Mapping Photovoltaic Soiling Using Spatial Interpolation Techniques

Leonardo Micheli,[1] Michael G. Deceglie,[1] Matthew Muller[1,2,3]

[1] National Renewable Energy Laboratory, Golden, Colorado, 80401, USA
[2] Universidad de Jaén, Jaén, 23071, Spain
[3] Leidos, Denver, Colorado, 80202, USA

***Abstract* — In this study, we present a new soiling map developed at NREL, showing data from 83 sites in the United States. Soiling has been measured through soiling stations or extracted by photovoltaic system performance data using referenced techniques. The data on the map have been used to conduct the first regional analysis of soiling distribution in the United States. We found that most of soiling occurs in the southwestern United States, with Southern California counties experiencing the greatest losses because of the high particulate matter concentrations and the long dry periods. Moreover, we employed five spatial-interpolation techniques to investigate the possibility of estimating soiling at a site using data from nearby sites. We found coefficients of determination of up to 78% between estimated and measured soiling ratios, meaning that, by using selective sampling, soiling losses can be predicted using the data on the map with a root-mean-square error of as low as 1.1%.**

***Index Terms* — photovoltaic systems, soiling, map, spatial interpolation.**

I. Introduction

The accumulation of dust, dirt, and particles on the surface of photovoltaic (PV) modules is a complex issue, known as *soiling*, that affects PV systems worldwide [1]. Soiling causes a reduction in the intensity of sunlight reaching the semiconductive material and therefore decreases the energy converted by the modules. Soiling causes drops in the annual energy yield of up to 7% in the United States, and power losses even higher than 50% have been reported worldwide [2]–[6].

For the last few years, the National Renewable Energy Laboratory (NREL) has been working to provide the solar community with tools and information to properly address this issue. In 2017, we presented a work on the key parameters for predicting soiling losses [7]: we found that particulate matter concentrations (expressed as PM_{10} and $PM_{2.5}$) and rainfall parameters describing the average and maximum length of the dry periods were the best predictors of soiling occurring at 20 sites in the United States. We also developed a *Stochastic Rate and Recovery* methodology to extract soiling losses directly from PV performance data [8]. This new model, based on the identification of soiling intervals between cleaning events and on a Monte Carlo simulation of the soiling profile, allows one to monitor soiling accumulated on PV systems without the need of soiling stations, thus dramatically increasing the amount of data already available on soiling and potentially limiting the costs associated with soiling detection.

In the present work, we introduce a new tool with the potential to help investors and operations and maintenance (O&M) teams better estimate soiling losses and, therefore, to improve the performance of their systems and increase their revenues, even if soiling or PV data are not available on site. Indeed, NREL has published an interactive soiling map (screenshot shown in Fig. 1), where soiling data from 83 locations in the United States have been collected [9]. In a previous world soiling map [10], data were shown as reported in literature, each calculated in a unique manner. The data in the new map are instead obtained with a systematic approach through consistent referenced methodologies.

The data shown in the map have been used to conduct a first-of-a-kind regional analysis of soiling in the United States to understand which regions are more susceptible to soiling losses. Moreover, we also investigated the usefulness of this map as a tool for estimating losses at new PV sites to understand if soiling at a site can be predicted by using data from nearby locations. In this effort, we employed five spatial interpolation techniques to estimate soiling losses at a site and compared them with actual soiling measurements.

Fig. 1. Screenshot of the soiling map published in October 2017 on www.nrel.gov/pv/soiling.html [9]. Triangular markers (▲) are soiling stations; squared markers (■) are PV installations. The markers are color-coded according to the severity of soiling.

II. Data and Methodology

A. Soiling: Quantification and Detection

Soiling is commonly quantified using two metrics: soiling ratio and soiling rate. The *soiling ratio* expresses the ratio between the output of the PV device and the same output in clean conditions. It has a value of 1 in clean conditions and decreases as soiling accumulates on the PV device's surface. If

TABLE I
METHODOLOGIES EMPLOYED TO DETERMINE THE SOILING LOSSES REPORTED ON THE SOILING MAP.

