Quantifying Soiling Loss Directly From PV Yield

Michael G. Deceglie, Leonardo Micheli, and Matthew Muller

***Abstract*—Soiling of photovoltaic (PV) panels is typically quantified through the use of specialized sensors. Here, we describe and validate a method for estimating soiling loss experienced by PV systems directly from system yield without the need for precipitation data. The method, termed the stochastic rate and recovery (SRR) method, automatically detects soiling intervals in a dataset, then stochastically generates a sample of possible soiling profiles based on the observed characteristics of each interval. In this paper, we describe the method, validate it against soiling station measurements, and compare it with other PV-yield-based soiling estimation methods. The broader application of the SRR method will enable the fleet scale assessment of soiling loss to facilitate mitigation planning and risk assessment.**

***Index Terms*—Field performance, Monte Carlo methods, photovoltaic cells, photovoltaic systems, soiling, solar energy, solarpanels, time series analysis.**

I. Introduction

SOLAR panels can get dirty, blocking light from reaching the cell and reducing photocurrent. Typically, risk for a given installation is quantified by deploying a soiling station for a period of time, along with other irradiance and meteorological sensors prior to plant construction. Such soiling stations generally consist of a solar cell or module that is periodically cleaned (daily, weekly, etc.) and a cell or module that is left to soil naturally. Comparison between the two sensors enables the estimation of soiling loss that would be expected for a photovoltaic (PV) system [1], [2].

An alternative to the soiling station approach is to estimate soiling loss directly from PV energy yield. Such yield-based methods provide an important supplement to station-based measurements. One advantage of the PV-yield-based estimation is that soiling stations are typically deployed for limited terms; using a yield-based method provides the opportunity to understand historical trends. This is particularly important because year-to-year variations in soiling loss can be substantial [3]. The realization of a validated yield-based soiling loss estimation method enables the quantification of soiling and associated risk at fleet scale using existing data without the need for additional equipment.

In this contribution, we describe such a yield-based soiling estimation method and compare it with existing methods. The method, which we term the stochastic rate and recovery (SRR) method, is based on the detection of soiling intervals between cleaning events and a Monte Carlo simulation of the soiling profile based on the characteristics of the intervals. We validate the method by considering the pseudoyield of the dirty sensor in 11 soiling stations located around the United States. By comparing the SRR estimate with the soiling loss calculated from the full soiling station datasets, we are able to validate the SRR method. In addition, we compare the SRR method with prior state-of-the-art yield-based estimates and show that it outperforms previous methods. It is also important to note that the SRR method automatically adjusts to the dataset without the need for parameter tuning and does not require precipitation data, which is often unreliable.

We begin by describing the data collection and reduction and the steps of the yield-based SRR soiling estimation algorithm we have developed. We also describe the steps of alternative yield-based estimates based on a fixed rate and precipitation threshold, termed fixed rate precipitation (FRP) here [4], [5]. Finally we discuss the validation results and their implications.

II. Methods

This section describes the data processing and methods used in our analysis. The PV-yield-based soiling estimates are designed to utilize a time series of daily performance index (PI), the ratio of realized to modeled PV energy yield. The estimates are most accurate when the yield is modeled in more detail, for example, to exclude seasonality and spectral biases. In addition, the filtering of problematic data points associated with clipping (due to high dc–ac ratio) or equipment outages is useful to ensure the most accurate results.

The goal of both the station-based and yield-based soiling loss calculations is the insolation-weighted soiling ratio ($r_{s,w}$), given by

$$r_{s,w} = \frac{\sum_i G_{\mathrm{poa},i} \times r_{s,i}}{\sum_i G_{\mathrm{poa},i}} \tag{1}$$

where $G_{\mathrm{poa},i}$ is the plane-of-array insolation on day i and $r_{s,i}$ is the soiling ratio on day i. The soiling ratio is the fraction by which the photocurrent is reduced due to soiling; e.g., an r_s of 0.9 represents 10% soiling loss. The insolation-weighed soiling ratio is superior to average soiling ratio or soiling rate as a loss metric because it captures the actual expected energy loss due to

Manuscript received October 19, 2017; revised December 1, 2017; accepted December 8, 2017. Date of publication January 23, 2018; date of current version February 16, 2018. This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308 with Alliance for Sustainable Energy, LLC, the Manager and Operator of the National Renewable Energy Laboratory. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. *(Corresponding author: Michael G. Deceglie.)*

M. G. Deceglie and M. Muller are with the National Renewable Energy Laboratory, Golden, CO 80401 USA (e-mail: michael.deceglie@nrel.gov; matthew.muller@nrel.gov).

