A Scalable Method for Extracting Soiling Rates from PV Production Data

Preprint

Michael G. Deceglie, Matthew Muller,
and Sarah Kurtz
National Renewable Energy Laboratory

Zoe Defreitas
SunPower Corporation

Presented at the 43rd IEEE Photovoltaic Specialists Conference
Portland, Oregon
June 5–10, 2016

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Conference Paper
NREL/CP-5J00-65763
June 2016

Contract No. DE-AC36-08GO28308

NOTICE

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at SciTech Connect http:/www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI http://www.osti.gov
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandra, VA 22312
NTIS http://www.ntis.gov
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content.

A Scalable Method for Extracting Soiling Rates from PV Production Data

Michael G. Deceglie[1], Matthew Muller[1], Zoe Defreitas[2] and Sarah Kurtz[1]

[1]National Renewable Energy Laboratory, Golden, Colorado, 80401, United States
[2]SunPower Corporation, San Jose, California, 95134, United States

Abstract—**We present a method for analyzing time series production data from photovoltaic systems to extract the rate at which energy yield is affected by the accumulation of dust, dirt, and other forms of soiling. We describe an approach that is based on prevailing methods, which consider the change in energy production during dry periods. The method described here builds upon these methods by considering a statistical sample of soiling intervals from each site under consideration and utilizing the robust Theil-Sen estimator for slope extraction from these intervals. The method enables straightforward application to a large number of sites with minimal parameterization or data-filtering requirements. Furthermore, it enables statistical confidence intervals and comparisons between sites.**

I. Introduction

The soiling of photovoltaic (PV) panels is an important factor affecting the energy output of PV systems. One approach to quantifying soiling rates at different locations involves comparisons between a naturally soiled and a frequently cleaned sensor [1]. The sensors can be either reference cells, modules, or some other type of sensors. The clean device must either be manually cleaned or automated cleaning equipment must be used. While these approaches can offer high accuracy, a challenge is the cost associated with deploying such soiling stations at a large number of sites. We consider an alternative approach: extracting soiling rates directly from PV system production data.

The use of PV production data enables the quantification of soiling risk across many sites without additional hardware. To make meaningful comparisons between sites, it is important that a soiling rate extraction method enable statistically rigorous comparisons while remaining flexible to the source of the data and the varying levels of metadata available about different sites. In this paper we will outline an approach to extracting soiling rates from PV production data that builds upon prevailing methods for soiling rate extraction [2]–[4], and explain how it supports these goals.

Generally, soiling of a PV system can be assessed by comparing actual PV production to some sort of performance model. These performance models can take on a continuum of complexities, from just considering nameplate rating, to simple temperature correction, to more detailed system modeling accounting for solar spectrum and shade. One goal of the method presented here is to be agnostic to the level of detail of the performance model, while providing meaningful information about the uncertainty of the soiling rate.

Being performance-model agnostic will enable soiling rate extraction from a large number diverse data sets from different sources. This, in turn, will support the quantification of soiling risk at global scale. In order to realize this breadth, it will be important to extract soiling rates from a large number of PV sites, without strict limitations of the data available for that site. Production data from different sources may have more or less information available for detailed performance modeling, but it is still desirable to make a meaningful statistical comparison between them.

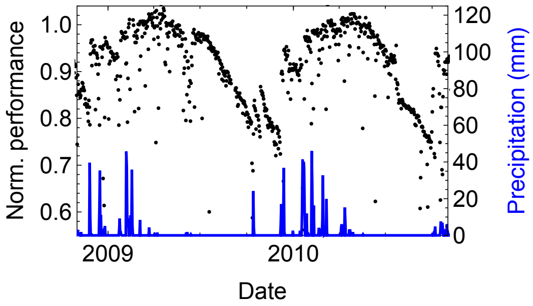

Fig. 1. A subset of energy production data from site A. The plot shows a time series of normalized performance metric, PM_{norm}, which has been insolation- and temperature-corrected to mitigate some seasonality effects (black) along with daily precipitation (blue). Soil accumulation during dry periods and recovery after rain events is apparent.

Existing methods for extracting soiling rates have been limited by the systems they could consider. For example, the analysis in [2] was only applied to sites with very strong soiling trends while that in [3] included weaker soiling trends, but heavy filtering was done to remove sites with data quality problems, inverter clipping, incorrect tilt angle specification, and other issues. Here, we only use a time series of production data that can be aggregated to daily production, along with minimal meteorological data; a detailed performance model of the system is not required.

