAD with irradiance filtering. Similarly, a 40% absolute improvement was achieved in the mean F_1 score (0.36 to 0.76) when the irradiance filter was applied to a tuned version of SRR. Filtering the PI data (irradiance or rolling window) before step detection (MAD or IQR thresholds) also resulted in consistently improved F_1 scores. This agrees with Skomedal et al.[28] and previous experience of the authors. While the rolling median approach was designed into SRR because it is robust against outliers, multiple outliers can occur in the rolling period and still can impact shift detection. Filtering therefore can be coupled with the rolling median to increase the ability to reject outliers. Table 2 shows that the irradiance filter outperforms the rolling window filter, but only by 1% to 2% absolute, depending on the model implemented. High-quality irradiance data are not always available; consequently, applying the rolling window filter is useful when irradiance data are not available.

The data in Table 2 also demonstrate that the SRR-MAD outperforms the SRR-IQR for each filter category. This is likely due to the 40-day rolling aspect of the SRR-MAD threshold, as compared with SRR-IQR being calculated for the entire time series. As discussed in Section 2.3.4, the noise in the PI can vary across the time series, and the SRR-MAD cleaning detection threshold is adjusted according to the local 40-day noise while the threshold is fixed for the entire data set per the current SRR-IQR approach. It can also be said that varying the shift threshold as is done with SRR-MAD (this could also be adopted with SRR-IQR) results in a model that is less dependent on tuning. This is demonstrated in that tuning β in SRR-MAD for each of the 22 PV systems resulted in only a 3% absolute improvement in the mean F_1 score beyond the best result in Table 2 while tuning α in SSR-IQR resulted in a 6% absolute increase in mean F_1 beyond the maximum value in Table 2.

The one change point detection algorithm that was tested, the Ruptures package's window segmentation algorithm with the "RBF" model, performed poorly at detecting labeled cleanings as compared with SRR-IQR and SRR-MAD. Furthermore, of the hundreds of variations tested for the change point algorithm, no patterns in parameter tuning, number of days for the rolling median, or filtering options stood out when detecting cleanings. There are various other change point algorithms that may have potential, but the results here are not

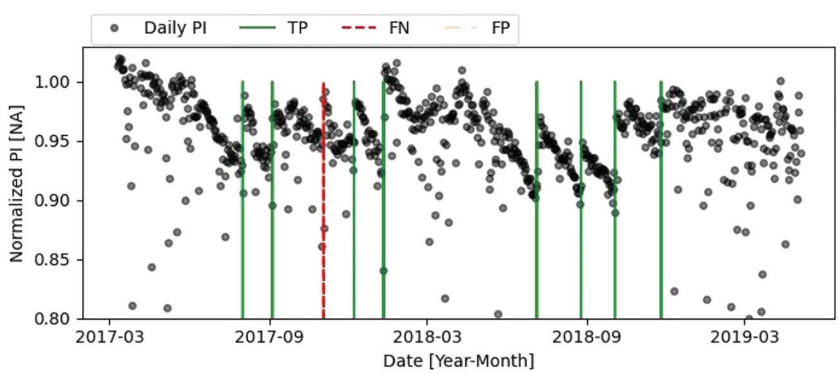

FIGURE 9 The same system as shown in Figure 8 but the shift detection results are for the SRR-MAD-Irradiance instead of the baseline SRR approach [Colour figure can be viewed at wileyonlinelibrary.com]

promising. The poor results could be due to that change point algorithms have been historically designed to detect any statistically significant change in the data set per a given cost function. PV data sets often have various embedded shifts while cleaning events are one specific subset of these shifts. Existing change point algorithms were not written to handle only a subset of significant shifts, and therefore, the results found in this work are not surprising.