L. Micheli is with the National Renewable Energy Laboratory, Golden, CO 80401 USA, and also with the Colorado School of Mines, Golden, CO 80401 USA (e-mail: leonardo.micheli@nrel.gov).

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

Digital Object Identifier 10.1109/JPHOTOV.2017.2784682

soiling, taking into account whether soiling occurs during high- or low-irradiance periods.

A. Soiling Station Data Collection and Calculations

Our analysis utilizes data from 11 soiling stations installed in the United States. The datasets ranged in length from 226 to 1063 days. A soiling station is a system composed of at least two PV cells (or modules) exposed under the same conditions. One of the PV devices is regularly cleaned (control device), and the second device is left to naturally soil (soiled device). This way, soiling accumulated over the soiled device surface can be quantified by comparing the electrical output of the two devices. The cleaning frequencies of the control devices of the stations considered in this study ranged between 2 and 14 days (detailed cleaning logs were not available).

Here, we use the hourly short-circuit current ($I_{\mathrm{sc},h}$) of each device. The data have been filtered and converted into daily values as in [6]: Only data recorded between 12 and 2 P.M. for irradiance values higher than 500 W/m^2 have been considered. This aggressive filtering is necessary based on the configuration of the stations, including the alignment of the devices and nearby sources of shade; less stringent filtering is appropriate when yield-based extraction is performed on PV system yield.

We calculated the daily plane-of-array insolation ($G_{\mathrm{poa},i}$) from National Solar Radiation Database (NSRDB) Physical Solar Model data that uses satellite data about clouds to model irradiance at the earth's surface [7], [8]. For one site, the NSRDB irradiance caused a high level of noise; therefore, we resorted to using ground-based irradiance measurements to calculate $G_{\mathrm{poa},i}$. Generally, the use of ground-based measurements is discouraged for soiling estimation because irradiance sensors can soil along with the PV system. The use of satellite-based data avoids this problem. We corrected the effects of spectral and angle-of-incidence losses using standard correction algorithms [9], [10]. For the FRP methods, we utilized precipitation data from [11].

We used the soiling station data to calculate a daily soiling ratio ($r_{s,i}$) and to extract a daily pseudoperformance index. The station-based daily soiling ratio is the ratio of the average daily short-circuit current of the soiled device to that of the control device [6]. We used this station-based $r_{s,i}$ to calculate $r_{s,w}$ according to (1). This station-based $r_{s,w}$ is used as a basis to validate and compare the yield-based estimates considered here.

To quantify soiling loss in PV systems, the yield-based methods we consider would be applied to the PI, the ratio of realized to modeled yield of the system. Here, we seek to validate the yield-based methods based solely on station data. Thus, we construct a pseudoperformance index from the current of the soiled device, which we term performance metric (PM) defined for the ith day as

$$\mathrm{PM}_i = \frac{\sum I_{\mathrm{sc},h}}{G_{\mathrm{poa},i}} \tag{2}$$

where the numerator is the sum of the soiled-device short-circuit current measurements over a day. We then normalized PM_i by its 95th percentile. We used this normalized daily time series of PM_i as the input into both the SRR and FRP yield-based estimation methods.

B. Stochastic Rate and Recovery Estimation Method

The SRR method consists of the following four steps, shown in Fig. 1, and described in detail below.

1) Automatically detect cleaning events, dividing the dataset into soiling intervals.
2) Estimate the slope of each interval and the cleaning magnitude between intervals.
3) Using the uncertainty of each interval's slope and the cleaning magnitudes, stochastically generate possible soiling profiles for the entire dataset.
4) Calculate $r_{s,w}$ for each randomly generated profile and use resulting distribution of $r_{s,w}$ values to calculate the median and confidence interval.

