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.

SAND2020-5826C

Modeling nonlinear photovoltaic degradation rates

Marios Theristis
Sandia National Laboratories
Albuquerque, NM, USA
mtheris@sandia.gov

Andreas Livera
University of Cyprus
Nicosia, Cyprus
livera.andreas@ucy.ac.cy

Leonardo Micheli
University of Jaén
Jaén, Spain
lmicheli@ujaen.es

C. Birk Jones
Sandia National Laboratories
Albuquerque, NM, USA
cbjones@sandia.gov

George Makrides
University of Cyprus
Nicosia, Cyprus
makrides.georgios@ucy.ac.cy

George E. Georghiou
University of Cyprus
Nicosia, Cyprus
geg@ucy.ac.cy

Joshua S. Stein
Sandia National Laboratories
Albuquerque, NM, USA
jsstein@sandia.gov

***Abstract*—It is a common approach to assume a constant performance drop during the photovoltaic (PV) lifetime. However, operational data demonstrated that PV degradation rate (R_D) may exhibit nonlinear behavior. Neglecting nonlinearities may increase financial risks. This study presents and compares three approaches, based on open-source libraries, which are able to detect and calculate nonlinear R_D. Two of these approaches include trend extraction and change-point detection methods, which are frequently used statistical tools. Initially, the processed monthly PV performance ratio (*PR*) time-series are decomposed in order to extract the trend and change-point analysis techniques are applied to detect changes in the slopes. Once the number of change-points is optimized by each model, the ordinary least squares (OLS) method is applied on the different segments to compute the corresponding rates. The third methodology is a regression analysis method based on simultaneous segmentation and slope extraction. Since the "real" R_D value is an unknown parameter, this investigation was based on synthetic datasets with emulated two-step degradation rates. As such, the performance of the three approaches was compared exhibiting mean absolute errors ranging from 0 to 0.46%/year whereas the change-point position detection differed from 0 to 10 months.**

Keywords—change-point analysis, modeling, nonlinear degradation, photovoltaics (PV).

I. Introduction

Precise knowledge of photovoltaic (PV) degradation rate (R_D) is important for projecting lifetime energy yield. Simplistic assumptions may cause detrimental effects increasing PV financial uncertainties and hence, investment risk [2]. Such assumptions may include: a) the usage of single R_D values from literature, b) values reported from different climatic conditions, c) values reported for a specific technology yet different module quality, d) assumption of constant performance loss over time, and so on.

On the other hand, relatively simple statistical analysis can be performed on PV performance time-series, in order to extract the degradation rate of a particular system. However, it is known that PV performance fluctuates due to a number of seasonally related factors such as temperature [3], spectrum [4], soiling [5], etc. Therefore, although the statistical tools are available and relatively easy to use, it is inherently challenging to extract reproducible PV degradation rates [6].

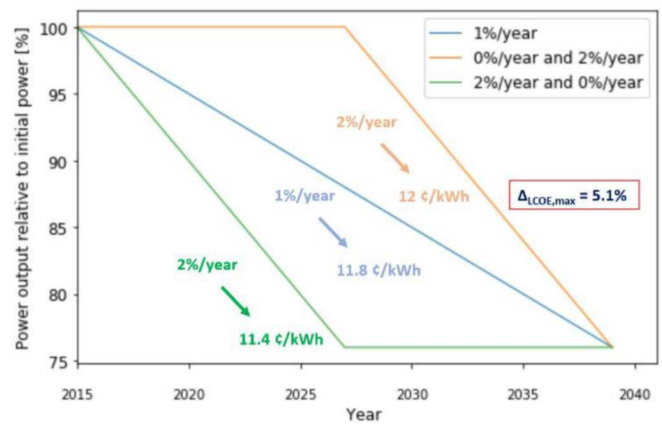

Fig. 1. Theoretical comparison of different degradation rates and the impact on *LCOE* (assuming discount rate 4%, operation &maintenance of 2%, installation costs of 3 $/W, annual irradiation of 1700 kWh/kWp). This figure was recreated from Stein *et al.* [1].

The way PV time-series data are handled and processed, adds to the complexity and uncertainty [7]. For example, if temperature correction is applied and the reported temperature coefficients are biased, a seasonality due to changing temperature ranges will be introduced. Furthermore, spatial and temporal variability in the actual temperature of the array could lead to seasonality in the performance metric, especially if the wind direction varies seasonally and affects spatial patterns in the array temperature. Moreover, in respect to data quality, integrity and processing, the R_D calculation can also be influenced by other factors such as missing data [8], sensor drift [9], filtering criteria [10], aggregation [11], etc.

