

# Improved PV Soiling Extraction through the Detection of Cleanings and Change Points

Leonardo Micheli, Marios Theristis, Andreas Livera, Joshua S. Stein,  
George E. Georghiou, Matthew Muller, Florencia Almonacid, Eduardo F. Fernández

**Abstract**—Photovoltaic (PV) soiling profiles exhibit a sawtooth shape, where cleaning events and soiling deposition periods alternate. Generally, the rate at which soiling accumulates is assumed to be constant within each deposition period. In reality, changes in rates can occur due to sudden variations in climatic conditions; e.g. dust storms or prolonged periods of rain. The existing models used to extract the soiling profile from PV performance data might fail to capture the change points and occasionally estimate incorrect soiling profiles. This work analyzes how the introduction of change points can be beneficial for soiling extraction. Data from nine soiling stations and a 1 MW site were analyzed by using piecewise regression and three change point detection algorithms. The results showed that accounting for change points can provide significant benefits to the modelling of soiling even if not all the change point algorithms return the same improvements. Considering change points in historical trends is found to be particularly important for studies aiming to optimize cleaning schedules.

**Index Terms**— Monitoring, Photovoltaic Systems, Regression Analysis, Soiling, Time-series Analysis.

## I. INTRODUCTION

Soiling consists of the accumulation of dust, dirt and contaminants on the surface of photovoltaic (PV) modules and causes significant losses to PV systems worldwide [1], [2]. Indeed, the layer of soiling absorbs, deflects, and reflects part of the incoming sunlight, reducing the amount of energy that the PV cell can convert into electricity [3].

A number of factors affects the soiling deposition and mitigation [4]–[6]. Generally, soiling deposits on the module surface during dry spells and can be removed either artificially or naturally [7]. For these reasons, the soiling loss profile is generally modelled as a sawtooth wave, caused by the alternation of cleaning events and deposition periods [8].

This work was supported by the European Union’s Horizon 2020 research and innovation programme under the NoSoilPV project (Marie Skłodowska-Curie Grant agreement no. 793120). The work of Marios Theristis, and Joshua S. Stein was supported by the U. S. Department of Energy’s Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 34366. The work of Andreas Livera, George E. Georghiou, Florencia Almonacid, and Eduardo F. Fernández was supported by the solar-era.net project ROM-PV. Part of the work was authored by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract DEAC3608GO28308. Matthew Muller’s work was supported by the DOE’s Office of Energy Efficiency and Renewable Energy under Solar Energy Technologies Office Agreements 34348. (Corresponding author: Leonardo Micheli, [lmicheli@ujaen.es](mailto:lmicheli@ujaen.es)).

Rainfall is the most common natural cleaning event [9]. In most cases, soiling is assumed to accumulate at a fixed rate (i.e. linear) during the deposition periods [10]–[13], whereas some authors have suggested exponential soiling deposition models [14].

Monitoring soiling is essential to better mitigate its effect on the energy production. Generally, this is performed through soiling stations or detectors, but the soiling profiles can also be “extracted” directly from PV performance data, without the need of installing any hardware. Such soiling extraction models are able to determine the cleaning frequencies and the severity of the soiling deposition rates throughout the various periods of the time-series. For this reason, they can actually be employed on both soiling measurement devices and PV performance data to analyze seasonality and variability of historical soiling trends.

Two soiling extraction models have been presented in literature; the fixed rate precipitation (FRP) model [10] and the stochastic rate and recovery (SRR) model [12]. The FRP model, which requires a performance metric and the rainfall pattern as inputs, generates a soiling profile by applying a fixed soiling rate to all dry periods [10]. The SRR model calculates the soiling rates between each pair of consecutive cleaning events and generates a number of potential soiling profiles by using a Monte Carlo simulation [12]. This method does not require any weather data as input, since cleanings are identified on the basis of positive shifts on the performance index. This weather-unaware approach has the advantages of not requiring the understanding of natural cleaning mechanisms, nor the knowledge of the operation and maintenance (O&M) schedule, which might be imperfectly reported or difficult to obtain. Indeed, not all rainfall events have cleaning effects. There have been discussions in literature regarding the existence of minimum rainfall thresholds [10], but no consensus has been reached yet on its value. Also, other meteorological parameters,

L. Micheli, E. F. Fernández, and F. Almonacid are with the Centro de Estudios Avanzados en Tierra, Energía y Medio Ambiente (CEACTEMA), University of Jaén, Jaén 23071, Spain (e-mail: [lmicheli@ujaen.es](mailto:lmicheli@ujaen.es); [eduardo.fernandez@ujaen.es](mailto:eduardo.fernandez@ujaen.es); [facruz@ujaen.es](mailto:facruz@ujaen.es)). M. Theristis, and J. S. Stein are with the Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: [mtheris@sandia.gov](mailto:mtheris@sandia.gov); [jstein@sandia.gov](mailto:jstein@sandia.gov)). A. Livera, and G. E. Georghiou are with the Photovoltaic Technology Laboratory and the FOSS Research Centre for Sustainable Energy, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus (e-mail: [livera.andreas@ucy.ac.cy](mailto:livera.andreas@ucy.ac.cy); [geg@ucy.ac.cy](mailto:geg@ucy.ac.cy)). M. Muller is with the National Renewable Energy Laboratory, Golden, CO, 80401 USA, and also with the CEACTEMA of the University of Jaén, Jaén 23071, Spain (e-mail: [matthew.muller@nrel.gov](mailto:matthew.muller@nrel.gov)).

such as wind and dew can naturally clean the PV modules. Therefore, a model based only on precipitation would not be able to detect non-rainfall related cleanings.

Both methods assume a sawtooth waved soiling profile. While this is often an acceptable approximation, significant benefits can be obtained by introducing change point analysis in soiling extraction models [15]. As such, changes in deposition rates, which can be caused by sudden variations in the climatic conditions can be detected; e.g. when a dust storm occurs or a rainy period ends.

Performing change point detection becomes particularly important when a weather-unaware approach is selected. For example, a cleaning-only extraction method would not be able to identify the switch from a rainy period, with no soiling accumulation, to a dry spell, where severe soiling deposition can occur. This means that it would model only a single soiling rate for the whole period, affecting the analysis of soiling seasonality and biasing the identification of the optimal cleaning schedules.

