An Investigation of the Key Parameters for Predicting PV Soiling Losses

Leonardo Micheli* and Matthew Muller

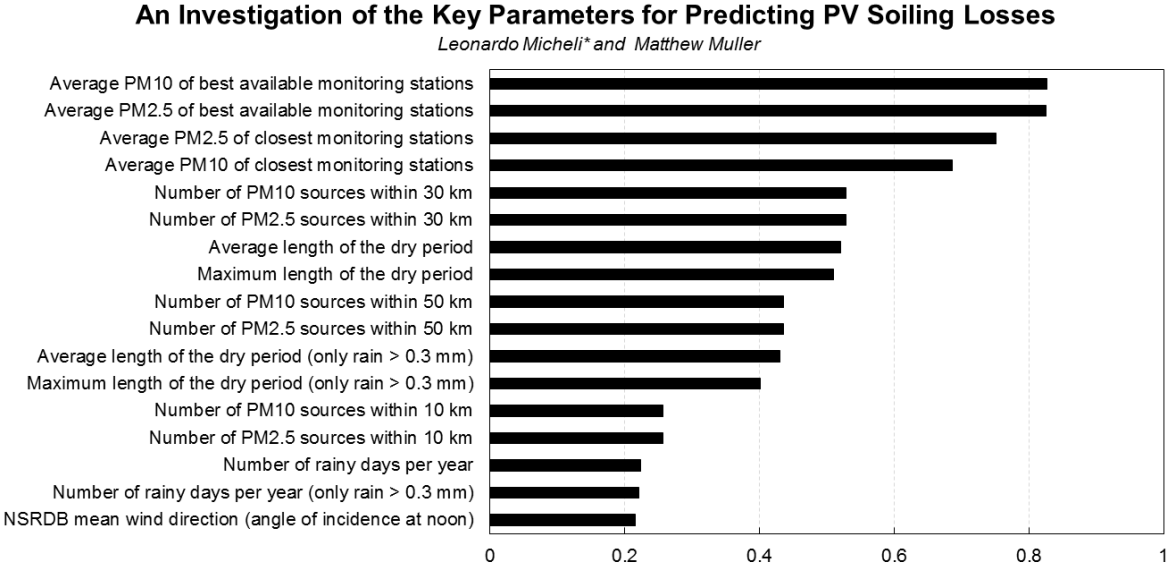
National Renewable Energy Laboratory, Golden CO, 80401

*Corresponding author: Leonardo.Micheli@nrel.gov

Abstract

One hundred and two environmental and meteorological parameters have been investigated and compared with the performance of twenty soiling stations installed in the USA, in order to determine their ability to predict the soiling losses occurring on PV systems. The results of this investigation showed that the annual average of the daily mean particulate matter values recorded by monitoring stations deployed near the PV systems are the best soiling predictors, with coefficients of determination (R^2) as high as 0.82. The precipitation pattern was also found to be relevant: among the different meteorological parameters, the average length of dry periods had the best correlation with the soiling ratio. A preliminary investigation of two-variable regressions was attempted, and resulted in an adjusted R^2 of 0.90 when a combination of $PM_{2.5}$ and a binary classification for the average length of the dry period was introduced.

Graphical table of contents



Coefficients of determination of the linear single-variable correlations between the most significant soiling predictors considered in this study and the soiling ratios registered by twenty PV systems deployed in the USA.

Keywords: soiling, photovoltaic performance, soiling losses, particulate matter, precipitation, linear regression.

Introduction

It is well known that PV performance is impacted by the accumulation of dust on the surface of PV panels, commonly referred to as “PV soiling losses”. PV soiling losses have been studied at various locations around the world; results are typically site-specific and related to mounting angle and other factors. Reported annual losses for the United States have ranged from 0 to 6% [1–4], while annual losses are not readily available for some of the dustiest regions of the world, like the Middle East, India or China. In these locations, peak losses have been reported ranging between 20 and 70% [5–10], but these values do not directly translate to annual losses, which are typically much lower. Annual PV soiling losses are generally the result of considering the soiling rate (typically the increase in loss per day) for a site in combination with rainfall patterns or other cleaning events. Kimber *et al.* [1] proposed one method for combining site soiling rates with rainfall patterns to derive time-series soiling losses and annual soiling losses.