The most commonly used assumption to statistically extract the R_D lies on the hypothesis of a linear performance drop over time. Operational data demonstrated that this is unrealistic in some cases [12] mainly due to the initial and wear-out degradation that may occur [13]. For example, Light and elevated Temperature Induced Degradation (LeTID) [14] and Light Induced Degradation (LID) [15] may occur initially in the PV module lifetime and then cease once an equilibrium has been reached, directly affecting the degradation rate.

The path traveled to a certain performance loss has a significant economic impact on the levelized cost of energy (*LCOE*), as theoretically shown in Fig 1. Another theoretical

Sandia National Laboratories is a multimission 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.

study conducted by Jordan *et al*. [13, 16] demonstrated *LCOE* differences of ~1.1 ¢/kWh, making the R_D the third most important factor influencing the *LCOE* [13]. Recently, a new methodology that takes into account nonlinear PV behavior was introduced by detecting change-points from the trends of decomposed PV time-series [12]. The methodology was also applied on real PV performance data verifying that some systems may exhibit nonlinear performance loss.

Expanding on this line of work, this study examines three methodologies based on open-source Python and R packages, which can be applied to PV time-series and are able to detect and calculate nonlinear R_D. Initially, the processed monthly PV performance ratio (*PR*) time-series are decomposed in order to extract the trend and change-point analysis techniques are applied to detect changes in the slopes. Once the number of change-points is optimized by each model, the ordinary least squares (OLS) method is applied to the different segments to compute the corresponding rates. Since the "real" R_D value is an unknown parameter, this investigation was based on synthetic datasets with emulated two-step degradation rates.

II. Methodology

In statistics, a change-point (or switch-point, or break-point) refers to a change in time-series properties (e.g., mean, variance, correlation, etc.) [17]. Such changes can be either continuous or discontinuous and in the case of nonlinear degradation, the change is considered as continuous since the two segments have the same R_D value at the change-point [18].

Trend extraction and change-point detection methods are frequently used statistical tools and several open-source algorithms are available in Python and R. Three of them were selected in this study based on the hypothesis that they can be applied on PV performance data. This hypothesis is mainly due to their ability to extract the trend prior to applying any change-point analysis instead of applying the change-point model on raw and highly fluctuating data. Highly variable/fluctuating time-series may result in too many change-points due to overfitting rate changes.

A. *Generation of synthetic datasets*

Five different scenarios of synthetic datasets were generated (see Table I) in order to examine the selected open-source packages. Their performance and effectiveness were evaluated based on the ability to detect the number and position(s) of change-points and also the precision errors in estimating the degradation rate values. Besides knowing the "real" degradation rate values and change-point positions, synthetic data also have the advantage of being independent of sensor drift, temperature uncertainty, soiling, maintenance issues etc., that may affect the accuracy and/or uncertainty of the calculations.

TABLE I. Five Different Scenarios Of Synthetic PV Performance Datasets Considered In The Comparative Analysis

Scenario	Change-point date/position	$R_{D,1}$ (%/year)	$R_{D,2}$ (%/year)
a	Jan-17 (24)	-5	-1
b	Jan-19 (48)	-0.5	-3.5
c	Jan-21 (72)	-1	-0.5
d	Jan-23 (96)	-1	-2.5
e	Jan-25 (120)	-3	-1

Similar to the procedure described by Theristis *et al*. [12], in order to generate synthetic PV performance datasets with annually varying meteorological conditions, 15 typical meteorological year (TMY) datasets from different locations in New Mexico (NM), USA were collected. These were used as consecutive inputs to a PV performance model of a monocrystalline silicon module using the Sandia PV Array Performance Model (SAPM) [19] from *pvlib-python* [20]. The PV performance time-series were then sliced into different segments where the different degradation rates were applied, and corresponding change-point locations were positioned to represent a two-step nonlinear performance loss. Temperature correction, normalization and monthly aggregation was then performed in order to create a dataset of temperature-corrected performance ratio (PR_{TC}) time-series.

B. *Description of selected open-source libraries*

Facebook Prophet Algorithm (*FBP*) [21]: 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; holidays are neglected in this study. A piecewise linear model is applied by default for the trend whereas the seasonal model is similar to the exponential smoothing in the Holt-Winters technique.