This work builds on a previous study from some of the authors, which demonstrated that change point detection can be beneficial for soiling extraction and analysis [15]. In that work, segmented (or piecewise) regression was employed to identify changes in the soiling deposition rates within dry periods. The segmented regression method required as input the number of change points in each deposition period, which had to be estimated in advance. In order to address this limitation, the present work explores alternative methods for change point detection that can be used for PV soiling analysis studies. Change point detection has been sparsely used in PV. Recently, such methods were used to identify changes in nonlinear degradation profiles [16], [17], while no applications were found for PV soiling analysis.

Compared to the aforementioned study [15], where only the piecewise regression was used to identify change points and fit the deposition rates, three change point identification algorithms were also tested in this work. These make it possible to consider different numbers of change points per deposition period. Furthermore, additional sites were considered in order to assess the change point techniques in a wider variety of soiling conditions and characteristics. This includes also the soiling analysis of a commercial 1 MW system in Southern Spain.

## II. METHODOLOGY

### A. Soiling Data

The soiling profiles measured by 9 soiling stations installed in the U.S. have been used. In particular, the daily soiling ratios and the daily rainfalls shown in Ref. [18] were provided. Soiling is quantified by the soiling ratio, which expresses the ratio of the energy output of a PV device to the energy output that the same PV device would produce if soiling was not present [19]. It has a value of 1 if no soiling losses occur, whereas it decreases otherwise. A second metric, named *soiling rate*, describes the daily derate of the soiling ratio. It has a value of 0%/day if there is no soiling deposition on the modules’ surfaces, otherwise it assumes negative values.

In order to prepare the data for the analysis, two data pre-processing steps were considered (see Fig. 1). Initially, a two-

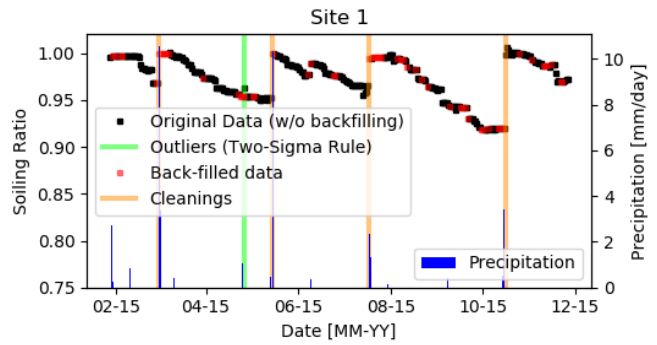


Fig. 1. Soiling profile for Site 1 in California, USA. The green vertical lines mark the outliers identified through the two-sigma rule method, whereas the red markers show the result of the back-filling procedure. Cleaning events detected by the SRR method are marked by orange vertical lines [12]. The rainfall (right y-axis) is indicated by the blue bars.

sigma rule filter was applied to remove outliers: any  $i$ -value outside of two standard deviations of the mean of the values within  $i-7$  and  $i+7$  was considered an outlier ( $i$  is days in this example). Furthermore, several daily soiling ratios were missing due to the irradiance filter ( $\geq 700 \text{ W/m}^2$ ) [20]. For this reason, back-filling of missing data was performed according to the Next Observation Carried Backward (NOCB) method [12].

In addition to the soiling stations, the 2019 DC performance ratio profile of a 1 MW PV system, installed near Granada, Spain, was considered in this work. The system consists of mono-crystalline modules facing south and mounted at a tilt angle of  $30^\circ$ . The energy yield and performance profiles were extracted from the hourly DC power of one of the strings using the methodology described in Refs. [20], [21] and the weather data from Ref. [22]. The site is characterized by a long dry summer, with no rainfalls between April and September. An artificial cleaning was performed on August 5<sup>th</sup> by the site’s O&M team but for the sake of this investigation, it is considered as an additional rain event.

### B. Change Point Classification

Change points are generally classified as “continuous” or “discontinuous” [23]. In the first case, only a change in slope

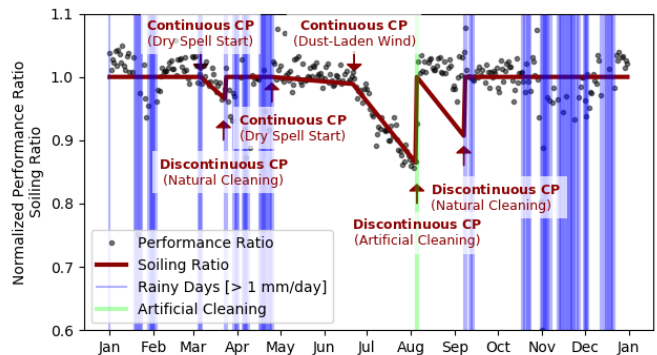


Fig. 2. Daily performance ratio in 2019 for a 1 MW system in Southern Spain. The soiling profile was extracted by considering the rainfall pattern, and a change point on June 22<sup>nd</sup> due to a dust laden wind. The applied methodology is described in Refs [20], [21] and the rainfall data were sourced from MERRA-2 [22]. Continuous and discontinuous change points (CP) are highlighted. Discontinuous change points are due to rainfalls (March 24<sup>th</sup> and September 8<sup>th</sup>) and an artificial cleaning (August 5<sup>th</sup>). Continuous change points are due to the start of dry spells (March 7<sup>th</sup> and April 26<sup>th</sup>) and a dust laden wind (June 22<sup>nd</sup>). Artificial cleanings are assumed to cause a positive shift in soiling ratio, but no change in soiling deposition rate [20], [21].

occurs (e.g. caused by a prolonged rainy period followed by a dry spell) whereas, in the second case, a step is present (e.g. caused by a cleaning event). Examples of the two classes of change points are shown in Fig. 2 for the soiling profile of the 1 MW PV system [20]. This work focuses on continuous change points only, as their identification has not been previously addressed in the literature. Therefore, if not otherwise specified, the term “change point” (CP) in this work refers to a “continuous change point”.

### C. Soiling Profile Extraction

Each soiling station’s time-series was divided in segments delimited by two consecutive events (i.e. cleanings and change points). A cleaning event produces a positive shift in soiling ratio and the time between two consecutive cleanings is called “deposition period”. On the other hand, a change point does not produce any abrupt change in soiling ratio and splits a deposition period in two segments of different slopes (i.e. soiling rates) that join at the change point (i.e. have the same soiling ratio at the change point).