In order for a soiling rate extraction method to scale to a large number of sites, it is also useful for it to avoid the use individually-determined site-specific parameters. Prevailing methods [2]–[4] for extracting daily soiling rates from PV production data rely on linear regression of periods between rainfall in conjunction with parameters including a minimum rainfall amount that cleans the panels, and a recovery period after rain during which time panels do not soil. These parameters must be determined for each system considered [4]. Using system-specific parameter values within the method reduces consistency and makes comparisons between sites less straightforward. We also use a robust regression method that reduces the effects of anomalies or lesser-quality data.

We begin by describing our method for quantifying daily soiling rates from PV production data. We then demonstrate the use of the method considering a case study of two different sites.

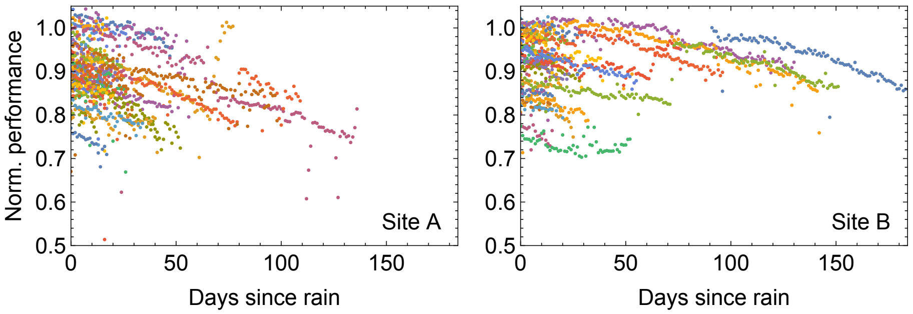

Fig. 2. The intervals of temperature-corrected and normalized performance metric, PM_{norm}, vs. days since rain, d, considered for the two sites. The different colors indicate different precipitation-free intervals. A slope is extracted from each of these intervals with the Theil-Sen method. The median of these extracted slopes in taken as a measure of the daily soiling rate for the site.

II. Soiling rate extraction method

The method we describe for determination of soiling rates from PV production data is a two-stage process. In the first stage, a daily performance metric is calculated based on a general model for the expected energy yield (for example, this could be performance index). In the second stage, the time series of the performance metric is analyzed along with precipitation data to determine a median daily soiling rate and the associated confidence interval. The focus of this work is on the second stage, which can be applied to various methods for calculating a daily performance metric with more- or less-detailed performance models.

A. *Performance metric calculation*

The first step in soiling rate extraction is to calculate a daily performance metric for the system under consideration. In this work we calculate a performance metric by temperature-correcting power measurements and comparing those to the daily plane-of-array insolation.

The inputs into our calculation are:

1) Time series of PV instantaneous power or energy production. In this study we used data with resolution of 15 minutes. Because we are concerned only with changes in performance relative to a system's peak performance, and thus will consider a normalized dimensionless performance metric, both power and energy measurements can be treated in the same way. In this paper we will use language assuming instantaneous power measurements for simplicity. The data considered here were not affected by inverter clipping, but the effects of clipping may be important for other systems.
2) Time series of ambient temperature. For this study we used 15-minute ambient temperature that was measured on-site. However there are other viable sources such as the National Solar Radiation Database in the United States (NSRDB) [5].
3) Time series of plane-of-array irradiance. We used 15-minute data that was measured on-site, however sources such as NSRDB could also be used. External sources can be particularly useful in cases where the irradiance sensor soils, is re-calibrated, or is otherwise adjusted.
4) Daily precipitation totals. We used data available from PRISM [6].

The first step in the performance metric calculation is to temperature-correct each power measurement, P, in the production time series to 25°C (T_0). In this work we used an empirical model for module temperature based on irradiance (neglecting wind speed) [7]. However, if module temperature measurements are available they can be used, relaxing the requirement for ambient temperature data. The temperature-corrected power, P_0, is calculated according to

$$P_0 = \frac{P}{1+\gamma(T-T_0)} \tag{1}$$

where γ is the power temperature coefficient for the modules in the array. Once calculated, the temperature-corrected power measurements are integrated over each day to give a temperature-corrected daily energy production for the system, E. This is combined with the daily insolation G (calculated by integrating the irradiance) and the array of the array A to give a daily performance metric, PM, according to

$$PM = \frac{E}{AG}. \tag{2}$$

To minimize impact of bias in the model and reduce the information required about a system, we normalize the daily values of PM to the 95th percentile of observed values for PM at a given site. This gives a dimensionless performance metric, PM_{norm}, appropriately sized relative to the near-peak performance of the system. In the results described here, area, A is not explicitly included in the calculations; it is normalized out. This approach avoids the need for a detailed performance model for the system and serves to isolate soiling losses from other system losses.