Type	Electrical Output	Conditions	Soiling Ratio		Soiling Rate	
			Mean Value	Uncertainty	Mean Value	Soiling Rate Range
Soiling Station	Short-circuit current	Only data between 11 a.m. and 1 p.m., for irradiance ≥ 500 W/m^2 and for which the normalized short-circuit current of the cleaned cell was at least 80% of its expected value were considered [7], [12]	Insolation-weighted mean of daily ratios of short-circuit current of dirty cell to short-circuit current of control cell [7]	Difference between calculated and modeled soiling ratio [24]	Median of slopes calculated using Theil-Sen estimator on each dry period longer than 14 days [11]	97.5th and 2.5th percentiles of soiling rate distribution calculated with method in [11]
PV System	Power or current at maximum power point	Clear-sky day (any day in which NSRDB's "Cloud Type" variable [25] has a value of 0)	Median of insolation-weighted soiling ratio range obtained by using the Stochastic Rate and Recovery method [8]	95% confidence interval of range of soiling ratios obtained by using the Stochastic Rate and Recovery method [8]		

soiling ratio is zero, no energy is produced by the PV device because soiling is blocking all the light from reaching the cell. The *soiling rate* measures daily rates of change in soiling ratio during dry periods [11]. It has a value of 0% per day if the soiling ratio does not change, and it assumes negative values when soiling accumulates on the surface of the PV device, thus lowering the soiling ratio.

The soiling data on the map have been sourced from both soiling station and PV system data, using the referenced methods listed in Table I. Soiling stations are used for a direct measurement of soiling at a site. They can have various geometries, but those used in this work are composed of two PV devices (reference cells or modules), with one device regularly cleaned (*control device*), while the second is left to naturally soil (*soiled device*). The soiling ratio for soiling stations can be quantified by dividing the electrical output of the soiled device to the same output of the control device. The methodology presented in [7], [12] has been followed: (i) only the central hours of the day have been considered so as to remove any bias due to the potential angular misalignment of the two reference devices and angle-dependent light scattering from soiling particles [13], [14], (ii) only data recorded for a minimum irradiance of 500 W/m^2 have been employed to filter out cloudy days that might increase the noise in the calculation of the soiling ratio, and (iii) only those hours in which the normalized short-circuit current of the cleaned cell was found to be higher than the 80% of its expected value have been considered in order to exclude periods in which the clean cell was malfunctioning or excessively soiled. These last two conditions were found to effectively reduce the noise of the soiling ratio calculation for the soiling stations installed in the USA; but they should be retuned if data from different climate conditions are considered.

Soiling occurring on PV systems has been determined by using the *Stochastic Rate and Recovery* method [8]. A performance metric is calculated as the ratio between the maximum power-point output and the plane-of-array irradiance; it is then used to determine the mean soiling ratio and median soiling rate through a Monte Carlo computation.

Note that different electrical outputs are available for the two types of sources. The soiling stations shown on the map monitor the short-circuit current of the two devices, whereas maximum-power-point power or current values are available for PV systems. Short-circuit current is known to be less affected by nonuniform soiling than the maximum-power-point data [15]; therefore, the soiling stations data reported on the map are expected to be less impacted by nonuniform soiling than PV systems.

B. Environmental Data

In our previous works, we showed that particulate matter concentrations and the length of dry periods were the best predictors of soiling measured at soiling stations in the United States [7], [12]. In the present analysis, particulate matter data have been sourced from the database of the U.S. Environmental Protection Agency (EPA) [16]. The mean annual concentrations recorded by the EPA monitoring stations closest to each soiling site have been averaged to get a mean value for each of the regions considered in this work (Table II). Generally, PM_{10} concentrations are recorded daily or every 6th day, whereas $PM_{2.5}$ concentrations are measured daily: in this work, we only accepted data recorded from monitoring stations that took at least 75% of the measurements scheduled in a year [17].

Site-specific daily precipitation data were downloaded from the University of Oregon's PRISM database [18]. The length of the dry periods of every region is obtained as the arithmetic mean of the length of the dry periods of each site. Rainfall data in PRISM are available for the continental U.S. only.