Cleaning dates were determined by using the *pv\_soiling* code [12], [24]. The distribution of the daily variations in soiling ratio was analyzed, and the third quartile ( $Q_3$ ) and the interquartile range (IQR) of the absolute values of the differences between neighbor soiling values were calculated. Cleanings were identified in those dates in which a positive shift in the performance data larger than  $Q_3 + 1.5 \cdot IQR$  occurred. While some authors have suggested site-specific tuning of this equation [13], in this work it has been used in its original version; same for all the sites.

Four different approaches were applied for detecting the change points (more details in Section II.D): segmentation (or piecewise regression), Pruned Exact Linear Time (PELT) algorithm, Facebook Prophet (FBP) algorithm, and the Bayesian estimation of abrupt change, seasonality and trend (BEAST). The last three methods are classified as change point detection algorithms.

In particular, the four methods were used to model the soiling ratio of any deposition period of at least 7 days. While segmentation has been employed to simultaneously detect the change points and fit the data, the other three algorithms were only used to detect the dates in which the change points occurred. The  $m$  number of dates identified for each deposition period were then given as inputs to the *curve\_fit* function in the *SciPy* library [25], which was used to fit the data in each deposition period with an  $m+1$  number of linear functions. Any change point detected within one week of a cleaning event was discarded.

### D. Change Point Detection Algorithms

*Segmented or Piecewise Regression (Segm)*. Compared to linear regression, in segmented or piecewise regression, the data are fit through more than just one line. Similar to the previous work [15], the soiling ratio in each deposition period is divided into two continuous segments:

$$\begin{aligned} \bullet \quad y &= a_0 \cdot x + b && \text{if } x < x_0 \\ \bullet \quad y &= a_0 \cdot x + b + a_1 \cdot (x - x_0) && \text{if } x \geq x_0 \end{aligned}$$

where  $x$  is the date,  $y$  is the soiling ratio,  $a_0$  is the soiling rate in the first segment,  $(a_0 + a_1)$  is the soiling rate in the second segment,  $b$  is the value of the soiling ratio at the start of the

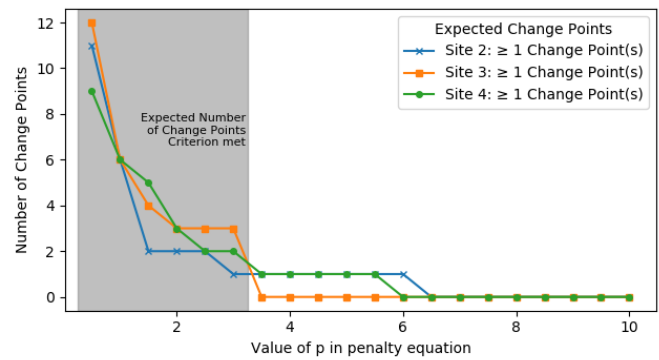


Fig. 3. Number of change points depending on the value of  $p$  in the penalty equation. The grey area marks the conditions in which the number of expected cleaning is respected for the three sites used for tuning.

deposition period and  $x_0$  is the date in which the rate change occurs. The segmented regression equations are defined through the piecewise function in the *NumPy* library [26]. The initial guesses were set to:  $x_0 = 7$  days,  $a_0 = 0$  %/day,  $a_1 = 0$  %/day and  $b = 1$ . In addition, these bounds were chosen  $0 \leq x_0 \leq N$ ,  $a_0 \leq 0$  and  $0 \leq b \leq 1$ , where  $N$  was the total number of days in each deposition period. No bound was set for  $a_1$ . When segmentation is used, change point identification and data fitting occur at the same time.

In the present form, segmented regression can only identify one change point per deposition period. For this reason, three change point algorithms, able to identify a flexible number of changes per period, were also tested.

*Pruned Exact Linear Time (PELT)*. PELT is an offline change point detection algorithm first presented in 2011 [27]. This method is based on the optimal partitioning algorithm [28]. The change points are detected through the minimization of the sum of the penalized cost functions of the segments in which the change points split the data series. PELT improves the computational efficiency of the original method through pruning (i.e. by removing at each iteration non-optimal change-point candidates) [27]. The PELT algorithm, in this work, has been implemented in Python 3.7.0 through the *ruptures* library [29]. It requires a penalty value as an input, which is used to prevent overfitting (e.g., identifying noise as change points) as it increases the cost of adding change points during the cost function minimization process. Higher penalty values decrease the computational cost and the number of detected change points, which might cause underfitting, while lower values are more at risk of overfitting. Its value is not defined *a priori*, and it is actually considered an “open question” by the research community, as it can significantly affect the results [30]. One of the most common approaches is to set the penalty  $\beta = p \cdot \ln(n)$ , where  $p$  represents the number of additional parameters introduced by adding a change point and  $n$  is the number of days in each time-series [31]. The penalty ( $\beta$ ) equation, in the aforementioned form, could have led to data overfitting [32] and, for this reason, the value  $p$  was adjusted by considering the data from three of the tested sites. For each of these sites (Sites 2, 3 and 4), the expected number of minimum change points was visually determined and set to 1. The analysis showed that any  $p \leq 3$  returns at least 1 change point for those three sites (see Fig. 3). As expected, the number of change points increases while the value of penalty decreases due to overfitting. On the other hand, excessively large penalty values

make it impossible to identify change points. Hence, the last value at which the criterion was met was selected ( $p = 3$ ). This value, and the equation used for the penalty, should be the focus of future studies, as those used in this work should not be assumed to be necessarily valid for any soiling study.

**Facebook Prophet Algorithm (FBP).** FBP is an open-source library, available in Python and R, used to forecast time-series based on an additive decomposition model, which combines trend, seasonality and holidays (neglected in this study) [33]. A piecewise linear model is applied by default for the trend whereas seasonality is modelled as an additive component. This method has been applied elsewhere for detecting change points in PV performance time-series in order to compute nonlinear degradation rates [16], [17]. However, since the rates of change for degradation rates differ from the soiling rates, the FBP was calibrated to provide meaningful results for PV soiling behavior. In order to achieve this, the flexibility of the extracted trend ( $change\_point\_prior\_scale = 0.4$ ), potential change points ( $n\_change\_points = 50$ ) and range ( $change\_point\_range = 1$ ) were adjusted, similar to the process reported by Theristis *et al.* [16]. Since not all investigated sites exhibit the same (or any) seasonal behavior, the FBP *seasonality* settings (i.e., daily, weekly, yearly, custom) were set as *False* to avoid recalibration at each site.