The method presented here is similar to prior work [3], but has been improved to capture uncertainty associated with cleaning events and outages or missing data along with improved automatic detection of cleaning events.

The first step is to automatically detect cleaning events. We achieve this by finding positive shifts in a centered 14-day moving median of daily PI or PM, as shown in Fig. 1(a). The moving median provides smoothing (similar to a moving average), but preserves steps. The series of differences (Δ) between neighboring values is calculated for the moving median. All positive shifts greater than $Q_3 + 1.5 \times \mathrm{IQR}$ are identified as cleaning events, where IQR is the interquartile range and Q_3 is the third quartile of all the difference magnitudes $|\Delta|$. Thus, only positive shifts that are outliers in magnitude (by this definition) are identified as cleaning events. This method of identifying cleaning events automatically adjusts to the noise level of the dataset to distinguish cleaning events. However, we note that the cleaning detection will be more effective when the noise is lower and artifacts are fewer in a given dataset. The periods between these detected cleaning events are referred to as soiling intervals.

The second step is to determine the slope of each interval and the recovery magnitude between them. The slope is determined using the Theil–Sen method [12], [13], which is robust to outliers. This method, visualized in Fig. 1(b), estimates the slope as the median of all slopes between pairs of points. Similarly, the intercept is estimated according to work presented in [14]. The Theil–Sen method also enables the calculation of a 95% confidence interval for the slope [12], which is used to enable the Monte Carlo simulation of step 3. The magnitude of each cleaning event is calculated from the difference between the fit value at the beginning of an interval and the end value of the fit of the preceding interval. This magnitude (M) is taken as a lower bound to the cleaning magnitude, in other words, we can say with confidence that the cleaning was at least of this magnitude.

The third step is to perform a Monte Carlo simulation, stochastically generating possible soiling profiles of the daily r_s for the entire length of the dataset. We used 1000 repetitions in each Monte Carlo simulation. An example of a Monte Carlo

