Local Variability in PV Soiling Rate

Michael Gostein[1], Kendra Passow[2], Michael G. Deceglie[3], Leonardo Micheli[3], Bill Stueve[1]

[1]Atonometrics, Austin, TX, 78757, USA
[2]First Solar, San Francisco, CA, 94105, USA
[3]National Renewable Energy Laboratory, Golden, CO, 80401, USA

***Abstract* — We examine the correlation between monthly PV system soiling rates over a multi-year period within two groups of First Solar soiling stations in separate regions of California as a function of site separation. The results demonstrate large seasonal variations in monthly soiling rates, with seasonal patterns that differ between sites. Monthly patterns at nearby sites were more highly correlated, suggesting that soiling rates can be strongly affected by local geography and weather. For the two regions studied, soiling rate pattern correlation diminished beyond a distance on the order of 50 km.**

***Index Terms* — photovoltaic systems, performance analysis, solar power generation, solar energy.**

I. Introduction

Soiling of PV modules is one of the principal energy loss factors for PV systems and has a significant impact on PV power plant economics. Accordingly, in recent years there has been a growing interest in measuring, predicting [1]-[2], and monitoring PV soiling rates, for pre-production prospecting as well as operations and maintenance (O&M) purposes, especially at utility-scale PV power plants [3].

An important question concerns the local variability of soiling rates. Is it sufficient to use isolated soiling measurements as representative of a larger region? If so, what are the appropriate distance scales for which a given station's results are considered representative? Conversely, on a very small scale, to what degree does soiling rate vary within a single large PV power plant?

First Solar has deployed an extensive network of soiling measurement stations, at both prospecting and power plant locations, providing multi-year data. In this study, we examine historical soiling data at sites within two regions of California, in order to elucidate trends in local and regional soiling variations.

II. Data Sources

Table I lists the 8 sites used as data sources in this study, each of which is an operating PV power plant. Region 1 and Region 2 are located, respectively, in southern and central California, and are separated by approximately 600 km. Listed distances are measured from the first site listed for each region, and are rounded to the nearest 10 km. Each site has stations collecting both meteorological (met) and soiling data, with varying number of stations per site as indicated. The study includes data collected from July 2013 to December 2017. Each soiling station comprises a pair of PV modules, one of which is cleaned every 1-2 weeks, and the other of which is allowed to soil naturally. A soiling ratio is calculated which quantifies the electrical output of the soiled module versus the cleaned module [3], and each soiling station is paired with a met station which collects irradiance and rain data. Table I lists the annual insolation-weighted soiling ratio measured at each site, details of which are provided below. We also list in Table I a "land type" that indicates the nature of the land surrounding the site, using a combination of climate and land use classifications. Although sites 1A-1C and 2B-2D are in areas with desert/arid climate classifications, there is nonetheless extensive agricultural activity in their surrounding areas so we note this in our land type as it may affect soiling rates and seasonal patterns.

TABLE I
SOILING STATION SITES IN THE STUDY

Region	Site	Land Type	Stations	Distance (km)	Annual Soiling Ratio
1	1A	Farming	6	0	0.976
	1B	Farming	12	10	0.983
	1C	Farming	2	50	0.956
	1D	Desert	3	120	0.995
2	2A	Arid	9	0	0.995
	2B	Farming	1	30	0.996
	2C	Farming	1	30	0.997
	2D	Farming	1	150	0.987

III. Automated Data Analysis

Completely automated analysis techniques based on [4]-[6] were used to process the data from the 35 stations. To illustrate the data analysis method, we show in Fig. 1 a 12-month segment of data from one of the stations located at site 1A.

The upper panel in Fig. 1 shows the soiling ratio, defined in this case as the ratio of temperature-corrected short-circuit current between the soiling station dirty module and clean module [3][7]. Light gray points in this plot are minute-by-minute measured soiling ratio values for the period of noon +/- 2 hours on each day. The minute-by-minute points for each day are filtered to remove non-representative points (using a 1σ filter) and values at irradiance less than 500 W/m^2; a daily irradiance-weighted average, shown in red, is then calculated as described in [3]. Note that the daily average soiling ratio