C. Spatial Interpolation Methods

We investigated the ability of spatial regression techniques to predict soiling losses at a site given soiling data collected in nearby locations by using the following procedure:

1. Extract from the soiling map only sites within a certain region.

2. Divide the selected sites randomly between a training (TrD) dataset and a test (TeD) dataset with the same number of data points.
3. Use the TrD soiling ratios to estimate the TeD's soiling ratios through the different spatial-interpolation techniques listed below.
4. Calculate R^2, root-mean-square error (RMSE), and normalized RMSE ($RMSE_n$) between the TeD soiling ratios measured at each site and those estimated by interpolation. $RMSE_n$ is calculated as the ratio of RMSE to the difference between the maximum and minimum soiling ratios among the sites selected in Step 1. To avoid any bias due to extremely low dataset population, calculate the three metrics only if soiling ratios can be estimated for at least four sites of the TeD.
5. Repeat points 2, 3, and 4 above 1000 times.
6. Calculate the average R^2, RMSE, and $RMSE_n$ from the outputs of the 1000 iterations. Ignore the results if R^2, RMSE, and $RMSE_n$ cannot be calculated for at least 50% of the iterations.

This approach is defined as "Random Sampling," because all the sites within a region are used in the analysis. In a different approach, named "Selective Sampling," only sites with certain features are used. This method, which adds a further site-selection step between Point 1 and Point 2 in the procedure, is of interest because, even if located nearby, soiling sites might be affected by dissimilar soiling because of different geometries or conditions. In this case, the $RMSE_n$ is calculated as the ratio of RMSE to the difference between the maximum and minimum soiling ratios among the sites remaining after the selective sampling.

We considered five spatial-interpolation methods, commonly used in air-quality studies [19]:

- Nearest neighbor (NN): each TeD site is assigned the soiling loss of the closest TrD site.
- Spatial averaging (SA): each TeD site is assigned a soiling ratio equal to the arithmetical mean of those of the TrD sites located within a set distance.
- Inverse distance weighting (ID): the soiling ratio of a TeD site is obtained as the mean of those of the TrD sites located within a set distance, and by using the inverse of the distance between the TeD and each TrD sites as a weighting factor.
- Inverse square distance weighting (ID2): the soiling ratio of a TeD site is obtained as the mean of those of the TrD sites located within a set distance, and by using the inverse of the squared distance between the TeD and each TrD sites as a weighting factor.
- Declustered distance estimation (DDE): the soiling ratio of a TeD site is obtained as the mean of those of the TrD sites located within a set distance, and by using as a weighting factor a parameter calculated considering both the inverse of the distance between the TeD and each TrD site, as well as the spatial distribution of the TrD sites, to avoid overweighting clustered data points [20].

Distances of 50 km and 250 km were considered in this study. Surface soiling maps of California, obtained by plotting the soiling ratio distribution with a resolution of 0.05°×0.05° using the different techniques, are shown in Fig. 2.