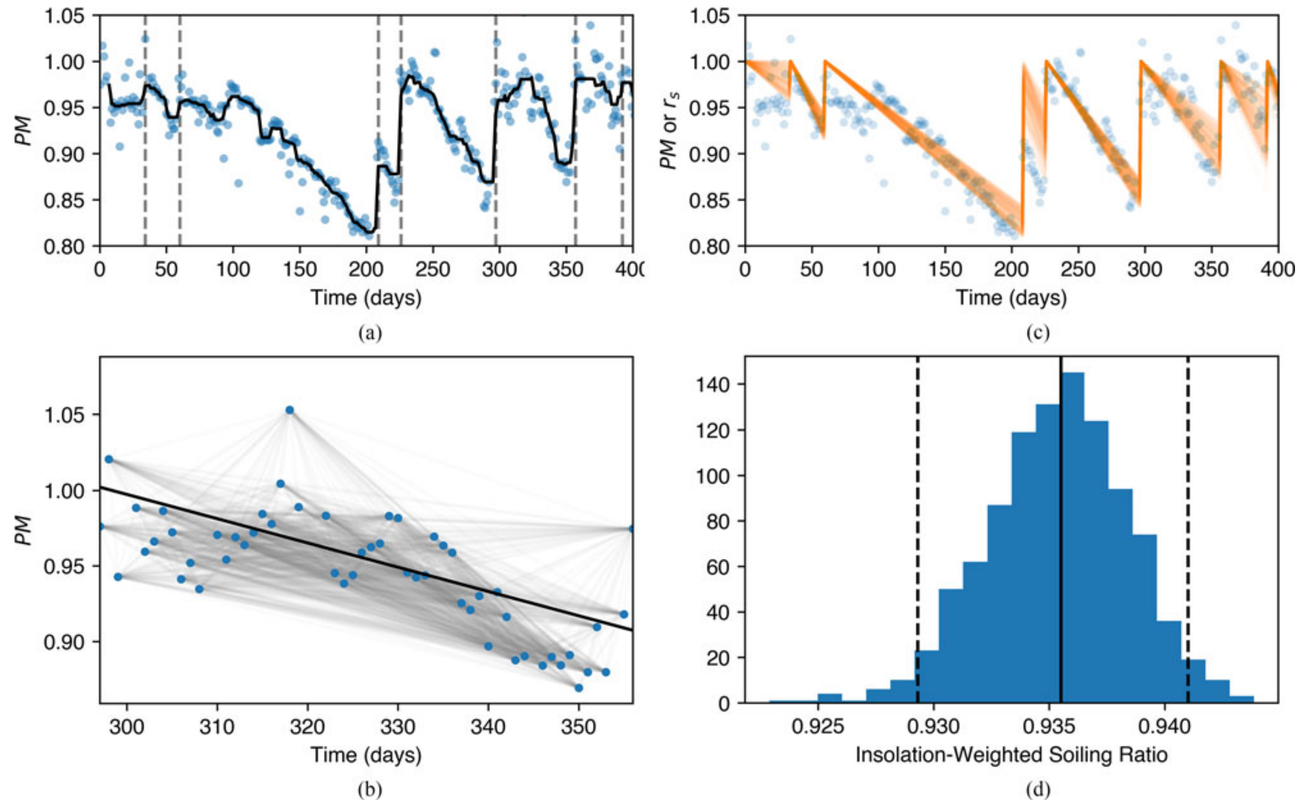

Fig. 1. Steps of the yield-based SRR soiling estimation method. Blue circles are the daily performance metric, PM. (a) Time series of PM is divided into soiling intervals by detecting positive shifts in the moving median (black line), which correspond to cleaning events (dashed lines). (b) Slope of each interval is extracted via the Theil–Sen method. The gray lines illustrate the lines between all possible pairs of points, the black line shows the best fit found from the median slope. (c) Possible soiling profiles are stochastically generated. The orange lines show a subset of these profiles. (d) Sample of $r_{s,w}$ values calculated from the stochastically generated profiles. The solid line shows the median, and the dashed lines indicate the 95% confidence interval.

simulation is shown in Fig. 1(c). Each profile is generated by starting the daily r_s at one for the beginning of the dataset and randomly selecting a slope for each soiling interval from a uniform distribution spanning the 95% confidence interval for each slope. If the confidence interval includes positive slopes, the uniform distribution of possible slopes is clipped to an upper bound of zero. If an interval is problematic in that it has greater than 500% uncertainty in slope, a positive slope, or a negative step in the moving median of greater than 5%, then the slope is taken to be zero for that interval. The recovery in r_s at each cleaning event is randomly selected from a half-normal distribution with its peak representing a perfect recovery to $r_s = 1$ and its 3σ point representing recovery by the inferred lower limit M.

The fourth and final step is to calculate $r_{s,w}$ for each randomly generated soiling profile. We then calculate the median and the 95% confidence interval (the 2.5 percentile to the 97.5 percentile) of the resulting sample of $r_{s,w}$, as shown in Fig. 1(d).

When an outage or period of rejected data longer than 14 days occurs, the r_s is taken as a constant for that time period; unchanging from the endpoint of the previous soiling interval in a given realization of the Monte Carlo simulation. If irradiance data is available during the outage, this constant but randomized soiling ratio will capture information about the likely soiling state during that period. If irradiance data is unavailable, we neglect the period in our analysis.

The implementation used for this contribution is not optimized for computational speed. However, a sense of the scaling potential can still be gained by considering performance on a laptop computer. For a typical multiyear dataset, steps 1 and 2 are completed <500 ms. Generation and analysis of each Monte Carlo realization takes <100 ms. For 1000 Monte Carlo realizations, this leads to a time of >1 min. However, the Monte Carlo realizations can be generated in parallel for applications where this speed becomes limiting.