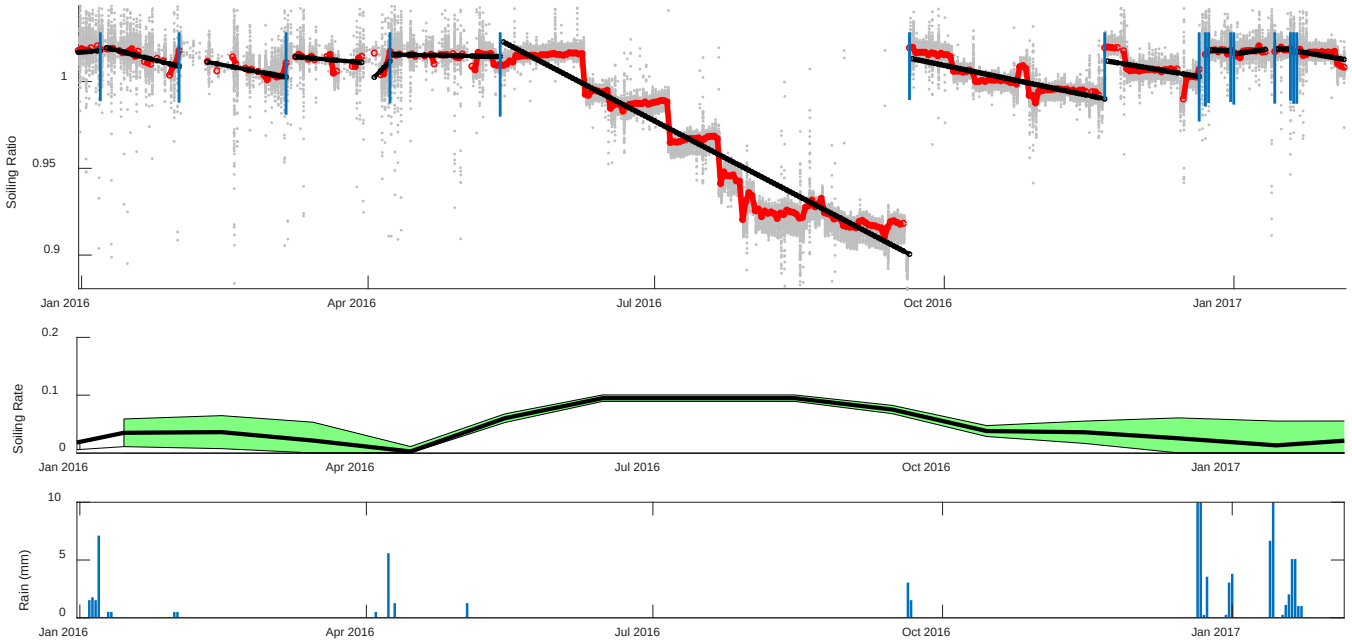


Fig. 1. Segment of data from a station at site 1A, illustrating the data analysis method. Upper panel: minute-by-minute measured soiling ratio (gray), daily average soiling ratio (red), detected soiling ratio reset events (blue), and linear fits between resets (black). Middle panel: soiling rate in %/day (black) and its confidence interval band (green). Lower panel: rain events.

values in this plot typically exceed 1.00 during clean periods. This is because the stations in this study were not on-site calibrated; rather, the measured values are normalized during analysis as will be described below.

During dry periods the daily average soiling ratio steadily decreases as soiling accumulates and losses increase. The soiling ratio appears to decrease in a step-wise fashion, due to the fact that cleaning of the clean module in the soiling station module pair was not automated and occurred only every 1-2 weeks. During rain events, plotted in the bottom panel, the soiling ratio value recovers or “resets”, due to cleaning of the entire system. Note that in some instances in Fig. 1 the daily average soiling ratio recovers to a clean-state value at times when there is no measurable rainfall. This can be caused by dew, non-measured rainfall, or even in a few cases washing of the entire power plant.

In order to quantify soiling rate in between reset events, reset events must first be detected from the soiling ratio data. We accomplish this by detecting outlier positive shifts in the 14-day moving median of the daily soiling ratio, as per [5], as well as designating any measured rain event as a reset event. Detected reset events are shown in the upper panel of Fig. 1 by vertical blue lines.