Ideally, if the soiling rate data and rainfall data were available for a new PV site, the data could be used to predict soiling derates for the site that could be directly incorporated into existing PV performance models. Rainfall data are typically available for most locations around the world, but soiling rate data are only sparsely available based on examples from the literature. IIT Bombay has attempted to gather all such rates available in the literature, and has compiled them on a world map [11]. Although this map is a great resource, the soiling loss rates were each measured in a unique way, and therefore cannot be used to directly calculate PV energy losses. It is desirable to move beyond site-specific soiling loss studies and to have a model or other technique to determine expected soiling losses at a given location without taking new measurements at that location. If a model is developed that can predict variation in PV soiling based on site characteristics and system architecture soiling can be included in site cost evaluation and selection. Engineers can better design the system to minimize soiling and operation and maintenance plans can be improved based on expected soiling losses and system economics. Financiers can also provide lower interest rates as soiling derates become data driven rather than generalized.

In this paper, the performance of PV soiling stations at 20 diverse sites in the USA are compared with existing meteorological data, pollution indexes, and land characteristics made available from national databases. The purpose is to determine if the differences in soiling losses at each site can be predicted using meteorological and other data. The present work extends the analysis reported in [12], by increasing the number of soiling stations considered, and by presenting a more in-depth analysis of each meteorological

parameter. Moreover, compared to the previous publication, more accurate particulate matter data are considered, and a detailed analysis of the land characteristics is introduced.

Data collection

Soiling stations

Data from 20 soiling stations installed in the USA have been investigated in this work. The stations have been operating in different time periods, ranging from 7 to 40 months, between 2013 and 2016: the data collection period of each station and the characteristics of their different installation locations are listed in Table 1. The stations have different structures and geometries, but they are all at the minimum composed of two reference cells (or modules) and a pyranometer. Each device on these stations is regularly cleaned, with the exception of one of the two cells. The two devices are here referred to as “cleaned cell” and “soiled cell,” regardless of whether the soiling station makes use of reference cells or full modules. The short circuit current output of each cell is recorded along with different weather conditions, depending on the soiling station. Five stations have a single axis tracker, and the remaining ones are mounted at a fixed tilt, facing south.

The soiling stations investigated in this work are located in eight different states of the USA, four of which have a coastline, including the Hawaiian island of Oahu. Several climate zones are represented in this study: each location has been characterized by a number of meteorological and environmental parameters. Each parameter has been identified as having potential impact on PV soiling losses. This identification is based either on the PV soiling analyses presented in the literature, or on a brainstorming effort by experts in the PV soiling community (The International PV Quality Assurance Task Force [PVQAT] Task Group 12) [13]. It is outside the scope of this work to provide a complete literature review of all of these parameters; an abbreviated list that warranted more discussion is included here, whereas the complete list, inclusive of description, ranges, and sources, is reported in the Appendix. Non-numeric parameters have been converted into both binary and numerical classifications to make them statistically analyzable. These new classifications have been based on the potential impact on soiling: higher binary and numerical values have been assigned to those categories that have potential for increased PV soiling.

Soiling metrics

The soiling losses at the 20 sites have been quantified using the short-circuit currents from both clean and dirty PV devices. Short-circuit current has been often used in literature and is still considered a good electrical parameter to estimate power loss in case of uniform soiling [50], which is the scope of the present investigation. All the PV

performance and weather data have been converted into and analyzed as daily values. The quality of each dataset has been independently checked, and data recorded by the soiling stations have all been processed in the same way. Each daily average current (Isc(i)) considered in this work has been calculated as:

$$Isc(i) = \frac{\sum_{h=j}^{k} Isc_h(h) \cdot \frac{1000\ W/m^2}{POA(h)}}{n} \quad (1)$$

where $Isc_h(h)$ is the mean hourly short-circuit current measured per each cell or module, POA is the mean hourly plane of array irradiance, h is the hour, j and k are the hours when the first and last current measurements are taken, and n is the number of hours used for the daily average. Where not available, the POA has been calculated using the hourly values of global horizontal irradiance and angle of incidence, by estimating the direct, the ground-reflected diffuse and the sky-diffuse components of global horizontal irradiance [47–49].

The same filtering criteria applied in the previous work [12] have been used in the present study. Data have been collected as 1-minute data and averaged into hourly values. Only data recorded between 11:00 AM and 1:00 PM and under clear sky conditions (irradiance ≥ 500 W/m^2) have been then considered for the calculation of the daily values, in order to remove data influenced by shading (which occurred in some stations). Although the soiling ratios registered during the central hours of the days are expected to be 1% higher than the daily averaged values [18], all sites are filtered in the same way, and therefore this bias should not impact the relative comparison between the sites investigated in this study. Moreover, those hours in which the normalized short circuit current of the cleaned cell was found to be lower than the 80% of its expected value have been removed.