C. Fixed Rate Precipitation Estimation Methods

Kimber *et al.* [4] proposed a method to calculate daily soling loss ($1 - r_s$). The method requires the following inputs:

1) Soiling rate: Slope of the performance metric/index profile during the longest dry period. It is obtained by extracting the slope of the linear regression of the performance metric over the longest dry period.
2) Cleaning threshold: Minimum amount of daily precipitation required to have a cleaning effect on PV modules. In this paper, precipitation events higher than cleaning threshold are referred to as *effective rainfalls*.
3) Refractory period length: A number of days after an effective rainfall for which no soiling occurs.

The soiling profile is determined by setting $r_s = 1$ on every effective rainfall and for the following refractory period. At the

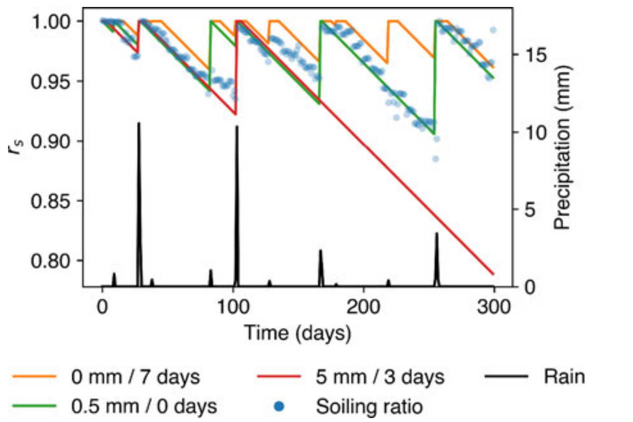

Fig. 2. Soiling ratio ($r_{s,w} = 0.970$) and the result of the FRP method under three different input combinations: 0 mm cleaning threshold and 7-day refractory period ($r_{s,w} = 0.989$), 5 mm cleaning threshold and 3-day refractory period ($r_{s,w} = 0.925$), and 0.5 mm cleaning threshold and 0-day refractory period ($r_{s,w} = 0.969$).

end of each refractory period, r_s is decreased each day according to the soiling rate until the next effective rainfall.

The refractory period can account for changing environmental conditions after wet periods. For example, it may be that dust is reduced for a while after significant rainfall until the local soil moisture content drops below a critical threshold. Prior work has shown that refractory periods (called "grace periods" in prior work) can vary in length and are only appropriate for some sites, complicating fleet-scale analyses [4].

In order to accurately estimate the soiling loss, the cleaning threshold and the refractory period length must be correctly chosen. These parameters may in fact be site specific and system specific. In this work, we consider a combination of different cleaning thresholds (0.0, 0.5, and 5 mm) and length of the refractory period (0, 3, 7, and 14 days). These produced 12 different combinations, whose results have been compared with those of the SRR method described in Section II-B. A subset of the FRP embodiments with different precipitation thresholds and refractory period lengths are shown in Fig. 2.

III. Results and Discussion

We compared the $r_{s,w}$ results of the SRR calculation described in Section II-B and the FRP methods described in Section II-C with the $r_{s,w}$ calculated directly from the soiling stations, as described in Section II-A. Although there is uncertainty associated with the station-based values of $r_{s,w}$, we treat the station-based values as the true values of $r_{s,w}$ for the purposes of comparing the methods presented in Sections II-B and II-C.

For each of the estimation methods considered, we calculated the root-mean-square error (RMSE), the coefficient of determination (R^2), and the mean signed deviation (MSD) relative to the station-based $r_{s,w}$. The results are presented in Table I. The SRR method outperforms the FRP methods in R^2 across all the FRP parameters considered here. The lowest RMSE was observed in the SRR method as well as three embodiments of the FRP method (rows 4, 11, and 13 in Table I). Several embodiments of the FRP method exhibit less bias than the SRR method, as evidenced by the MSD values. However, we believe the bias is likely due to an artifact in the station-based $r_{s,w}$ from cosoiling of the clean and dirty sensor between cleanings [15]. This would cause station $r_{s,w}$ values to be biased high, leading to a negative MSD, which is observed for the SRR method.