B. *Soiling rate calculation*

To calculate daily soiling rate, the daily PM data is considered along with daily precipitation totals. An example subset of the daily values for PM_{norm} along with daily precipitation totals is shown in Fig. 1. The figure illustrates the soiling signal to be extracted from the dataset; dry periods are associated with a decline in system output as soil accumulates on the panel surfaces. Rain events clean the system, but sometimes only partially. It is the rate of change in PM_{norm} during dry periods that we seek to extract as a daily soiling rate. Fig. 1 also highlights a challenge in quantifying annualized soiling; since not all rain events clean the system entirely, the effects

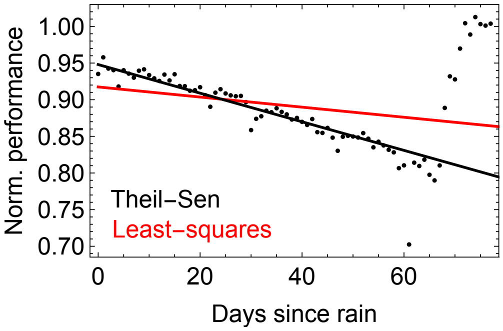

Fig. 3. PM_{norm} vs. d for an example precipitation-free interval from site A showing the advantage of the Theil-Sen method (black line) as compared to a least-squares linear regression (red line). This interval appears to have an unaccounted-for cleaning event near its end. This causes substantial skewing of linear regression, but the Theil-Sen method successfully extracts the soiling rate of interest. The use of the Theil-Sen estimator makes the proposed method more easily scalable to large numbers of sites without the need for anomaly filtering.

of partial cleaning and persistent soiling must be taken into account to quantify annualized soiling loss.

To extract the daily soiling rate, we proceed to calculate the number of days elapsed, d, between each day in the production dataset and the most recent preceding precipitation event. We place no threshold on the magnitude or intensity of precipitation event. We then partition the dataset into precipitation-free intervals and select only those intervals longer than 14 days. Plots of these intervals for the two sites considered in this study are shown in Fig. 2.

For each interval longer than 14 days, we use the Theil-Sen estimator [8] to extract a slope of PM_{norm} vs. d for that interval. The Theil-Sen estimator is calculated for a collection of points by calculating the slopes between all pairs of points in a dataset and then taking the median value of those calculated slopes. It is more robust to outliers than a least-squares linear regression. An example comparing the use of the Theil-Sen estimator to a least-squares linear regression is shown for one dry interval in Fig. 3. This example illustrates how the Theil-Sen estimator is more robust to anomalies; in this case an apparent cleaning event not associated with precipitation. This robustness is an advantage in scaling our soiling-rate extraction method to more sites, as it alleviates filtering requirements that may otherwise need to be tuned on a site-by-site basis.

Finally, once slopes have been calculated, the median of these slopes is taken as the metric for soiling at that site. The samples from different sites can also be statistically compared to one another. We use bootstrapping [9] to estimate confidence intervals for the median daily soiling rate.

C. Irradiance data considerations

An important consideration in carrying out these calculations is the source and nature of the irradiance data used in calculating the performance metric. If an irradiance sensor soils concurrently with the PV system, there is potential for bias to be present in the slopes of the performance metric during periods without rain. It's also important to note that different irradiance sensors may soil differently; for example a domed thermopile pyranometer may soil less than a reference cell. Because of the potential for the PV and irradiance sensor to soil concurrently, comparisons between different sites where the irradiance sensors are cleaned with different frequencies is not straight forward. In the two systems considered here, the irradiance sensors were cleaned annually.

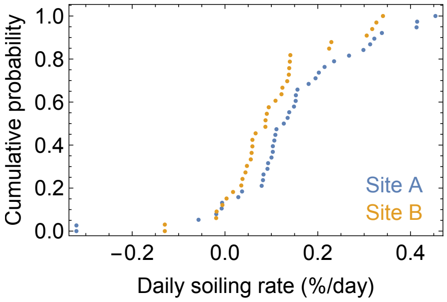

Fig. 4. Cumulative distribution functions for the soiling rates extracted from every interval at both sites. This shows that site A has more severe soiling than site B. Note that scatter and anomalous intervals sometimes give negative rates. To minimize the effect of such anomalies, we take the median of the observed rates as the soiling metric.