**Bayesian Estimation of Abrupt Change, Seasonality and Trend (BEAST).** The BEAST is an analytical option for robust change point detection and nonlinear trend analysis developed as a MATLAB library and an R package called “RBeast”. It is a time-series decomposition algorithm that utilizes the ensemble modelling technique for pooling the results of multiple fitted models and combining them into an improved final model via a Bayesian model averaging [34]. The BEAST algorithm decomposes the time-series into three components (abrupt changes, periodic/seasonal changes and trends), enabling the extraction of trend and detection of change points. Some of the advantages of this algorithm include the ability to: a) quantify the likelihood (probability) of the detected change points and b) reduce overfitting due to the combined use of multiple models. However, it requires high computational needs/power when using large datasets with multiple time-series. In this work, the BEAST algorithm tuning was performed using the provided dates for Site 1 and the CP method of the RBeast algorithm was provided with the input values shown in Table I. For the investigated soiling stations’ time-series, the period of the cyclic component was determined by specifying an integer number that indicates the number of observations per cycle (e.g., for complete and daily sampled annual time-series, the period of the cyclic/seasonal component is set to 365).

TABLE I  
INPUT PARAMETERS USED FOR RBEAST MODEL

Input Parameter	Value
Period of the cyclic/seasonal component	Number of observations per cycle
Minimum and maximum polynomial order to fit the trend	0 and 4
Minimum and maximum harmonic order allowed in fitting seasonal component	0 and 4
Minimum separation distance between neighboring trend change points	7

A different soiling profile was generated for each change point approach and site by modelling individually each soiling deposition period in the time-series. The fitting was performed by providing as input the change point dates and by modelling a number of linear segments equal to the number of change points +1. Consecutive segments were modelled to join at the change points.

### E. Evaluation Criteria

The extracted soiling profiles were compared using the coefficient of determination ( $R^2$ ) and root mean square error (RMSE) as criteria. Extractions were performed through both the cleanings-only (CO) and the cleanings-and-change-points (CCP) approaches.

It should be noted that the addition of change points increases the number of intervals in which each time-series is divided and the degrees of freedom. Unavoidably, this leads to a reduction in the modelling error despite whether the change points are true or the result of overfitting. For this reason, the CO method was mainly employed as a baseline for the comparison of the different CCP profiles.

## III. RESULTS AND DISCUSSION

### A. Change Points Identification

The impact of considering change points for the extraction of soiling profiles is shown in Fig. 4. The models demonstrated an RMSE reduction over 40% in the best-case scenario, depending on the applied algorithm. Although none of the approaches systematically outperformed the others, PELT demonstrated the least significant improvements, especially for Sites 1 and 9. These results might be improved in future with a finer selection of the penalty value. It is worth mentioning that despite its simplicity and the assumption of a single change point, the results of segmentation are similar to those obtained by the change point algorithms.

Fig. 5 shows the soiling profiles generated with the change point methods for four representative sites. These were selected as examples of maximum (Site 1 and 9), minimum (Site 7) and average (Site 6) CCP extraction improvements.

Despite its strong sawtooth profile, Site 1 benefits considerably from considering change points. As can be seen, after some cleaning events, the start of the soiling deposition is delayed by a few days. In particular, this occurs in Fig. 5 for the

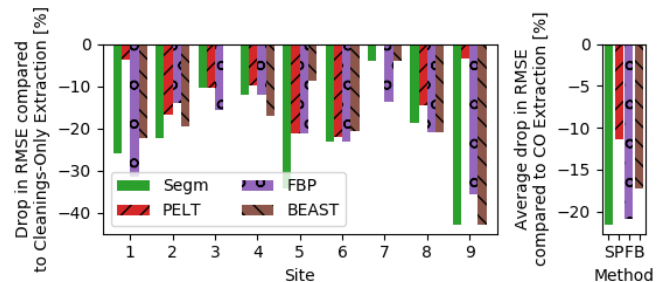


Fig. 4. Relative reduction in RMSE obtained with the cleaning-and-change-points (CCP) approaches compared to the cleanings-only (CO) method. Left plot: RMSE drop obtained by each CCP method for each site. Right plot: average RMSE drop of each CCP method (S: Segmentation, P: PELT, F: FBP, B: BEAST). If no bar is shown, the method does not provide a significant improvement for the specific site.

detected change points 1A<sub>S</sub>, 1C<sub>S</sub>, 1D<sub>S</sub>, 1A<sub>F</sub>, 1D<sub>F</sub> and 1C<sub>B</sub> (first number indicates the site). The identification of this phenomenon through segmentation makes it possible to lower the RMSE by more than 20%. Both FBP and BEAST detect at

least one change point in June 2015 (1B<sub>F</sub>, 1C<sub>F</sub>, 1B<sub>B</sub>), when a positive shift occurs in the soiling ratio profile. This might be due to a cleaning event that remained undetected and that could have been identified if the SRR cleaning detection equation was