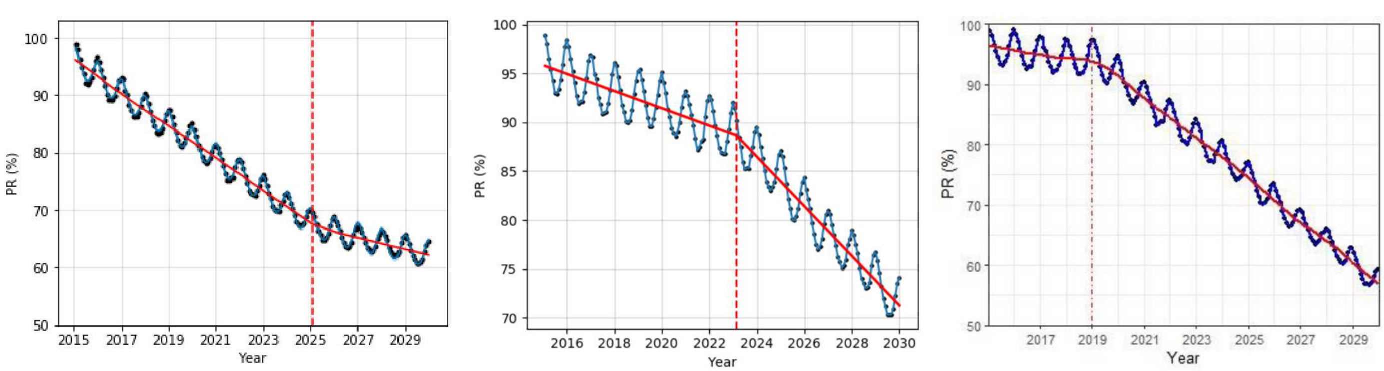

Fig. 2. Change-point detection on monthly *PR* time-series with different nonlinear degradation rate combinations and change-point positions. The *PR* time-series were extracted from synthetic datasets, which were simulated using the SAPM on *pvlib-python* using 15 TMY datasets from around NM, USA. From left to right, the plots were generated using results from FBP (Scenario *e*), SegmR (Scenario *d*) and RBeast (Scenario *b*). The red dashed vertical lines indicate the positions (locations) of the detected change-points.

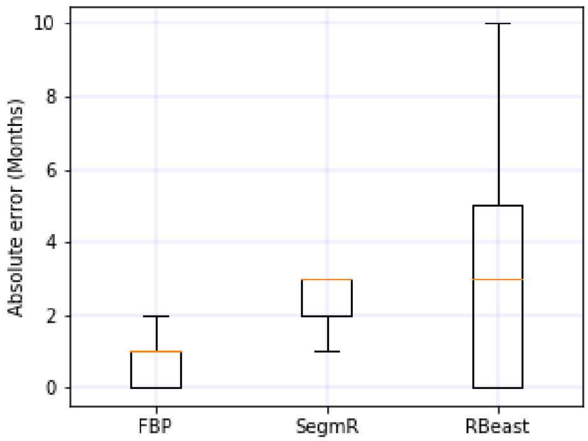

Fig. 3. Boxplot of performance comparison for the three different models in respect to the absolute error in locating the change-point position.

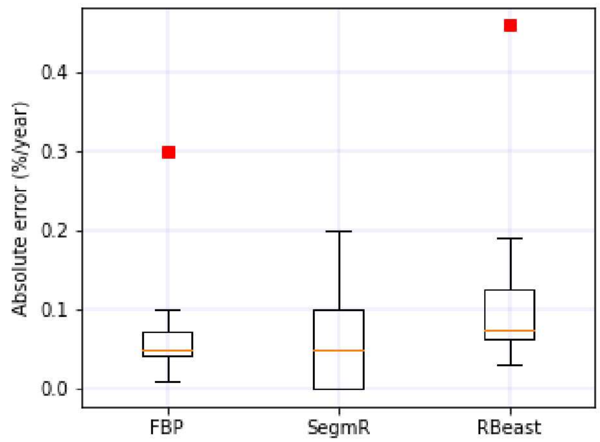

Fig. 4. Boxplot of performance comparison for the three different models in respect to the absolute error in estimating the degradation rates in different segments.

Once the trend is extracted for the *PR* time-series, change-point analysis is performed to identify the number and location of change-points by capturing statistical changes in the slopes of pre-defined segments of the time-series. Once the nonlinear trend is "sliced", the methodology treats each segment in a linear manner. This is achieved by applying the ordinary least squares (OLS) method [22] in order to compute the different degradation rates for each segment, accordingly.