An interesting potential solution to the challenges around irradiance data is the use of modeled irradiance based on satellite data, such as that available from the NSRDB [5]. Noise introduced by the use of such satellite data will be reflected in the confidence intervals calculated in the soiling rate extraction method. However, bias will not be automatically captured by the soiling rate calculation. Detailed understanding of the application of satellite-based irradiance data to this soiling rate extraction method will be an important consideration for future efforts.

III. Application to field data

As an example application we consider results of the above calculations for two sites, A and B. We consider more than six years of production data from each site. Cumulative probability functions for the soiling rates extracted from each dry interval are shown in Fig. 4. It is also interesting to note in Fig. 4 that the distributions extend below zero. This is a result of noise in the data. Because of such results, we take the median value as the soiling metric for the site. The sign test on each sample gives p-values less than 10^{-4}, indicating that for both sites, the null hypothesis that the median of the population is equal to 0 can be rejected with high confidence. Thus we can conclude that there is statistically significant soiling at each site.

The distributions show that Site A generally has more severe soiling than Site B. The confidence intervals for the median soiling rate, calculated via bootstrapping [9] are compared for the two sites in Fig. 5. The confidence intervals shown in Fig. 5 support the conclusion that Site A has more severe soiling than Site B.

Finally, this approach allows useful statistics to be calculated around risk for a given site. For example, applying the bootstrap to the slopes from all the dry periods at each site allows us to conclude, with 97.5% confidence, that the median soiling rate at Site A is no worse than 0.17%/day and that the soiling rate for Site B is no worse than 0.12%/day.

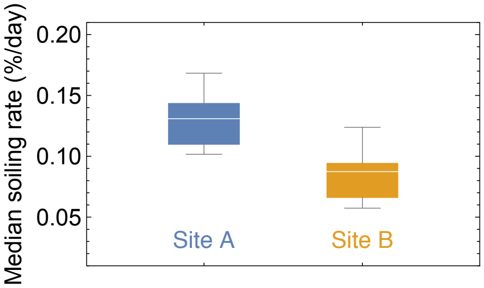

Fig. 5. Box and whisker chart illustrating the uncertainty in the median daily soiling rate at each site. The uncertainty was calculated via bootstrapping. The whiskers indicate the 95% confidence interval, the boxes indicate the central 50% confidence interval, and the white lines indicate the median value.

IV. Effects of PV modeling and seasonality

One of the strengths of the soiling rate calculation described in Section II-B is that it can be applied to performance indices calculated with varying levels of detail in a meaningful way. To demonstrate this, we carry out the soiling rate calculation described in Section II-B on synthesized time series of daily *PM* values. This enables us to calculate confidence intervals and compare them to a known soiling rate. One factor that performance models handle with varying degrees of accuracy is seasonality; that is, changes in the performance metric over the different seasons. We vary the amount of residual seasonality (that unaccounted-for in the performance model used to calculate *PM*) present in the synthesized time series and demonstrate that the soiling rate calculation method and associated statistics appropriately account for this uncertainty.

We synthesized datasets based on real precipitation data from Site A in order to capture the non-random seasonal nature of rainfall. The synthesized daily performance metric, PM_{synth} on the i^{th} day of operation is given by Equation 3.

$$PM_{synth} = NS\left(1 - Y\sin(2\pi i/365. - \phi)\right) \quad (3)$$

Here, N is a random noise factor drawn from a normal distribution about one with a standard deviation of 0.02. S is a soiling factor. S is calculated assuming a linear reduction in *PM* during precipitation-free periods. The slope of the reduction for each period is randomly drawn from normal distribution of soiling rates with a mean soiling rate of 0.15%/day and a standard deviation of 0.075%/day. On each day with precipitation, S recovers by a randomly chosen fraction between 0 and 1. The final term in Equation 3 represents residual seasonality in *PM*, where ϕ represents the phase relationship between the residual seasonality and the seasonal rain patterns and Y is the fractional amplitude of the residual seasonality. Here we consider the worst case scenario for ϕ, where the downward slope of the seasonality coincides with the dry season causing constructive interference in the PM_{synth} signal. An example of one such synthesized *PM* time series is shown in Fig. 6a.