We assume soiling rate to be linear in between the reset events. Soiling rate is calculated by determining the slope of the daily soiling ratio values for each data segment in between resets, using the robust Theil-Sen method for slope estimation [4]-[6]. Ninety-five percent confidence intervals of the Theil-Sen slope are determined using bootstrapping [4]; for simplicity, we assumed confidence intervals are symmetric about the mean value. The middle panel of Fig. 1 shows determined soiling rates in %/day (solid line) and their associated confidence intervals (green shading). Soiling rates were determined for each segment of data in between reset events, and were then binned into monthly values by calculating the mean and combined confidence interval of all segments overlapping a given month, weighted by the fraction of the month corresponding to each segment. Note that the width of the confidence interval is primarily determined by the length of time between cleaning events; during the long dry period in the summer of 2016 the confidence intervals are very narrow, indicating a precise determination of the soiling rate, while during the winter months with their more frequent rains the confidence intervals on the soiling rate determination are wider.

Detection of reset events was also used to automatically normalize the daily average soiling ratio values by determining the soiling ratio measured when the station was clean. For each soiling station, the daily average soiling ratios measured 2 days after each reset event were collected into a distribution. This distribution was filtered with a 1σ filter to remove non-representative points, corresponding, for example, to resets that

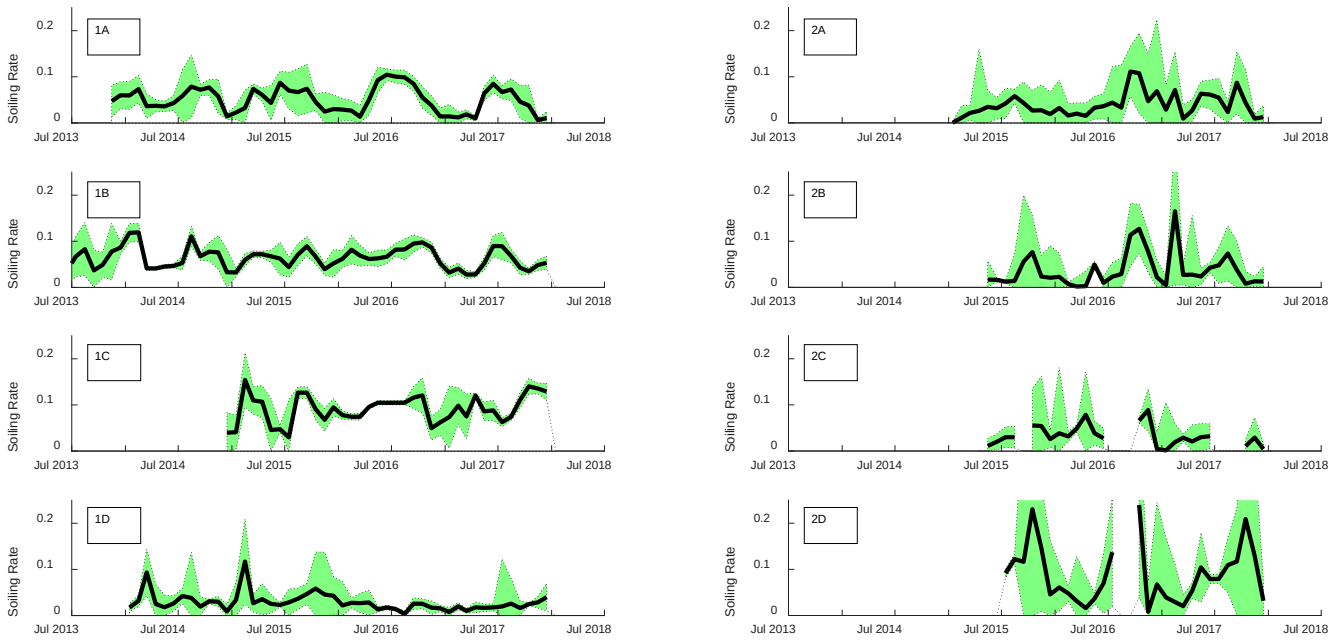


Fig. 2. Monthly soiling rates for the 8 sites in %/day (black) and their confidence interval bands (green).

resulted in only a partial cleaning, and the mean value of the resulting distribution was used as the measured clean-state soiling ratio for the station, typically differing from 1.00 by up to ~2%. This value was then used to normalize soiling ratio values for the station.

IV. Results

Fig. 2 shows the soiling rates and their confidence intervals for each of the 8 sites over the multi-year period. Stations within a given site generally showed similar trends, although with differences that may be explored in a later work. For Fig 2, all stations at a given site were averaged together for each month of data using a Monte Carlo resampling method to determine both the mean soiling rate at the site and its confidence interval, taking into account the soiling rates and confidence intervals of the individual stations.