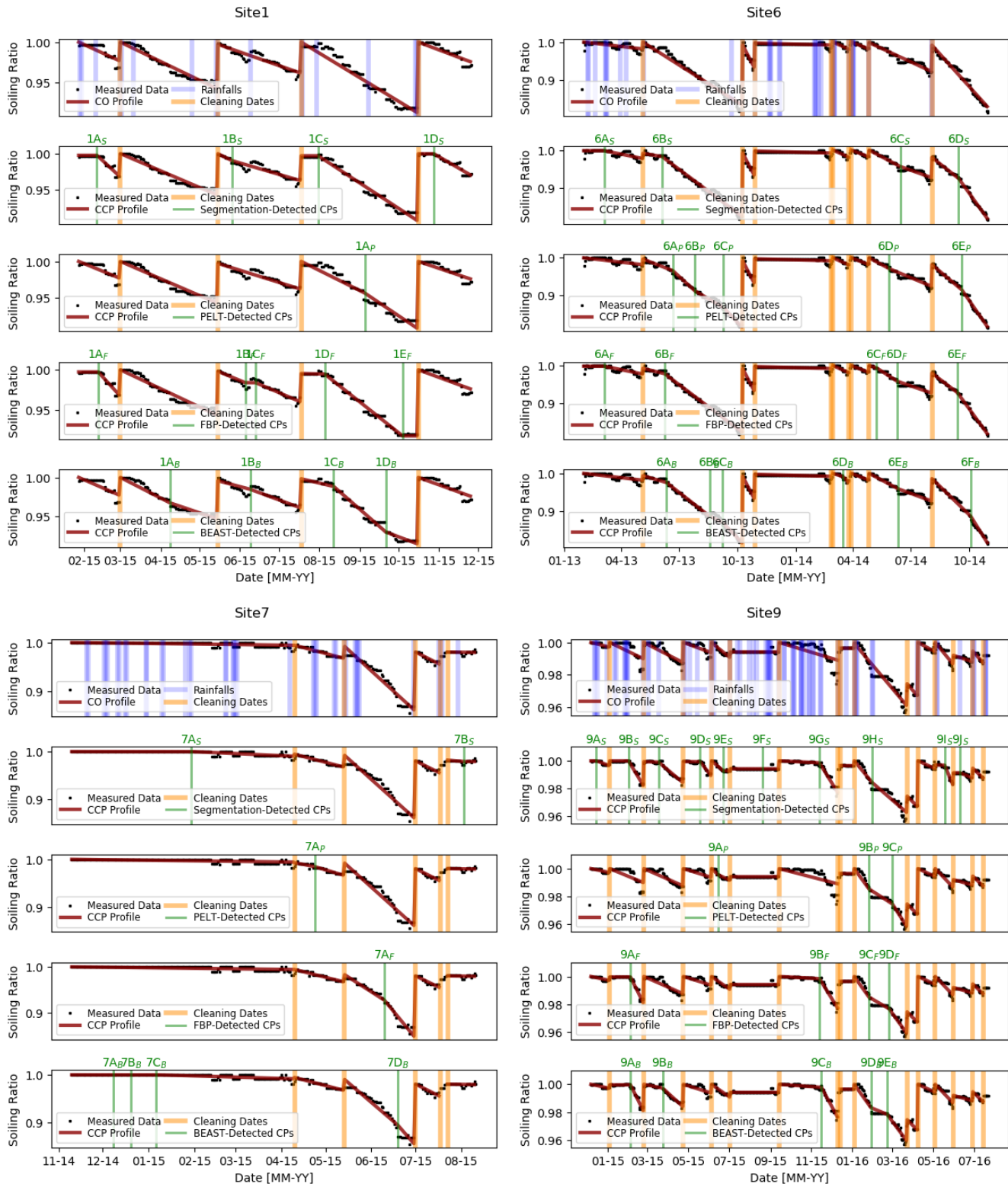


Fig. 5. Cleanings-Only (CO) and Cleanings-and-Change-Points (CCP) profiles for Site 1 (upper left plot), Site 6 (upper right plot), Site 7 (lower left plot) and Site 9 (lower right plot). Each plot shows five graphs, from top to bottom: CO profile, Segmentation-based CCP, PELT-based CCP, FBP-based CCP and BEAST-based CCP. Change points are labelled with a number (corresponding to the site number), a letter from A to F assigned in chronological order, and a subscript referring to the change point detection algorithm (S: Segmentation, P: PELT, F: FBP, B: BEAST).

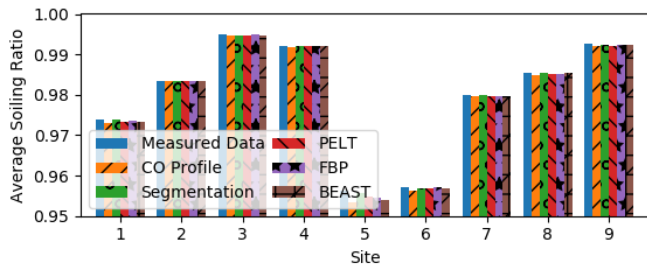


Fig. 6. Average soiling ratio measured by the soiling stations (“Measured Data”) and obtained with the cleaning-only method (“CO Profile”) and with the cleaning-and-change-points approaches.

tuned. Nevertheless, this also confirms that the possibility of using change point detection algorithms for cleaning identification should be further investigated. Lastly, it is worth mentioning that, compared to segmentation, change point detection algorithms were capable of identifying multiple change points within the same deposition period; this occurs between August and October 2015 (1D<sub>F</sub>, 1E<sub>F</sub>, 1C<sub>B</sub>, 1D<sub>B</sub>).

Similar results for all methodologies were obtained for Site 6, with drops in RMSE between 20% and 25%. This is mainly achieved by segmenting the dry period between May and October 2013, which starts with a small deposition rate, followed by a more intense soiling period (6B<sub>S</sub>, 6A<sub>P</sub>, 6B<sub>P</sub>, 6C<sub>P</sub>, 6B<sub>F</sub>, 6A<sub>B</sub>, 6B<sub>B</sub>, 6C<sub>B</sub>). As such, the splitting of shorter deposition periods made by all the approaches in June 2014 (6C<sub>S</sub>, 6D<sub>P</sub>, 6D<sub>P</sub>, 6E<sub>P</sub>) and October 2014 (6D<sub>S</sub>, 6E<sub>P</sub>, 6E<sub>P</sub>, 6F<sub>P</sub>) enabled improvements in the soiling extraction.

Site 7 exhibited the lowest RMSE improvements by the change point detection algorithms. In this case, FBP yielded a difference of 14% compared to the worst performing CP algorithm, PELT. With respect to BEAST, the 9-day delay (7A<sub>B</sub> compared to 7A<sub>F</sub>) reduced the RMSE drop to 4%.

The largest improvements in RMSE were exhibited in Site 9. This is mainly due to the fact that each cleaning is followed by a tail of rain events that prevents soiling from depositing on the modules immediately after the cleaning. This is particularly clear in the period between September and December 2015. PELT was the only method that failed to identify the change in soiling rate during that period. As such, it returned the lowest drop in RMSE of around 3%.

Of the investigated methods, Segm and FBP showed the highest modelling improvements for the considered sites, with average RMSE drops > 20% (right plot of Fig. 4). Segmentation is the only method that does not need any tuning, making it easy to apply, but requires in advance the exact number of change points per fit. The other methods automatically determine the number of change points, but require calibration. Different calibration parameters than those detailed in II.D can result in different estimates. Based on these results, a simple Python code that can be used to apply segmentation to soiling deposition periods has been made available on GitHub [35].

As mentioned, soiling extraction methods can find application to both soiling station and PV performance data. However, it should be noted that such data can be exposed to different levels of noise and issues that can affect the quality of

the analysis. This means that change point algorithms will likely need to be recalibrated when used for different PV and/or soiling datasets. Therefore, no universally-valid calibration settings might exist.