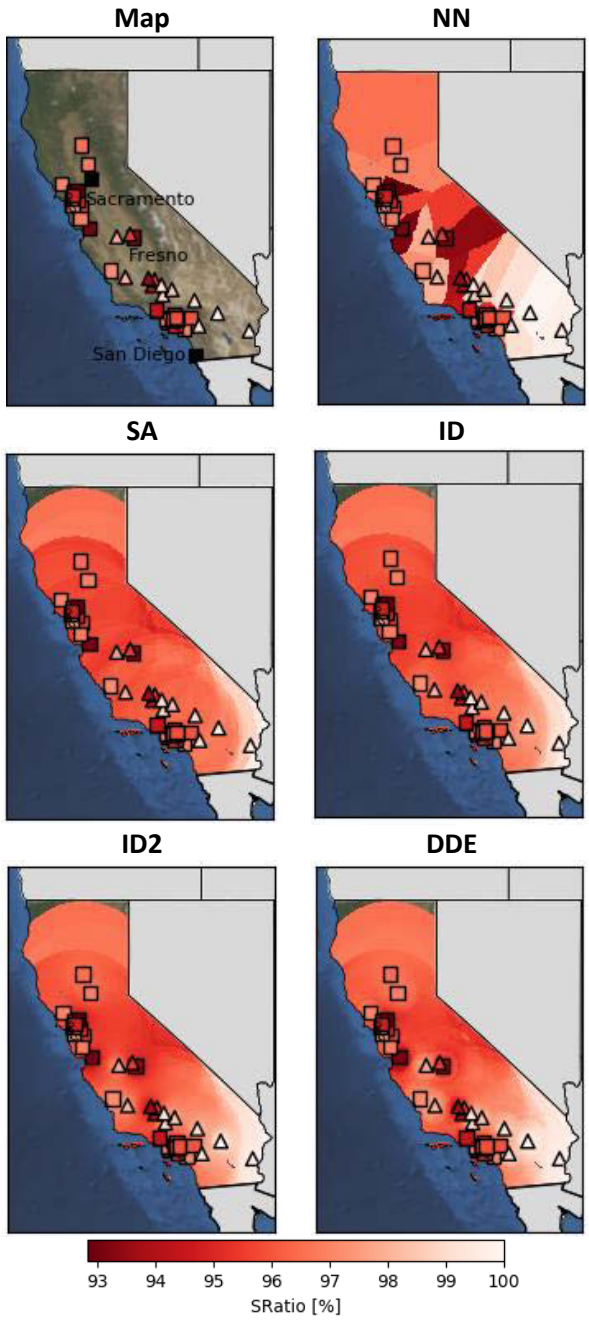

Fig. 2. Soiling surface maps of California obtained by using the different spatial interpolation techniques considered in this study. Squared makers (■) represent PV systems, triangular markers (▲) represent soiling stations. For SA, ID, ID2, and DDE, we considered a radius of 250 km. The pixel size is 0.05°×0.05°, which corresponds to vertical dimensions of 5.5 km and horizontal dimensions ranging between 4.2 km (at northern latitudes) and 5.3 km (at southern latitudes).

III. Soiling Map and Regional Analysis

A. Soiling Map

The current version of the soiling map, published in October 2017, contains 83 data points, measured from 41 soiling stations and 42 PV systems [9]. It will be updated as new locations are shared with NREL. The following information is currently reported for each site (see Table I for references):

- Years of operation,
- County where the site is located,
- Soiling Ratio: insolation-weighted mean of the daily soiling ratios,
- Uncertainty on the calculation of the soiling ratio,
- Soiling Rate: median of the daily rates of change in soiling ratio during dry periods, and
- Soiling Rate range: the 97.5th and the 2.5th percentiles of the soiling rate distribution.

The soiling ratios shown in the map are obtained as insolation-weighted averages of the daily mean values. This approach was chosen because insolation-weighted values reflect the impact of soiling on the energy yield, which, we believe, is the main interest of those who will use the map. The distribution of soiling experienced in multi-inverter systems will be addressed in future versions of the map: in the current version, only the value of one inverter is presented, even for sites in which multiple inverters are installed. No information is available on the cleaning schedule of the PV systems.

B. Soiling Regions

All the locations shown on the map were divided into six regions (Table II). Four of them mirror the aerosol regions described by the EPA [21], and the last two regions group together the data points collected either in Hawaii or in the U.S. Virgin Islands. New regions can be added, existing ones can be amended, and the state lists can be updated as new data points become available.

The results of the analysis are shown in Fig. 3. The Southwestern (SW) states are those with the most soiling, with

TABLE II
SOILING REGIONS.

Region	States	Sites
Hawaii (HI)	Hawaii	5
Northeast (NE)	Connecticut, New Jersey	3
Southeast (SE)	Florida, Georgia, Maryland, North Carolina	4
Southern California (SC)	California (selected counties) *	25
Southwest (SW)	Arizona, California,** Colorado, Nevada, New Mexico, Texas	44
Virgin Islands (VI)	U.S. Virgin Islands	2

* Counties: Imperial, Kern, Los Angeles, Orange, Riverside, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, Ventura.
** Only counties not included in the SC region.