In order to setup FBP to provide meaningful results for PV behavior, the flexibility of the extracted trend, number of potential change-points, and range had to be adjusted according to the process and settings reported by Theristis *et al*. [12]. An example of a FBP application is demonstrated in Fig. 2 (left) for Scenario *e* of Table I.

Segmented or Piecewise Regression (*SegmR*) [18]: In segmented regression, the data are fit with more than just one line, separated by the change-point(s). In this work, the degradation curve is divided into two continuous segments and each segment is fitted in a way that the sum of the squared error of the complete time-series is minimized (see Fig. 2, middle for Scenario *d* of Table I). The two segments are forced to join at the change-point.

Compared to other change-point algorithms, this approach allows to simultaneously identify both change-points and corresponding slopes (i.e., degradation rates). One of the drawbacks in the formulation used in this study is that SegmR is able to identify only the pre-determined number of change-points (i.e. one change-point in this case). Also, because of the noise in the data, the algorithm would return a change-point even in conditions of a linear degradation. Future work will work on introducing filters able to address this issue.

The model is developed in Python 3.7.0. Seasonal adjustment is first performed on the data by using the *seasonal_decompose* function in *statsmodels* to extract the trend component of the time-series. The segment regression equations are defined through the piecewise function in the *NumPy* library. The curve fitting is performed with the *curve_fit* function in the *SciPy* library [23], which employs a Trust Region Reflective algorithm. The initial guesses were set to: x_0=N/2, a_0=0, a_1=0 and b=1, where N is the total number of months in the time-series. In addition, these bounds were chosen $0 \leq x_0 \leq N$ and $0 \leq b \leq 1$. It should be noted that the returned change-point date, in some of the cases, can be affected by the initial guesses.

RBeast [24]: Bayesian estimation of abrupt change, seasonality and trend (BEAST) applies the Bayesian ensemble time-series decomposition algorithm. By utilizing the ensemble modelling technique, the results of the multiple fitted models are determined and incorporated into the final averaged Bayesian model. According to the authors [24], this makes the BEAST a universal and robust algorithm for change-point and complex nonlinear trend analysis. More specifically, the "BEAST" enables the determination of abrupt changes (i.e. change-points), cyclic variations (e.g. seasonality) and nonlinear trends in time-series observations by decomposing the time-series data into three components: abrupt changes, trends and cyclic/seasonal variations. The algorithm also quantifies the likelihood (probability) of the detected changes. The BEAST algorithm is applicable to time-series data of all kinds and it was developed as a MATLAB library and an R package called "RBeast".

In this work, the period of the cyclic/seasonal component of the time-series and the minimum separation time between the neighboring season change-points were set to 12. The maximum number of trend changepoints allowed was 1. Furthermore, the minimum and maximum polynomial order to fit the trend were set to 0 and 1 respectively. Finally, the maximum harmonic order for fitting the seasonal component was set to 5. A demostration of the output generated by RBeast is illustrated in Fig. 2 (right) for Scenario *b* of Table I.

III. Results And Discussion

The application of the three different methodologies on the synthetic monthly PR_{TC} time-series revealed different slopes, mutually verifying the presence of a nonlinear, two-step behavior.

With respect to locating the change-point positions, the absolute error varied from 0 to 10 Months with the FBP outperforming the other models by demonstrating a median

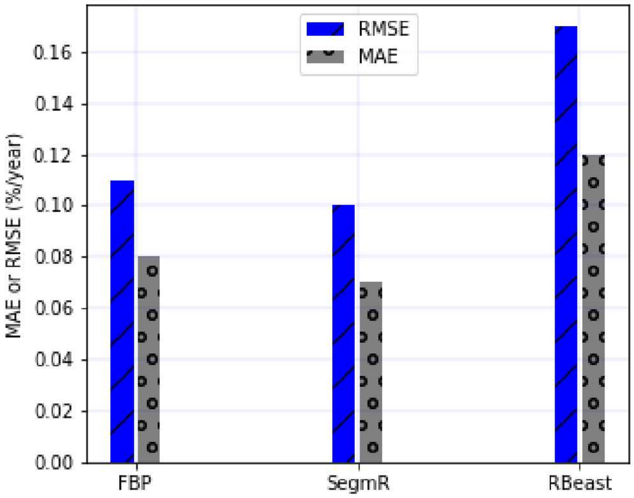

Fig. 5. Comparison of the *RMSE* and *MAE* metrics in estimating nonlinear degradation rates using FBP, SegmR and RBeast.

absolute error of 1 Month compared to 3 Months of SegmR and RBeast (see Fig. 3). Furthermore, the Mean Absolute Error (*MAE*) was 0.8, 2.4, 3.6 Months for FBP, SegmR and RBeast, respectively.