In particular, tuning of change point algorithms is expected to become more important as the data get more noisy and can change depending on the goal of the analysis. For example, in some of the cases analyzed in this study, it is possible to observe abrupt downward shifts in the soiling station profiles (Fig. 5). Such sudden changes are caused by less frequent cleaning of the “reference” cell in a soiling station. Soiling builds up almost equally on the two devices of the stations when no cleaning occurs and, therefore, the effect of soiling deposition can only be measured when soiling is removed from the reference device. These abrupt changes have been occasionally detected as change points by the algorithms (e.g. 1A<sub>P</sub>, 1D<sub>B</sub>, 7D<sub>B</sub>, 9I<sub>S</sub>): finer tunings might be able to solve this problem. On the other hand, it might be possible, in future, to tune the algorithms in order to identify the soiling station’s cleaning days (abrupt downward shifts) and correct the soiling measurement [36]. Similarly, the same algorithms could also be used to recognize those dates in which positive shifts in performance occur (due to natural cleaning events or O&M activities).

### B. Cleaning Schedule Optimization

As discussed in Ref. [15], considering change points had a limited impact on the calculation of the average soiling ratios, obtained as mean of the daily values over a time period (Fig. 6). However, the most significant effect of change points was found in terms of the RMSE between daily modelled profiles and measured profiles. This becomes particularly important when the extracted soiling profiles are used for cleaning schedule optimization and prediction studies.

In a previous publication [8], weather generators were used to stochastically estimate potential rainfall patterns of a site [37]. In the same work, the importance of understanding the monthly variability of soiling rates was also highlighted. Lastly,

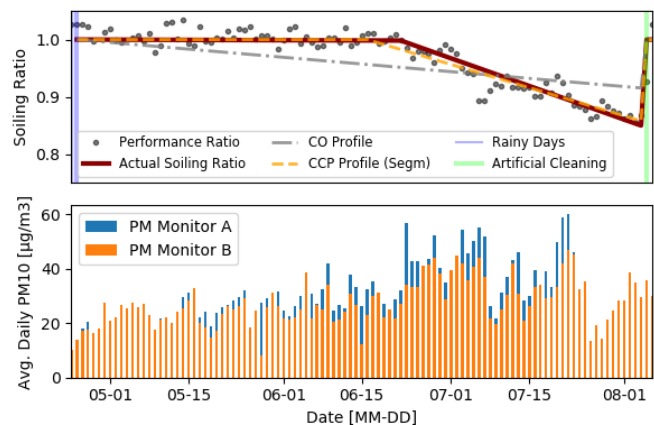


Fig. 7. Top Plot: Soiling Ratio profiles for a 1MW PV plant in Southern Spain, between April 24<sup>th</sup> and August 6<sup>th</sup>, 2019. The brown line shows the soiling profile generated when the actual soiling rate change date (June 22<sup>nd</sup>, 2019) is considered. The grey dash dotted line shows the soiling profile extracted by using only cleaning events. The orange dashed line shows the soiling profile extracted through segmentation. Bottom Plot: Average daily PM<sub>10</sub> concentrations measured by two PM<sub>10</sub> monitors in Granada [38]. No data are available for Monitor A from July 25<sup>th</sup> to 6<sup>th</sup> of August.

that investigation showed that typical soiling profiles could be predicted by combining the generated rainfall patterns and the monthly soiling rates statistically determined from historical data. These profiles could be used to estimate in advance the optimal cleaning frequency of a site and to determine the times in which cleanings are typically more economically viable. Change points can play a key role in this effort, as they can differentiate between high and low soiling rate intervals within the same deposition periods, which can last for several months, rather than assuming constant soiling rates. An example of their use is provided by analyzing the performance of the 1 MW plant described in Section II.A, whose profile was shown in Fig. 2.

The daily soiling ratio profile of the Spanish site recorded in the dry spell between April and August is shown in the upper plot of Fig. 7. This summer dry spell takes place from the rainfall event on April 25<sup>th</sup> to the cleaning event on August 5<sup>th</sup>. A change in soiling deposition occurred on June 22<sup>nd</sup>, in the middle of the period. This was due a dust and sand-laden wind, as also confirmed by the recordings of two particulate monitors installed in the town of Granada (lower plot of Fig. 7), less than 20 km from the PV site [38]. The average of the daily mean coarse particle matter concentrations (PM<sub>10</sub>) ranged from  $25 \pm 6 \mu\text{g}/\text{m}^3$  to  $41 \pm 10 \mu\text{g}/\text{m}^3$  for Monitor A and  $24 \pm 5 \mu\text{g}/\text{m}^3$  to  $33 \pm 9 \mu\text{g}/\text{m}^3$  for Monitor B, for the periods between April 26<sup>th</sup> - June 22<sup>nd</sup> and June 22<sup>nd</sup> - August 4<sup>th</sup>.

The “actual” summer soiling profile, built through segmentation by forcing June 22<sup>nd</sup> as a change date, is depicted with a brown line in Fig. 7. Compared to the soiling profile shown in Refs. [20], [21], where the soiling rate was extracted first from the first deposition period, and only subsequently from the second period, the two periods have been fitted simultaneously to reduce the overall modelling error.

If the whole period was modelled through a cleanings-only approach, the soiling ratio would have a profile shown by the dash dotted grey line in Fig. 7, characterized by a single soiling rate for the whole period. If segmentation was used to identify a change point, this would be found on June 16<sup>th</sup> and the soiling profile would have the more-realistic profile shown by the dashed orange line.

Table II summarizes the characteristics of the soiling profiles for that long dry summer spell when change points are considered versus neglected. For simplicity, segmentation was the only method employed for change point detection; this is the easiest method to use, as it does not require any tuning, and the one showing the best results for the considered set of sites (Fig. 4). While the average soiling ratio was not affected significantly by neglecting the change points, a cleaning-only approach was found to underestimate the maximum loss.

The optimal cleaning day is when a cleaning generates the largest energy recovery (i.e. maximum average soiling ratio) in a single cleaning scenario. If one had to visually determine the most convenient cleaning day for this example, a date in the middle of the most severe soiling deposition period (June 22<sup>nd</sup> to August 4<sup>th</sup>) would be selected. In particular, a date in the middle of this period is returned (July 14<sup>th</sup>) by analyzing the “actual” soiling profile with methodology proposed in Ref. [8] and by assuming that the cleaning would have not changed the

TABLE II  
CHARACTERISTICS OF THE EXTRACTED SOILING RATIO PROFILES BETWEEN APRIL 26<sup>TH</sup> AND AUGUST 4<sup>TH</sup>.