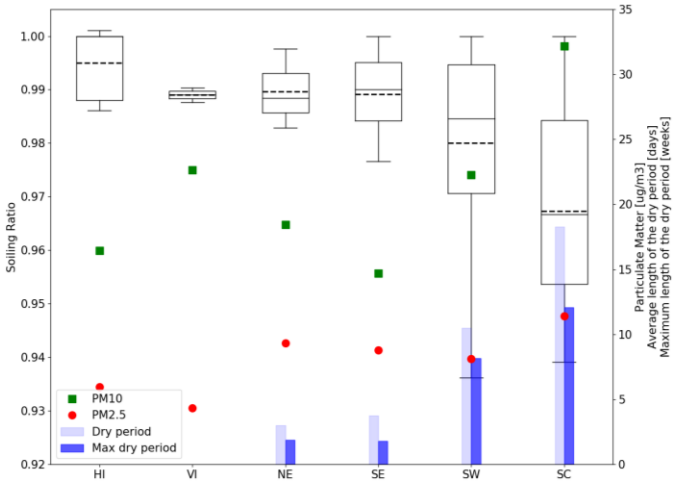

Fig. 3. Average soiling ratios (boxplot, left y-axis), particulate matter concentrations, and dry-period lengths (right y-axis) for the six U.S. soiling regions considered in this study. Each boxplot shows minimum, first quartile, median (bold continuous line), mean (grey broken line), third quartile, and maximum soiling ratios. PM_{10} (green square markers) and $PM_{2.5}$ (red round markers) are obtained as the mean of the values recorded by the EPA monitoring station closest to each site in the region. Dry-period data (light blue bars) and maximum dry-period data (dark blue bars) are from PRISM [18]; rain data for Hawaii and U.S. Virgin Islands are not available.

Southern California (SC) having the highest soiling losses (i.e., minimum soiling ratios). This is not surprising because this region has the highest PM_{10} and $PM_{2.5}$ concentrations (32 $\mu g/m^3$ and 11 $\mu g/m^3$, respectively) as well as the longest dry periods. However, the particulate matter does not follow the same trend as soiling in other regions. A clearer correlation is obtained instead if we consider the average or maximum length of the dry periods. The average length is three and five times higher in the SW and SC, respectively, than the average values on the East coast. Similarly, the maximum length in the western regions is four and seven times higher, respectively, than the average values on the East coast.

IV. Spatial Interpolation

A. Random Sampling

Soiling occurring at a site might experience strong inter-annual variations [22], and the impact of soiling on different inverters of the same system has yet to be quantified. For these reasons, the analysis presented in this section does not include PV system data, which are recorded over long and different periods of time (ranging between 2001 and 2016) and do not consider potential nonuniformities occurring among inverters at the same sites. This investigation considers only the soiling station data available on the soiling map, which have all been recorded between 2013 and 2016 using short-circuit current measurements. An analysis, inclusive of PV system data, will be presented in future works and will include new sites as they become available. In addition, we considered only the two regions with most data points: Southern California, and

Southwest (inclusive of Southern California counties).

A random sampling spatial interpolation is first attempted. The mean R^2, obtained as the arithmetic averages of R^2 calculated for each of the 1000 iterations, are shown in Table III. The results show the existence of a correlation between the interpolated and measured data, with R^2 achieving maximum values higher than 70%. These correspond to RMSE in the range of 1.3% to 2.3% and to a minimum normalized RMSE of 21.2%. The best results are obtained if a spatial-interpolation technique is employed, instead of considering the soiling of the nearest site only, independently of its distance. Moreover, a smaller radius is found to return better results in the Southwest region: the closer the sites considered for spatial interpolation, the higher the correlations. Unfortunately, no valid results can be shown for Southern California for a distance of 50 km because most of the iterations did not meet the requirement set on Point 4 of the procedure detailed in Section II.C.

To further understand the impact of the distance between a soiling data point and the site of interest, the NN analysis is repeated considering as valid only TrD locations within a set distance of the TeD site. In the Southwest, the results show that considering NN only if the data point is within 50 km of the site instead independently of its distance raises the R^2 from 42% to 63%. No significant difference is found if we consider a maximum distance of 250 km, confirming that only shorter distances should be considered to increase the quality of the estimation.