Significant improvement has been demonstrated in the detection of cleanings within labeled PV data sets, but these results should also be considered within the context of complete soiling extraction algorithms. Recall that 95% of the labeled cleanings for all 22 data sets were identified by the default version of SRR within RdTools. At the same time, the F_1 score was only 0.36 due to a high number of FPs (extra cleanings identified that were not labeled as such). This result leads to the following questions:

1. Were these FPs actually minor cleanings that were not identified due to the labeling rubric, or due to lack of knowledge of the true soiling signal?
2. What is the consequence of these FPs on soiling rate and soiling loss extraction?

The first question cannot be answered within the context of the current data, but synthetic data sets or future data sets accompanied with soiling stations may provide answers. The second question can be explored within the context of complete soiling extraction algorithms; this will be the next step beyond this work. For example, SRR uses identified cleanings to subdivide the data set into fitting periods for soiling rates; therefore, additional FPs imply shorter soiling intervals. Shorter intervals could lead to improved soiling rate fits, but at the same time, this has potential to bias overall soiling losses to lower values.

5 | CONCLUSION

Cleaning events were manually labeled for inverter time series data from 22 PV systems per a fully documented process. The data sets were carefully selected to include varying degrees of soiling, cleaning events, and noise. The RdTools-SRR algorithm served as a baseline for detecting the prelabeled cleanings while several methods were examined to improve cleaning detection, including (1) optimization of SRR user parameters (SRR-IQR), (2) an alternate statistical method for distinguishing cleaning events (SRR_MAD), (3) the application of the Ruptures-Window change point detection to determine cleaning events, and (4) the application of noise and irradiance filters on the time series data in conjunction with the above methods. The baseline RdTools-SRR resulted in a mean F_1 score of 36% while the F_1 score improved to 55%, 63%, and 72% for Ruptures-Window, SRR-IQR, and SRR-MAD, respectively. Furthermore, applying irradiance or noise filtering continued to significantly improve both the SRR-IQR and SRR-MAD approaches. SRR-MAD in combination with irradiance filtering achieved the most accurate detection of the cleaning events within the 22 data sets, with a mean F_1 score of 79% and a 43% improvement over the baseline RdTools-SRR algorithm. While these results demonstrate that cleanings can be more accurately detected, future work must consider the impact of cleaning detection methodology within complete algorithms for extracting critical soiling rate and soiling loss metrics.

ACKNOWLEDGEMENTS

The authors would like to thank Kevin Anderson for his support in extracting data from the NREL PV Fleets project and Michael Deceglie for his work in making updates to the RdTools-SRR algorithm that were relevant to this work. This work was authored in part by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the US Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding was provided by the DOE's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Numbers 34348 and 35893. The views expressed in the article do not necessarily represent the views of the DOE or the US Government. The US Government retains, and the publisher, by accepting the article for publication, acknowledges that the US Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for US Government purposes. Leonardo Micheli was funded through the European Union's Horizon 2020 research and innovation program under the NoSoilPV project (Marie Skłodowska-Curie Grant 793120). Eduardo F. Fernández thanks the Spanish Ministry of Science, Innovation and Universities (RYC-2017-21910).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Duramat at https://datahub.duramat.org/project/example-data.

ORCID

Matthew Muller https://orcid.org/0000-0002-3835-4096
Leonardo Micheli https://orcid.org/0000-0001-7986-7560

REFERENCES

1. Costa S, Diniz A, Kazmerski L. Solar energy dust and soiling R&D progress: literature review update for 2016. *Renew Sustain Energy Rev*. 2018;82:2504-2536. doi:10.1016/j.rser.2017.09.015
2. Costa S, Diniz A, Kazmerski L. Dust and soiling issues and impacts relating to solar energy systems: Literature review update for 2012–2015. *Renew Sustain Energy Rev*. 2016;63:33-61. doi:10.1016/j.rser.2016.04.059
3. Toth S, Hannigan M, Vance M, Deceglie M. Enhanced photovoltaic soiling in an Urban environment. In: *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*. IEEE; 2019:2904-2907.
4. Boyle L, Flinchpaugh H, Hannigan M, Ambient airborne particle concentration and soiling of PV cover plates. *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*. IEEE. 2014:3171-3173. doi:10.1109/PVSC.2014.6925609
5. Boyle L, Flinchpaugh H, Hannigan M. Assessment of PM dry deposition on solar energy harvesting systems: measurement–model comparison. *Aerosol Sci Tech*. 2016;50(4):380-391. doi:10.1080/02786826.2016.1153797