The most critical part of the performance comparison in this study is the prediction error in quantifying the degradation rates at each segment. Fig. 4 illustrates the performance comparison in the form of a boxplot where, initially, the FBP and SegmR methods seem to outperform RBeast. Specifically, a median absolute error of 0.05%/year was found for FBP and SegmR whereas RBeast exhibited 0.075%/year. The modeling performance was further compared using the *MAE* and root mean square (*RMSE*) metrics (see Fig. 5). With respect to *MAE*, 0.08%/year, 0.07%/year, 0.12%/year were demonstrated for FBP, SegmR and RBeast, respectively, which highlights a better performance of SegmR, in this case. Furthermore, the *RMSE* values followed the same trends as in the case of *MAE*, with SegmR exhibiting 0.10%/year whereas FBP and RBeast resulted in 0.11%/year and 0.17%/year, respectively.

Overall, all methods demonstrated relatively low prediction errors, even when the change-point detection error was relatively high. From a preliminary interpretation standpoint, SegmR is clearly the winner, if accuracy on degradation rate prediction is the priority assuming a two-step behavior. On the other hand, change-point analysis models are also useful for detecting abrupt changes in PV performance time-series, which may not necessarily be due to a change in degradation rate. Such models can be applied for detecting failures, maintenance events, or any other trend-based performance losses such as soiling [25], especially when cleaning events are not recorded or soiling profiles are unknown. Therefore, depending on the application, the optimum change-point detection model may differ.

IV. Conclusions

A comparative analysis was performed to investigate different open-source methodologies for detecting and quantifying nonlinear PV degradation rate. Two of the methodologies (i.e., FBP and RBeast) consisted of decomposition models coupled to change-point analysis techniques whereas the third one (SegmR) performed simultaneous segmentation and extraction of slopes in the different segments enabling the calculation of the corresponding degradation rates.

Although all methods demonstrated good performance (highest *MAE* and *RMSE* of 0.12%/year and 0.17%/year), the results indicate two main conclusions. First, FBP exhibited the lowest prediction errors in locating the positions of change-points. Second, SegmR was the most accurate in computing the corresponding degradation rates. This indicates that different change-point detection models may be more appropriate depending on the particular case, although this needs further investigation by applying the models to failure detection, soiling, etc.

Future work will expand on this study by investigating additional change-point techniques and a longer list of scenarios including three-step or greater degradation rate behavior to incorporate the wear-out phase. Furthermore, synthetic data generated for other PV module technologies will also be used to demonstrate further the robustness of these methods.

Acknowledgment

This material is based upon work 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 34364. Sandia National Laboratories is a multimission 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 UCY team would like to acknowledge the financial support received from the University of Cyprus through the internal PV-4LIFE project. Leonardo Micheli's contribution was supported by the European Union's Horizon 2020 research and innovation programme under the NoSoilPV project (Marie Skłodowska-Curie Grant agreement no. 793120).

References

[1] J. S. Stein, C. Robinson, B. King, C. Deline, S. Rummel, and B. Sekulic, "PV Lifetime Project: Measuring PV Module Performance Degradation: 2018 Indoor Flash Testing Results," in *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*, 2018, pp. 0771-0777.

[2] D. C. Jordan and S. R. Kurtz, "Photovoltaic Degradation Rates—an Analytical Review," *Progress in Photovoltaics: Research and Applications,* vol. 21, pp. 12-29, 2013.

[3] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Solar Energy,* vol. 83, pp. 614-624, 2009.

[4] J. A. Caballero, E. F. Fernández, M. Theristis, F. Almonacid, and G. Nofuentes, "Spectral Corrections Based on Air Mass, Aerosol Optical Depth, and Precipitable Water for PV Performance Modeling," *IEEE Journal of Photovoltaics,* vol. 8, pp. 552-558, 2018.

[5] L. Micheli and M. Muller, "An investigation of the key parameters for predicting PV soiling losses," *Progress in Photovoltaics: Research and Applications,* vol. 25, pp. 291-307, 2017.