TABLE I
Performance of Yield-Based Soiling Estimation Methods

	Method	R^2	RMSE	MSD
1	SRR	0.87	0.009	−0.007
2	FRP: 0 mm, 3 days	0.81	0.010	0.007
3	FRP: 0 mm, 7 days	0.80	0.012	0.009
4	FRP: 0 mm, 0 days	0.80	0.009	0.005
5	FRP: 5 mm, 14 days	0.79	0.035	−0.019
6	FRP: 5 mm, 7 days	0.78	0.039	−0.023
7	FRP: 0 mm, 14 days	0.78	0.014	0.011
8	FRP: 0.5 mm, 7 days	0.77	0.010	0.005
9	FRP: 5 mm, 3 days	0.77	0.042	−0.025
10	FRP: 0.5 mm, 14 days	0.77	0.012	0.008
11	FRP: 0.5 mm, 3 days	0.77	0.009	0.003
12	FRP: 5 mm, 0 day	0.77	0.044	−0.027
13	FRP: 0.5 mm, 0 day	0.76	0.009	0.001

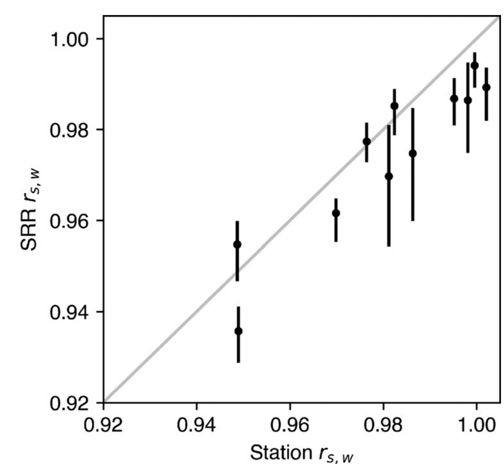

Fig. 3. Comparison between the SRR soiling estimates of $r_{s,w}$ and those calculated from soiling station data for 11 stations around the United States. The points represent the median SRR $r_{s,w}$ and the bars show the 95% confidence interval. The gray line illustrates perfect agreement.

In addition to its performance advantages, another advantage of the SRR method relative to the FRP methods is illustrated by the range of performance for FRP methods given in Table I. For example, the RMSE varies from 0.009 (equivalent to the SRR method) to 0.044, depending on the cleaning threshold and refractory period length used. The advantage of the SRR method is that there are no such parameters to tune. The only parameter in the SRR method is the length of time used for the moving median. We found a period of 14 days works well for this parameter in all the datasets we have considered.

Fig. 3 shows a comparison between the SRR estimates of $r_{s,w}$ and those calculated from the full station measurements. Generally, good agreement and a high degree of linearity are observed. However, we note that the SRR values tend to be lower than the station values. We attribute this to the cosoiling effect on stations [15]. When the clean and dirty cells/modules soil together, the r_s appears unchanged, even though soiling may still be occurring. This would tend to overestimate the true r_s and, thus, $r_{s,w}$. The yield-based metrics are generally more resistant to this problem, because they are based on comparison with theoretical yield, not a measured ratio in the field. The cosoiling bias can be addressed by frequently cleaning the reference device in a soiling station. Future development and, in particular, the validation of yield-based soiling estimates, such as the SRR method, will benefit from further study utilizing station data with more frequently cleaned control devices.

IV. Conclusion

We have described a method (SRR) for automatically estimating the soiling loss directly from PV yield data. Using a pseudoyield-based performance metric derived from the soiled device in soiling stations, we validated the method against soiling losses calculated from full soiling station datasets. We compared the method with the commonly employed FRP methods and observed that it outperforms the FRP methods in coefficient of determination while achieving the lowest RMSE value observed in this study.

In addition to the performance advantage over the FRP methods, the SRR method has the advantage of not requiring parameters, such as threshold precipitation or refractory period length. The SRR method also takes into account different soiling rates that may change over time or seasonally. Another advantage is that it detects cleaning events with an automatically determined threshold based on the noise in the dataset.