6. Micheli L, Muller M. An investigation of the key parameters for predicting PV soiling losses. *Prog Photovolt Res Appl*. 2017;25(4):291-307. doi:10.1002/pip.2860
7. Micheli L, Deceglie MG, Muller M. Predicting photovoltaic soiling losses using environmental parameters: an update. *Prog Photovoltaics Res Appl*. 2019;27(3):210-219. doi:10.1002/pip.3079
8. Micheli L, Muller M, Kurtz S, Determining the effects of environment and atmospheric parameters on PV field performance. *2016 IEEE 43rd Photovoltaic Specialist Conference (PVSC)*, IEEE; 2016, 1724-1729. doi:10.1109/PVSC.2016.7749919
9. Javed W, Guo B, Figgis B. Modeling of photovoltaic soiling loss as a function of environmental variables. *Sol Energy*. 2017;157(August):397-407. doi:10.1016/j.solener.2017.08.046
10. Goossens D, Offer ZY, Zangvil A. Wind tunnel experiments and field investigations of eolian dust deposition on photovoltaic solar collec- tors. *Sol Energy*. 1993;50(1):75-84. doi:10.1016/0038-092X(93)90009-D
11. Fernández EF, Chemisana D, Micheli L, Almonacid F. Spectral nature of soiling and its impact on multi-junction based concentrator systems. *Sol Energy Mater Sol Cells*. 2019;201:110118. doi:10.1016/j.solmat.2019.110118
12. Ilse K, Figgis B, Naumann V, Hagendorf C, Bagdahn J. Fundamentals of soiling processes on photovoltaic modules. *Renew Sustain Energy Rev*. 2018;98(January):239-254. doi:10.1016/j.rser.2018.09.015
13. Gostein M, Caron JR, Littmann B, Measuring soiling losses at utility-scale PV power plants. *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*. IEEE. 2014:885-890. doi:10.1109/PVSC.2014.6925056
14. Gostein M, Düster T, Thuman C, Accurately measuring PV soiling losses with soiling station employing module power measurements. *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*. 2015:1-4. doi:10.1109/PVSC.2015.7355993
15. Gostein M, Faullin S, Miller K, Schneider J, Stueve B, Mars Soiling Sensor™, *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)*. IEEE, 2018:3417-3420. doi:10.1109/PVSC.2018.8547767
16. Valerino M, Bergin M, Ghoroi C, Ratnaparkhi A, Smestad G. Low-cost solar PV soiling sensor validation and size resolved soiling impacts: A comprehensive field study in Western India. *Sol Energy*. 204:307-315. doi:10.1016/j.soler.2020.03118
17. Micheli L, Ruth D, Deceglie M, Muller M. *Time Series Analysis of Photovoltaic Soiling Station Data: Version 1.0*. National Renewable Energy Laboratory (NREL); 2017.
18. Muller M, Micheli L, Martinez-Morales AA, 2017. A method to extract soiling loss data from soiling stations with imperfect cleaning schedules. *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*. IEEE 2017:2881-2886. doi:10.1109/PVSC.2017.8366214
19. Muller M, Micheli L, Solas A, et al. An in-depth field validation of "DUSST": a novel low-maintenance soiling measurement device. *Prog Photovoltaics Res Appl*. 2021;29(8):953-967. doi:10.1002/pip.3415
20. Fernandez-Solas A, Micheli L, Muller M, Almonacid F, Fernandez E. Design, characterization and indoor validation of the optical soiling detector "DUSST". *Sol Energy*. 211:1459-1468. doi:10.1016/j.solener.2020.10.028
21. Micheli L, Fernández EF, Muller M, Smestad GP, Almonacid F. Selection of optimal wavelengths for optical soiling modelling and detection in photovoltaic modules. *Sol Energy Mater Sol Cells*. 2020;212:110539. doi:10.1016/j.solmat.2020.110539