For the nearby sites 1A and 1B, it is apparent that soiling rates display a strong seasonal trend, peaking around 0.1 %/day in the summer months and falling to near 0 in the winter. Possibly this seasonal pattern relates to agricultural activity in the surrounding region, which could generate airborne particulates. Sites 1C and 1D, which are farther away, show different patterns. Site 1D, which is located in a desert area across a mountain range from the other three sites, shows a completely different pattern.

In region 2, sites 2A and 2B appear to show soiling rates peaking to ~0.15 %/day in late summer or early autumn, although their pattern is less regular and contains inter-annual variations, and this pattern changes somewhat in moving to the farthest site, 2D.

To quantify the correlation of the monthly soiling rate patterns, we show in Fig. 3 the R-squared correlation coefficient for the linear regression of the monthly data series for site A in each region (i.e. site 1A or 2A) versus the other sites in the same region, plotted as a function of distance. The point at distance 0 corresponds to the A site versus itself and is identically 1. Site 2C was excluded from the plot because unfortunately its one soiling station was offline during the summers of 2016 and 2017 and consequently displayed only very limited range in its soiling rate data (as seen in Fig. 2). Fig. 3 illustrates that the monthly soiling rate patterns of the closest sites are correlated, and that the correlation between

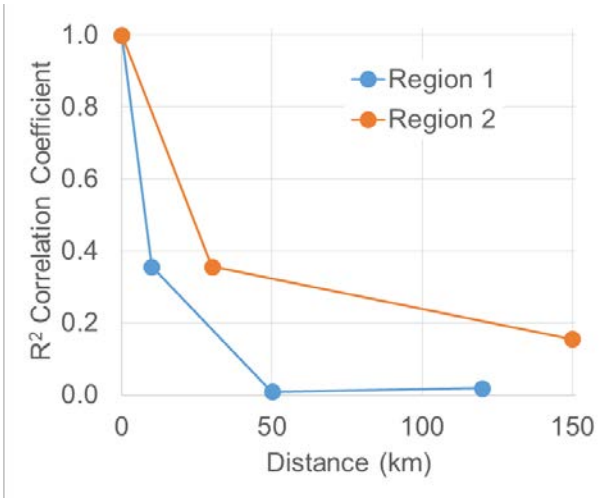


Fig. 3. Correlation between monthly soiling rate patterns versus distance for the sites in Table I.

sites diminishes with distance. The nearest sites have an R-squared coefficient of approximately 0.35, indicating at least partial correlation. In fact this value is similar to the correlation coefficients between stations within the large multi-station sites, which in good cases were in the range 0.3-0.5. However, note in Fig. 3 that the R-squared values decrease beyond a correlation distance on the order of ~50 km. These results suggests that, for the regions considered, a soiling station must be located within at least ~50 km of a PV plant for its data to be meaningful in soiling prediction at the plant. Of course this distance cannot be regarded as universal for other regions, but would be specific to the local geographic features of any region. We note however that the finding of a correlation length on the order of tens of kilometers is in agreement with a previous study that correlated soiling rates to airborne particulate matter density ($PM_{2.5}$ and PM_{10}) measured at varying distances from soiling stations [1].

Note that the R-squared coefficient provides us only a relative measure of correlation and does not directly quantify prediction accuracy. Ideally, we wish to know what would be the uncertainty in predicting PV power plant losses using soiling station data measured at varying distances from a PV plant. Analysis of our results to determine this prediction uncertainty will require further work.

We also attempted to correlate the monthly soiling rate of each site in the present study with the nearest ground-based measurements of airborne particulate matter density ($PM_{2.5}$ and PM_{10}). The particulate matter data were downloaded from the U.S. Environment Protection Agency (EPA) database [8], which contains data from an EPA network of particulate matter monitoring stations, measuring PM_{10} and $PM_{2.5}$ daily or every 6^{th} day. We calculated the monthly mean PM concentrations for each PV soiling site from the PM concentrations recorded at the closest monitoring station each month. In order to get the most complete data, only stations that took at least 75% of the yearly scheduled measurements have been considered.