Continued advancement of the SRR method will likely follow the analysis advances made in the degradation rate analysis, including the use of clear-sky-filtered and clear-sky-modeled yield [16] and improved automatic filtering of, for example, outages and clipping. These advances apply to the generation of the PI time series (a key for degradation rate analyses as well) rather than to the SRR method itself. The performance of the SRR method is expected to improve for cleaner PI time series. Other future improvements may include the inclusion of abrupt soiling events, nonlinear soiling trends, and metrics associated with soiling seasonality.

The SRR method enables the fleet-scale analysis of historical soiling loss. This unlocks the potential of PV data that the community has already been collecting for decades. Broader applications of the method will enable the development of soiling predictions and risk quantification based on site characteristics.

Acknowledgment

The authors would like to thank S. Kurtz, T. Silverman, and L. Simpson for insightful discussions. The authors would also like to thank partners who shared soiling station data used in this analysis.

References

[1] J. R. Caron and B. Littmann, "Direct monitoring of energy lost due to soiling on first solar modules in California," in *Proc. IEEE 38th Photovolt. Spec. Conf.*, 2012, vol. 6, pp. 1–5.

[2] M. Gostein, J. R. Caron, and B. Littmann, "Measuring soiling losses at utility-scale PV power plants," in *Proc. IEEE 40th Photovolt. Spec. Conf.*, 2014, vol. 6, pp. 0885–0890.

[3] M. G. Deceglie, L. Micheli, and M. Muller, "Quantifying year-to-year variations in solar panel soiling from PV energy-production data," in *Proc. IEEE 44th Photovolt. Spec. Conf.*, 2017, to be published.

[4] 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 *Proc. IEEE 4th World Conf. Photovolt. Energy Conf.*, 2006, vol. 2, no. 5, pp. 2391–2395.

[5] F. A. Mejia and J. Kleissl, "Soiling losses for solar photovoltaic systems in California," *Solar Energy*, vol. 95, pp. 357–363, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X13002582

[6] L. Micheli, M. Muller, and S. Kurtz, "Determining the effects of environment and atmospheric parameters on PV field performance," in *Proc. IEEE 43rd Photovolt. Spec. Conf.*, 2016, vol. 6, pp. 1724–1729.

[7] National Renewable Energy Laboratory, "National Solar Radiation Data Base (NSRDB)." [Online]. Available: https://nsrdb.nrel.gov/. Accessed on: Nov. 21, 2016.

[8] M. Sengupta *et al.*, "A physics-based GOES satellite product for use in NREL's National Solar Radiation Database," 2014. [Online]. Available: https://www.nrel.gov/docs/fy14osti/62237.pdf

[9] M. Lee and A. Panchula, "Spectral correction for photovoltaic module performance based on air mass and precipitable water," in *Proc. IEEE 43rd Photovolt. Spec. Conf.*, 2016, vol. 6, pp. 1351–1356.

[10] A. Souka and H. Safwat, "Determination of the optimum orientations for the double-exposure, flat-plate collector and its reflectors," *Solar Energy*, vol. 10, no. 4, pp. 170–174, 1966.

[11] "PRISM Climate Group." [Online]. Available: http://prism.oregonstate.edu. Accessed on: Sep. 28, 2016.

[12] P. K. Sen, "Estimates of the regression coefficient based on Kendall's tau," *J. Amer. Statist. Assoc.*, vol. 63, no. 324, pp. 1379–1389, 1968.

[13] H. Theil, "A rank-invariant method of linear and polynomial regression analysis," in *Henri Theil's Contributions to Economics and Econometrics*. Berlin, Germany: Springer, 1992, pp. 345–381.

[14] W. Conover, *Practical Nonparametric Statistics*, 2nd ed. New York, NY, USA: Wiley.

[15] M. Muller, L. Micheli, and A. A. Martinez-Morales, "A method to extract soiling loss data from soiling stations with imperfect cleaning schedules," in *Proc. IEEE 44th Photovolt. Spec. Conf.*, 2017, to be published.

[16] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV degradation methodology and application," *IEEE J. Photovolt.*, to be published, doi: 10.1109/JPHOTOV.2017.2779779.

Authors' photographs and biographies not available at the time of publication.