TABLE III

MEAN R^2, IN %, OBTAINED FOR DIFFERENT SPATIAL INTERPOLATION METHODS IN THE VARIOUS REGIONS AFTER 1000 ITERATIONS. R^2 ARE NOT REPORTED IF SOILING RATIOS OF FOUR OR MORE SITES OF THE TEST DATASET COULD NOT BE CALCULATED IN AT LEAST 50% OF THE ITERATIONS.

Method	NN	SA		ID		ID2		DDE	
		50 km	250 km	50 km	250 km	50 km	250 km	50 km	250 km
SC	37		27		44		46		46
SW	42	74	41	74	47	73	48	74	49

B. Selective Sampling

Soiling is known to depend on the characteristics of the site and can vary with the geometry of the system. A random sampling approach does not consider these factors, because it only accounts for the distance between a site and the available data points. In this light, selective sampling has the potential to increase the results of spatial interpolation because it will remove data points with the features different than those of the site of interest. The features we considered in this study for the site selection are the following:

1) Tracking configuration: fixed tilt or single-axis tracked.
2) Mounting type: roof- or ground-mounted.
3) Land cover: developed or non-developed. The categories have been selected by using the USDA soil survey [23]. In the presence of a mixture of constructed materials and vegetation, we considered any site where impervious surfaces accounted for less than 20% of the total cover as "Non-Developed."

The results are shown in Fig. 4: we considered only the parameters that returned dataset counting at least the half of

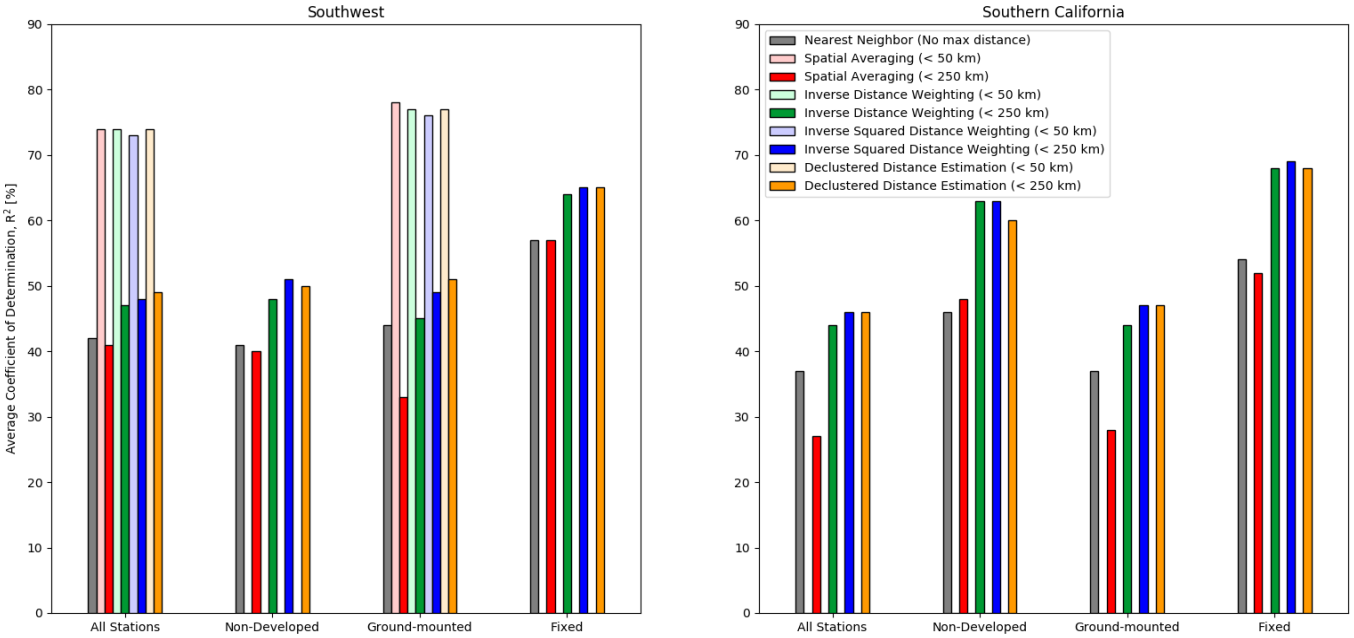

Fig. 4. Comparing the results of random sampling ("All Stations") with the results of selective sampling in two regions. Only data respecting the requirement of Step 6 of the procedure are shown. Each spatial interpolation method is reported using a single color: data obtained for radii of 50 km are reported in lighter shades as they are not available for all the selections and the regions.