This approach allows us to carry out the soiling rate calculation described in Section II-B with varying levels of residual seasonality. Fig. 6b shows the 95% confidence intervals for the median daily soiling rate calculated as described in Section II-B. We see that the confidence interval tends to expand at higher levels of residual seasonality, but continues to bracket the known median soiling rate in the synthesized time series. This demonstrates that the soiling rate calculation is robust to different performance models that will leave different levels of residual seasonality in the daily *PM* time series. This is an important feature of the approach, as it will allow meaningful comparisons between many different sites, even in the presence of varying levels of metadata available for performance modeling and *PM* calculation.

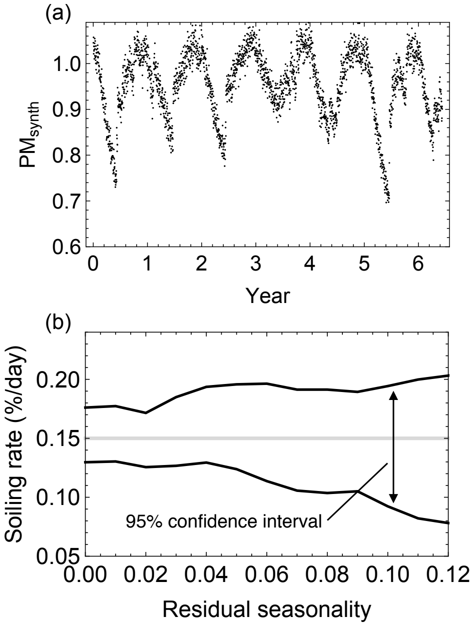

Fig. 6. (a) Time series of daily performance synthesized according to Equation 3 with a value for Y of 0.05. (b) 95% confidence interval in the median daily soiling rate calculated by applying the method described in Section II-B to synthesized *PM* time series for varying levels of residual seasonality (Y in Equation 3). Residual seasonality is that which remains in the data after application of a performance model for calculation of performance. The known true median soiling rate is indicated by the gray line. With varying levels of residual seasonality, the soiling rate extraction method yields a meaningful confidence interval that appropriately brackets the underlying true median soling rate.

V. Conclusion

We have presented a method for extracting median daily soiling rates from PV production data. The method is designed to scale to large numbers of sites in a straightforward, consistent, and robust way. As we have shown, the method can extract statistically significant daily soiling rates and allow comparisons of confidence intervals.

The method described here is also agnostic to the performance model used to calculate a performance metric. We demonstrated this by analyzing synthesized data sets and showing that the confidence intervals respond appropriately to increased seasonality which is not accounted for by the performance model. In practice, different PV systems have different level of detail in their available metadata, facilitating different levels of detail in their associated performance models. The method described here enables meaningful comparison between such diverse systems and supports the goal of building a world-wide and coherent understanding of risk factors for soiling.

ACKNOWLEDGMENT

The authors thank Timothy J Silverman and Dirk Jordan of the National Renewable Energy Laboratory and Michael Anderson of SunPower for insightful conversations in support of this work.

This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory. 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.

REFERENCES

[1] M. Gostein, J. R. Caron, and B. Littmann, "Measuring soiling losses at utility-scale pv power plants," in *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*, June 2014, pp. 0885–0890.

[2] 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*, vol. 2, May 2006, pp. 2391–2395.

[3] F. A. Mejia and J. Kleissl, "Soiling losses for solar photovoltaic systems in california," *Solar Energy*, vol. 95, pp. 357 – 363, 2013.

[4] J. Caron and B. Littmann, "Direct monitoring of energy lost due to soiling on first solar modules in california," *Photovoltaics, IEEE Journal of*, vol. 3, no. 1, pp. 336–340, Jan 2013.

[5] NREL, National Solar Radiation Database, https://nsrdb.nrel.gov/.

[6] Oregon State University, PRISM Climate Group, http://prism.oregonstate.edu/.

[7] D. L. King, J. A. Kratochvil, and W. E. Boyson, "Photovoltaic array performance model," Sandia National Laboratory, Tech. Rep. SAND2004-3535, 2004.

[8] P. K. Sen, "Estimates of the regression coefficient based on kendall's tau," *Journal of the American Statistical Association*, vol. 63, no. 324, pp. 1379–1389, 1968.

[9] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, ser. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.