[6] D. C. Jordan, C. Deline, M. G. Deceglie, A. Nag, G. M. Kimball, A. B. Shinn, J. J. John, A. A. Alnuaimi, A. B. A. Elnosh, W. Luo, A. Jain, M. U. Saleh, H. v. Korff, Y. Hu, J. Jaubert, and F. Mavromatakis, "Reducing Interanalyst Variability in Photovoltaic Degradation Rate Assessments," *IEEE Journal of Photovoltaics,* pp. 1-7, 2019.

[7] D. C. Jordan and S. R. Kurtz, "The Dark Horse of Evaluating Long-Term Field Performance-Data Filtering," *IEEE Journal of Photovoltaics,* vol. 4, pp. 317-323, 2014.

[8] A. Livera, M. Theristis, E. Koumpli, S. Theocharides, G. Makrides, J. Sutterlueti, J. S. Stein, and G. E. Georghiou, "Data processing and quality verification for improved photovoltaic performance and reliability analytics," *Under review,* 2020.

[9] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV Degradation Methodology and Application," *IEEE Journal of Photovoltaics,* vol. 8, pp. 525-531, 2018.

[10] A. J. Curran, C. B. Jones, S. Lindig, J. Stein, D. Moser, and R. H. French, "Performance Loss Rate Consistency and Uncertainty Across Multiple Methods and Filtering Criteria," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, 2019, pp. 1328-1334.

[11] B. Paudyal, M. Bolen, and D. Fregosi, "PV Plant Performance Loss Rate Assessment: Significance of Data Filtering and Aggregation," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, 2019, pp. 0866-0869.

[12] 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 Journal of Photovoltaics,* pp. 1-7, 2020.

[13] D. C. Jordan, T. J. Silverman, B. Sekulic, and S. R. Kurtz, "PV degradation curves: non-linearities and failure modes," *Progress in Photovoltaics: Research and Applications,* vol. 25, pp. 583-591, 2017.

[14] F. Kersten, P. Engelhart, H. Ploigt, A. Stekolnikov, T. Lindner, F. Stenzel, M. Bartzsch, A. Szpeth, K. Petter, J. Heitmann, and J. W. Müller, "A new mc-Si degradation effect called LeTID," in *42nd IEEE Photovoltaic Specialist Conference (PVSC)*, 2015, pp. 1-5.

[15] M. Theristis, V. Venizelou, G. Makrides, and G. E. Georghiou, "Chapter II-1-B - Energy yield in photovoltaic systems," in *McEvoy's handbook of photovoltaics (third edition)*, S. A. Kalogirou, Ed.: Academic Press 2018, pp. 671-713.

[16] D. C. Jordan, S. R. Kurtz, K. VanSant, and J. Newmiller, "Compendium of photovoltaic degradation rates," *Progress in Photovoltaics: Research and Applications,* vol. 24, pp. 978-989, 2016.

[17] S. Sharma, D. A. Swayne, and C. Obimbo, "Trend analysis and change point techniques: a survey," *Energy, Ecology and Environment,* vol. 1, pp. 123-130, 2016.

[18] W. W. Piegorsch and A. J. Bailer, *Analyzing environmental data*: John Wiley & Sons, 2005.

[19] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Photovoltaic Array Performance Model," *Sandia Report, SAND2004-3535,* 2004.

[20] W. F. Holmgren, R. W. Andrews, A. T. Lorenzo, and J. S. Stein, "PVLIB Python 2015," in *42nd IEEE Photovoltaic Specialist Conference (PVSC)*, 2015, pp. 1-5.

[21] S. J. Taylor and B. Letham, "Forecasting at Scale," *The American Statistician,* vol. 72, pp. 37-45, 2018/01/02 2018.

[22] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting methods and applications*: John Wiley & sons, 2008.

[23] E. Jones, T. Oliphant, and P. Peterson, "SciPy: Open source scientific tools for Python," 2001.

[24] K. Zhao, M. A. Wulder, T. Hu, R. Bright, Q. Wu, H. Qin, Y. Li, E. Toman, B. Mallick, X. Zhang, and M. Brown, "Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm," *Remote Sensing of Environment,* vol. 232, p. 111181, 2019.

[25] L. Micheli, M. Muller, E. F. Fernández, and F. Almonacid, "Segmentation of Deposition Periods: An Opportunity to Improve PV Soiling Extraction," in *47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020.