the initial site number for each region. The best results are still obtained for shorter distances (lighter colors); but more data points are needed to confirm this trend because results valid according to Step 6 of the procedure described in Section II.C are obtained for only two categories in the Southwest. Overall, maximum R^2 values between 78% and 76% (RMSE between 1.4% and 1.6%) are achieved if ground-mounted stations are considered.

For larger distances, the site selection shows the ability to improve the R^2 of the correlations by up to 25%. The greatest enhancements are obtained if stations located in "Non-Developed" sites or fixed stations are considered. Fixed soiling stations return the most consistent improvements, probably because most of the fixed stations have similar tilt angles, ranging between 20° and 25°. Table IV shows how the minimum RMSE and $RMSE_n$ values can be lowered in all the regions if we consider a selective approach.

The results of this first investigation confirm that the map can be a useful tool for predicting soiling losses at sites where no soiling data are available. However, more data points are required to draw more generalized conclusions and to analyze a larger number of selective parameters.

TABLE IV
MINIMUM RMSE AND NORMALIZED RMSE, IN %, OBTAINED FOR THE TWO REGIONS CONSIDERED IN THE ANALYSIS DEPENDING ON THE SELECTED SAMPLING APPROACH.

Region	SC		SW	
Sampling	Random	Selective	Random	Selective
RMSE	2.0	1.7	1.4	1.1
$RMSE_n$	30.7	28.1	22.3	20.1

IV. CONCLUSIONS

In this work, we presented a new soiling map that can help the community address the soiling of photovoltaic modules. The map collects soiling data from 83 sites, either soiling stations or PV systems, analyzed using referenced techniques; it will be updated with new locations as they become available.

The data on the map have been used to conduct the first regional analysis of soiling across the United States. The western states, which experience the longest dry periods, are those affected by the highest soiling losses, with the sites located in southern California having the maximum losses.

The soiling map can be used to estimate losses at a site given soiling data from nearby locations. As proof, we employed spatial-interpolation techniques to estimate soiling losses in the most soiled regions of the United States. We found coefficients of determination as high as 74% when soiling data at randomly sampled locations are replaced with those measured at the closest-available locations. This means that the average soiling ratio at a site can be estimated with RMSE as low as 1.4%. In particular, the best results were obtained if only locations within 50 km of the investigated sites were considered: the addition of new data will make it possible to estimate with better accuracy soiling that occurs at sites where no data are available.

The outcomes of the analysis suggest that the coefficients of determination can be enhanced up to 78% and the RMSE lowered to 1.1% if selective sampling is performed to only consider data points with features like those of the investigated sites. Selective sampling requires a larger dataset and more information on the geometry of the systems, and the conditions of the sites. The analysis will be repeated when new data become available, including, in addition, data from PV systems and additional selective parameters.

ACKNOWLEDGMENTS

This work was authored by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract 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] S. C. S. Costa, A. S. A. Diniz, and L. L. Kazmerski, "Solar energy dust and soiling R & D progress : Literature review update for 2016," *Renew. Sustain. Energy Rev.*, vol. 82, no. April 2017, pp. 2504–2536, 2018.

[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*, 2006, pp. 2391–2395.

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

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

[5] T. Sarver, A. Al-Qaraghuli, and L. L. Kazmerski, "A comprehensive review of the impact of dust on the use of solar energy: History, investigations, results, literature, and mitigation approaches," *Renew. Sustain. Energy Rev.*, vol. 22, pp. 698–733, 2013.