At first look, we observed only a very weak correlation between monthly soiling rates and monthly particulate matter of each site, in contrast to the strong correlation observed in a previous study [1]. This is not surprising, as the previous study correlated yearly average data rather than monthly average and other researchers [9]-[10] have already discussed the challenges of predicting short-term soiling using single variable regression compared to longer-term predictions, due to the more complex interactions among the environmental parameters. In addition, in the present study, only data from the closest stations were considered; considering that the closest station could be as far as 93 km from a site, different spatial interpolation techniques could be employed in the future for better correlations. Moreover, the large confidence interval bands in some cases might have affected the results. Indeed, we noted that the correlations between soiling rates and PM_{10} could be enhanced for all the sites in Region 2, with a maximum R^2 higher than 0.7, if high-uncertainty soiling rate data were removed. Despite that, the same improvement could not be found in Region 1. Further work is needed to elucidate the correlation between particulate matter concentration and soiling rate, and to understand the impact that different weather patterns and soil characteristics can have on this correlation.

Finally, we also determined the average soiling ratio of each site. To do this, we first normalized the measured daily average soiling ratios of each soiling station by the clean-state soiling ratio value for that station determined from post-reset measurements, as described above. We then used daily insolation values to calculate an insolation-weighted average of all the daily soiling ratio values for each station during the years 2016 and 2017. Values for all stations within a site were averaged, in order to obtain the site-average soiling ratios shown in Table I. The overall soiling related losses of the sites in this study were in the range of <5%.

V. Conclusions

Our results demonstrate an example of large seasonal variations in monthly soiling rates in two nearby regions, with seasonal patterns that differ markedly between sites. Monthly patterns at two pairs of sites within approximately 50 km of each other were more strongly correlated, while farther sites showed significantly lower correlation. The results suggest that soiling rates can be strongly affected by local geography, land use, and weather, resulting in specific regional patterns. Therefore, there is a limited range of applicability for soiling station data. Even within a single site, we observed variations in soiling station data, indicating that a larger number of stations is beneficial for determining soiling at a given site more accurately. In additional work we plan to investigate more thoroughly the within-site variations in soiling station data and uncertainties associated with prediction of soiling loss as a function of station distance.

VI. Acknowledgements

This work was authored in part 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 No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 30311. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. 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.

REFERENCES

[1] L. Micheli and M. Muller, "An investigation of the key parameters for predicting PV soiling losses," *Prog. In Photovoltaics*, v. 25, Apr. 2017, pp. 291-307.

[2] W. Javed, B. Guoa, and B. Figgis, "Modeling of photovoltaic soiling loss as a function of environmental variables," *Solar Energy*, v. 157, Nov. 2017, pp. 397-407.

[3] M. Gostein, J. R. Caron, B. Littmann, "Measuring soiling losses at utility-scale PV power plants," in *40th IEEE Photovoltaic Specialists Conference*, Denver, CO, 2014.

[4] M. G. Deceglie, M. Muller, S. Kurtz, Z. Defreitas, "A scalable method for extracting soiling rates from PV production data," in *43rd IEEE Photovoltaic Specialists Conference*, Portland, OR, 2016.

[5] M. G. Deceglie, L. Micheli, M. Muller, "Quantifying soiling loss directly from PV yield," *IEEE Journal of Photovoltaics*, v. 8, March 2018, pp. 547-551.

[6] L. Micheli, D. Ruth, M. G. Deceglie, M. Muller, "Time series analysis of photovoltaic soiling station data," NREL/TP-5J00-69131, September 2017.

[7] M. Gostein, B. Littmann, J. R. Caron, L. Dunn, "Comparing PV power plant soiling measurements extracted from PV module irradiance and power measurements," in *39th IEEE Photovoltaic Specialists Conference*, Tampa, FL, 2013.

[8] US Environmental Protection Agency, "Air quality system data mart [internet database]." [Online]. Available: https://www.epa.gov/airdata. [Accessed: 15-May-2018]

[9] L. Boyle, H. Flinchpaugh, and M. Hannigan, "Assessment of PM dry deposition on solar energy harvesting systems: measurement–model comparison," *Aerosol Sci. Technol.*, vol. 50, no. 4, pp. 380–391, Apr. 2016.

[10] W. Javed, B. Guo, and B. Figgis, "Modeling of photovoltaic soiling loss as a function of environmental variables," *Solar Energy*, 157, 397 (2017).