22. Curtis T, Tatapudi S, TamizhMani G, Design and operation of a waterless PV soiling monitoring station, *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*, IEEE, 2018:3407-3412. doi:10.1109/PVSC.2018.8547820
23. Korevaar M, Mes J, Nepal P, Snijders G, van Mechelen X, Novel soiling detection system for solar panels. *33rd European Photovoltaic Solar Energy Conference and Exhibition*, 2017:2349-2351. doi:10.4229/EUPVSEC20172017-6BV.2.11
24. Kimber A, Mitchell L, Nogradi S, Wenger H. The effect of soiling on large grid-connected photovoltaic systems in California and the southwest region of the United States. In: *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*. IEEE; 2006:2391-2395.
25. Mejia FA, Kleissl J. Soiling losses for solar photovoltaic systems in California. *Sol Energy*. 2013;95:357-363. doi:10.1016/j.solener.2013.06.028
26. Besson P, Munoz C, Ramirez-Sagner G, Salgado M, Escobar R, Platzer W. Long-term soiling analysis for three photovoltaic technologies in Santiago Region. *IEEE J Photovolt*. 2017;7(6):1755-1760. doi:10.1109/JPHOTOV.2017.2751752
27. Deceglie M, Micheli L, Muller M. Quantifying soiling loss directly from PV yield. *IEEE J Photovolt*. 2018;8(2):547-551. doi:10.1109/JPHOTOV.2017.2784682
28. Skomedal Å, Haug H, Marstein ES. Endogenous soiling rate determination and detection of cleaning events in utility-scale PV plants. *IEEE J Photovolt*. 2019;9(3):858-863. doi:10.1109/JPHOTOV.2019.2899741
29. Skomedal Å, Deceglie MG. Combined estimation of degradation and soiling losses in photovoltaic systems. *IEEE J Photovolt*. 2020;10(6):1788-1796. doi:10.1109/jphotov.2020.3018219
30. Micheli L, Fernandez EF, Muller M, Almonacid F. Extracting and generating PV soiling profiles for analysis, forecasting, and cleaning optimization. *IEEE J Photovolt*. 2020;10(1):197-205. doi:10.1109/JPHOTOV.2019.2943706
31. Micheli L, Muller M, Fernández EF, Almonacid FM. Change point detection: an opportunity to improve PV soiling extraction. In: *IEEE 45th Photovoltaic Specialist Conference (PVSC)*. IEEE; 2020.
32. Ilse K, Micheli L, Figgis B, et al. Techno-economic assessment of soiling losses and mitigation strategies for solar power generation. *Joule*. 2019;3(10):2303-2321. doi:10.1016/j.joule.2019.08.019
33. RdTools, NREL/rdtools: Version 2.1.0-beta.1, 2020. doi:10.5281/ZENODO.4307010
34. Deline C, White R, Muller M, Anderson K, Perry K, Deceglie M, Simpson L, Jordan D, PV fleet performance data initiative program and methodology, *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. IEEE. 2020:1363-1367. doi:10.1109/pvsc45281.2020.9300583
35. PRISM (2020) Climate Group—Oregon State University, PRISM Gridded Climate Data (AN81d). http://www.prism.oregonstate.edu/explorer/
36. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. *Signal Process*. 2020;167:107299. doi:10.1016/j.sigpro.2019.107299

How to cite this article: Muller M, Perry K, Micheli L, Almonacid F, Fernández EF. Automated detection of photovoltaic cleaning events: A performance comparison of techniques as applied to a broad set of labeled photovoltaic data sets. *Prog Photovolt Res Appl*. 2022;1-11. doi:10.1002/pip.3523