Method	Avg. Soiling Ratio	Min Soiling Ratio	Soiling Rates before and after the CP	Change Point Date	Recommended Cleaning Date
“Actual” Profile	0.97	0.86	-0.01% -0.34%	June 22 <sup>nd</sup>	July 14 <sup>th</sup>
CO Profile	0.96	0.92	-0.08% -0.08%	N.A.	June 16 <sup>th</sup>
CCP profile	0.97	0.87	-0.00% -0.29%	June 16 <sup>th</sup>	July 11 <sup>th</sup>

soiling rate. On the other hand, the results in Table II show that neglecting the change in soiling deposition rate would lead instead to selecting June 16<sup>th</sup> as the optimal cleaning day. This date in reality is part of a period of no or limited soiling, as it is more than a month earlier than the sand laden wind. For this reason, cleaning on June 16<sup>th</sup> would have none or limited effect on the soiling ratio. According to an economic analysis conducted on the investigated PV site [20], cleaning the system 28 days earlier than the optimal date would make soiling mitigation almost no profitable (< 0.1% increase in Net Present Value compared to a no-cleaning scenario). In contrast, the CCP profile generated through segmentation returned a date within 3 days from the optimal cleaning day. This would maximize the returns of soiling mitigation (> 3.5% increase in Net Present Value), with only a limited difference compared to the benefits of an optimal cleaning schedule [20].

The CCP ability to analyze historical soiling trends more accurately than a CO approach makes change points an important tool toward the development of methods to characterize the seasonality of soiling, to generate future soiling profiles and to also identify in advance the cleaning frequencies and dates that are typically more cost effective.

This simple example aims to further demonstrate that change points can be a useful tool for automated PV soiling extraction. In the given example, change points made it possible to identify a variation in soiling rate due to a change in suspended particle concentration. As mentioned earlier, they might be of use in a number of other conditions as well, such as in the case of a switch from a prolonged rainy period to a dry spell, or in the case of seasonal harvesting or tilling activities taking place near a PV site.

#### IV. CONCLUSIONS

This paper investigated the impact of identifying change points in PV soiling extraction. Segmentation and three change point algorithms were employed to detect the locations of change points using field data from nine soiling stations.

Overall, considering change points was found to improve the soiling extraction compared to a cleaning-only approach. Not all sites benefitted from the change point detection algorithms and not all algorithms behaved similarly. In particular, among the investigated approaches, PELT exhibited the lowest RMSE improvements, although it should be further investigated since refined penalty values could potentially enhance the results. The relatively simplest approach of segmentation returned

comparable results with the best performing change point algorithm (i.e. FBP).

Furthermore, segmented regression was also applied on a 1 MW PV plant for benchmarking purposes. It was shown that a change-point-unaware soiling extraction can lead to identifying non-optimal cleaning dates. This can negatively impact the analysis of historical soiling data and the development of cleaning optimization and forecasting methods.

Future research directions are suggested throughout the paper. Improving the tuning of the change point algorithms can lead to different results. Although the change point detection algorithms were tested for variations in soiling rate deposition (i.e. continuous change points), they could also be used to identify cleaning events (i.e. discontinuous change points). Finally, change point algorithms might also be tuned to recognize infrequent cleaning conditions and provide useful information to correct soiling station measurements.

#### ACKNOWLEDGMENT

The authors acknowledge Sonnedix for sharing commercial PV performance data. In particular, they wish to thank Juan M. Fernández and Ruth Prieto for the support in accessing and analyzing the data.

NoSoilPV project is a Marie Skłodowska-Curie Actions Individual Fellowship grant (agreement no. 793120) funded by the European Union’s Horizon 2020 research and innovation programme.

Sandia National Laboratories is a multitechnology laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. 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 U.S. Government purposes.

Project ROM-PV is supported under the umbrella of SOLAR-ERA.NET Cofund by the General Secretariat for Research and Technology (GSRT), the Ministry of Economy, Industry and Competitiveness – State Research Agency (MINECO-AEI) and the Research and Innovation Foundation (RIF) of Cyprus. SOLAR-ERA.NET is supported by the European Commission within the EU Framework Programme for Research and Innovation HORIZON 2020 (Cofund ERA-NET Action, N° 691664).

#### REFERENCES

[1] K. Ilse *et al.*, “Techno-Economic Assessment of Soiling Losses and Mitigation Strategies for Solar Power Generation,” *Joule*, vol. 3, no. 10, pp. 2303–2321, Oct.