[6] S. C. S. Costa, A. S. A. C. Diniz, and L. L. Kazmerski, "Dust

and soiling issues and impacts relating to solar energy systems : Literature review update for 2012 – 2015," *Renew. Sustain. Energy Rev.*, vol. 63, pp. 33–61, 2016.

[7] L. Micheli and M. Muller, "An investigation of the key parameters for predicting PV soiling losses," *Prog. Photovoltaics Res. Appl.*, vol. 25, no. 4, pp. 291–307, 2017.

[8] M. G. Deceglie, L. Micheli, and M. Muller, "Quantifying Soiling Loss Directly from PV Yield," *IEEE J. Photovoltaics*, 2018.

[9] National Renewable Energy Laboratory, "Photovoltaic modules soiling map," 2018. [Online]. Available: https://www.nrel.gov/pv/soiling.html. [Accessed: 18-May-2018].

[10] IIT Bombay, "SERIIUS Site Soiling Rate of the World." [Online]. Available: http://www.ncpre.iitb.ac.in/pages/SERIIUS_Soiling_rate_of_the_World.html. [Accessed: 21-Nov-2016].

[11] M. G. Deceglie, M. Muller, Z. Defreitas, and S. Kurtz, "A Scalable Method for Extracting Soiling Rates from PV Production Data," in *2016 IEEE 43rd Photovoltaic Specialist Conference (PVSC)*, 2016.

[12] L. Micheli, M. Muller, and S. Kurtz, "Determining the effects of environment and atmospheric parameters on PV field performance," in *2016 IEEE 43rd Photovoltaic Specialist Conference (PVSC)*, 2016, vol. 2016-Novem, pp. 1724 – 1729.

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

[14] International Electrotechnical Commission, "Photovoltaic system performance – Part 1: Monitoring (IEC 61724-1, Edition 1.0, 2017-03)." IEC, Geneva, Switzerland, 2017.

[15] M. Gostein, T. Duster, and C. Thuman, "Accurately Measuring PV Soiling Losses With Soiling Station Employing Module Power Measurements," in *IEEE 42nd Photovoltaic Specialist Conference (PVSC)*, 2015.

[16] US Environmental Protection Agency, "Air Quality System Data Mart [internet database]." [Online]. Available: https://www.epa.gov/airdata. [Accessed: 15-May-2018].

[17] U.S. Environmental Protection Agency, "About Air Data Reports." [Online]. Available: https://www.epa.gov/outdoor-air-quality-data/about-air-data-reports. [Accessed: 01-May-2018].

[18] PRISM Climate Group - Oregon State University, "PRISM Climate Group - Oregon State University." [Online]. Available: http://www.prism.oregonstate.edu/explorer/. [Accessed: 03-Jun-2017].

[19] D. W. Wong, L. Yuan, and S. A. Perlin, "Comparison of spatial interpolation methods for the estimation of air quality data," *J. Expo. Anal. Environ. Epidemiol.*, vol. 14, no. 5, pp. 404–415, 2004.

[20] S. Falke, R. Husar, and B. Schichtel, "Mapping Air Pollutant Concentrations from Point Monitoring Data I: Declustering and Temporal Variance." [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.612.5742&rep=rep1&type=pdf. [Accessed: 18-May-2018].

[21] U.S. Environmental Protection Agency, "Air Quality Criteria for Particulate Matter (Final Report, 2004)," Washington, DC, 2004.

[22] M. G. Deceglie, L. Micheli, and M. Muller, "Quantifying Year-to-Year Variations in Solar Panel Soiling from PV Energy-Production Data," in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*, 2017.

[23] U. S. D. of A. Soil Survey Staff, Natural Resources Conservation Service, "Web Soil Survey." [Online]. Available: http://websoilsurvey.nrcs.usda.gov/. [Accessed: 18-May-2018].

[24] M. Muller, L. Micheli, and A. A. Martinez-Morales, "A Method to Extract Soiling Loss Data From Soiling Stations with Imperfect Cleaning Schedules," in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*, 2017.

[25] National Renewable Energy Laboratory, "National Solar Radiation Data Base (NSRDB)." [Online]. Available: https://nsrdb.nrel.gov/. [Accessed: 18-May-2018].