- 2019.
- [2] S. C. S. Costa, A. S. A. Diniz, and L. L. Kazmerski, “Solar energy dust and soiling R & D progress : Literature review update for 2016,” *Renew. Sustain. Energy Rev.*, vol. 82, pp. 2504–2536, 2018.
- [3] G. P. Smestad *et al.*, “Modelling photovoltaic soiling losses through optical characterization,” *Sci. Rep.*, vol. 10, no. 1, p. 58, Dec. 2020.
- [4] L. Micheli and M. Muller, “An investigation of the key parameters for predicting PV soiling losses,” *Prog. Photovoltaics Res. Appl.*, vol. 25, no. 4, pp. 291–307, Apr. 2017.
- [5] L. Micheli, M. G. Deceglie, and M. Muller, “Predicting photovoltaic soiling losses using environmental parameters: An update,” *Prog. Photovoltaics Res. Appl.*, vol. 27, no. 3, pp. 210–219, Mar. 2019.
- [6] B. Figgis, B. Guo, W. Javed, S. Ahzi, and Y. Rémond, “Dominant environmental parameters for dust deposition and resuspension in desert climates,” *Aerosol Sci. Technol.*, vol. 52, no. 7, pp. 788–798, Jul. 2018.
- [7] K. K. Ilse, B. W. Figgis, V. Naumann, C. Hagendorf, and J. Bagdahn, “Fundamentals of soiling processes on photovoltaic modules,” *Renew. Sustain. Energy Rev.*, vol. 98, no. January, pp. 239–254, 2018.
- [8] L. Micheli, E. F. Fernández, M. Muller, and F. Almonacid, “Extracting and Generating PV Soiling Profiles for Analysis, Forecasting, and Cleaning Optimization,” *IEEE J. Photovoltaics*, vol. 10, no. 1, pp. 197–205, Jan. 2020.
- [9] F. A. Mejia and J. Kleissl, “Soiling losses for solar photovoltaic systems in California,” *Sol. Energy*, vol. 95, pp. 357–363, 2013.
- [10] A. Kimber, L. Mitchell, S. Nogradi, and H. Wenger, “The Effect of Soiling on Large Grid-Connected Photovoltaic Systems in California and the Southwest Region of the United States,” in *Photovoltaic Energy Conversion, Conference Record of the 2006 IEEE 4th World Conference on*, 2006, pp. 2391–2395.
- [11] P. Besson, C. Munoz, G. Ramirez-Sagner, M. Salgado, R. Escobar, and W. Platzer, “Long-Term Soiling Analysis for Three Photovoltaic Technologies in Santiago Region,” *IEEE J. Photovoltaics*, vol. 7, no. 6, pp. 1755–1760, 2017.
- [12] M. G. Deceglie, L. Micheli, and M. Muller, “Quantifying Soiling Loss Directly From PV Yield,” *IEEE J. Photovoltaics*, vol. 8, no. 2, pp. 547–551, Mar. 2018.
- [13] Å. Skomedal, H. Haug, and E. S. Marstein, “Endogenous Soiling Rate Determination and Detection of Cleaning Events in Utility-Scale PV Plants,” *IEEE J. Photovoltaics*, vol. 9, no. 3, pp. 858–863, 2019.
- [14] R. K. Jones *et al.*, “Optimized Cleaning Cost and Schedule Based on Observed Soiling Conditions for Photovoltaic Plants in Central Saudi Arabia,” *IEEE J. Photovoltaics*, vol. 6, no. 3, pp. 730–738, 2016.
- [15] L. Micheli, M. Muller, E. F. Fernández, and F. M. Almonacid, “Change Point Detection: An Opportunity to Improve PV Soiling Extraction,” in *IEEE 45th Photovoltaic Specialist Conference (PVSC)*, 2020.
- [16] M. Theristis, A. Livera, C. B. Jones, G. Makrides, G. E. Georghiou, and J. S. Stein, “Nonlinear Photovoltaic Degradation Rates: Modeling and Comparison Against



- Conventional Methods,” *IEEE J. Photovoltaics*, vol. 10, no. 4, pp. 1112–1118, Jul. 2020.
- [17] M. Theristis *et al.*, “Modeling nonlinear photovoltaic degradation rates,” in *IEEE 47th Photovoltaic Specialist Conference (PVSC)*, 2020.
- [18] L. Micheli, D. Ruth, M. G. Deceglie, and M. Muller, “Time Series Analysis of Photovoltaic Soiling Station Data: Version 1.0, August 2017,” Golden, CO, 2017.
- [19] I. E. C. (IEC), “IEC 61724-1 - Photovoltaic System Performance Monitoring—Guidelines for Measurement, Data Exchange, and Analysis (Part 1),” 2017.
- [20] L. Micheli, E. F. Fernández, J. T. Aguilera, and F. Almonacid, “Economics of seasonal photovoltaic soiling and cleaning optimization scenarios,” *Energy*, vol. 215, p. 119018, Jan. 2021.
- [21] L. Micheli, M. Theristis, D. Talavera, F. Almonacid, J. S. Stein, and E. F. Fernandez, “Photovoltaic Cleaning Frequency Optimization Under Different Degradation Rate Patterns,” *Renew. Energy (In Press)*.
- [22] Global Modeling and Assimilation Office (GMAO), “MERRA-2 tavg1\_2d\_slv\_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Single-Level Diagnostics V5.12.4,” 2015. .
- [23] J. D. Toms and M. A. Villard, “Détection des seuils: Choix de la méthode statistique en fonction des questions écologiques et des objectifs de planification de la conservation,” *Avian Conserv. Ecol.*, vol. 10, no. 1, 2015.
- [24] National Renewable Energy Laboratory, “PV\_soiling: code for extracting soiling loss from PV plant data,” 2018. [Online]. Available: [https://github.com/NREL/pv\\_soiling](https://github.com/NREL/pv_soiling). [Accessed: 29-Apr-2018].
- [25] E. Jones, E. Oliphant, P. Peterson, and et al., “SciPy: Open Source Scientific Tools for Python.” 2001.
- [26] C. R. Harris *et al.*, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
- [27] R. Killick, P. Fearnhead, and I. A. Eckley, “Optimal detection of changepoints with a linear computational cost,” *J. Am. Stat. Assoc.*, vol. 107, no. 500, pp. 1590–1598, 2012.
- [28] B. Jackson *et al.*, “An algorithm for optimal partitioning of data on an interval,” *IEEE Signal Process. Lett.*, vol. 12, no. 2, pp. 105–108, 2005.
- [29] C. Truong, L. Oudre, and N. Vayatis, “Selective review of offline change point detection methods,” *Signal Processing*, vol. 167, p. 107299, 2020.
- [30] R. Killick and I. A. Eckley, “changepoint: An R Package for Changepoint Analysis,” *J. Stat. Softw.*, vol. 58, no. 3, 2014.
- [31] G. Schwarz, “Estimating the Dimension of a Model,” *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
- [32] K. Haynes, “Detecting Abrupt Changes in Big Data,” Lancaster University, 2017.
- [33] S. J. Taylor and B. Letham, “Forecasting at Scale,” *PeerJ Prepr.*, 2017.
- [34] K. Zhao *et al.*, “Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm,” *Remote Sens. Environ.*, vol. 232, no. July, p. 111181, 2019.
- [35] L. Micheli, “segm\_soiling\_profile: a function of the NoSoilPV Python package,” 2020. [Online]. Available: <https://github.com/lmicheli/NoSoilPV>. [Accessed: 21-Aug-2020].
- [36] M. Muller, L. Micheli, and A. A. Martinez-Morales, “A Method to Extract Soiling Loss Data From Soiling Stations with Imperfect Cleaning Schedules,” in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*, 2017.
- [37] D. S. Wilks and R. L. Wilby, “The weather generation game: a review of stochastic weather models,” *Prog. Phys. Geogr.*, vol. 23, no. 3, pp. 329–357, 1999.
- [38] Ayuntamiento de Granada, “Calidad del aire,” 2020. [Online]. Available: <https://www.granada.org/inet/calidadaire.nsf>. [Accessed: 07-May-2020].