The daily soiling ratio (daily SRatio) measures the ratio between the normalized short-circuit currents of soiled and cleaned PV cells for each *i*th-day, as follows:

$$Daily\ SRatio(i) = \frac{Isc_{soiled}(i)}{Isc_{cleaned}(i)} \quad (2)$$

where Isc_{soiled} and $Isc_{cleaned}$ are the irradiance-corrected average daily short-circuit currents of the dirty and clean cells, respectively. The daily SRatio is always expected to be less than or equal to 1. The closer the daily SRatio is to 1, the lower the soiling losses registered at the site.

In order to compare the site soiling losses with the parameters listed in the Appendix, the average of the daily soiling ratios recorded at each site during the data collection period (SRatio) has been calculated as follows:

$$SRatio(site) = \frac{\sum_{i=1}^{n_{days}} daily\ SRatio(i)}{n_{days}} \quad (3)$$

where n_{days} is the number of days the soiling station has been operating. Note that the SRatio(site) is indicative of annual soiling losses at a given site, but it should not be interpreted as an annual energy loss, as equation (3) does not weight the daily SRatio by the irradiation available each day.

A second soiling parameter that has been included in the investigation is the soiling rate, which describes the rate at which the daily soiling ratio varies during non-rainy periods at each site. It has been determined using the method proposed by Deceglie *et al.* [51], and is the median of the slopes of the daily soiling ratio profile for any dry period longer than 14 days. The slopes have been calculated using the Theil-Sen method.

Meteorological parameters

There are no definitive conclusions in the literature as to which measures of precipitation have the highest correlation with PV soiling losses. For the six sites investigated in the previous work [12], it was found that the average number of days between rain events had the best correlation with the soiling losses. As this was a preliminary investigation, it was considered valuable to continue to take into account a range of metrics for capturing the precipitation statistics of a site. Therefore the precipitation profile of each site has been described through a larger number of parameters (listed in the Appendix). Hourly rainfall has been measured directly at most of the 20 soiling stations, but the precipitation statistics are derived from PRISM datasets [20–22] to treat each site consistently, and also to avoid using rain data from a small number of sites that posed quality control issues. Moreover, by comparing the soiling ratio behavior and the rainfall pattern at some sites, it was found that no cleaning effect was recorded on the soiled cell after several rain events of 0.3 mm or less. For this reason, a second set of precipitation parameters has been calculated: daily accumulated precipitation less than or equal to 0.3 mm is not considered a rain event in this set.

Wind speed, ambient temperature and relative humidity statistics have been determined directly from onsite measurements, with the exception of a few sites that did not include relative humidity data. Where not available, the relative humidity data were determined from the National Solar Radiation Database (NSRDB) [23–25]. The same database has been used to source a second set of wind data. The mean wind direction has been

calculated using the methodology reported in Ref. [26]. In order to be able to analyze the impact of wind direction, along with the standard wind direction convention (0° to 360°, where 0° represents calm wind and 360° wind blowing from north), a variable consisting of the absolute value of the difference between the mean wind direction and the azimuth orientation of the soiling stations at noon (south) has been introduced. This variable is equal to 0° if the mean wind direction is south, to 90° if either west or east, and to 180° if north.

Pollution

The US Environmental Protection Agency (EPA) defines particulate matter (PM) as a mixture of solid particles and liquid droplets suspended in air [27]. PM_{10} and $PM_{2.5}$ represent respectively the concentrations, in 1 m^3 of air, of airborne particulate matter less than 10 microns and less than 2.5 microns in diameter. Therefore, $PM_{2.5}$ is a subset of PM_{10}. These smaller particles can typically remain airborne for longer times and travel for longer distances than the larger particles included in the PM_{10} because of their lower weight [28]. The effects of particulate matter on the performance of PV have already been described in literature [29–31]. A linear correlation between PM_{10} and the soiling accumulation rate (in grams of particulate accumulated on 1.0 m^2 of PV cover plate per day) was first presented in Ref. [32], where two sites in Colorado were investigated. In the previous communication [12], it was found that, among the investigated parameters, PM_{10} and $PM_{2.5}$ had better correlations with soiling losses than any other parameter. For this work, the particulate matter concentrations of each location have been directly sourced from the US EPA databases [33], taking into account only the time period when each PV system was operating, and then processed according to the methodology presented in Ref. [34].

The US EPA database reports the annual average of the daily mean values of PM_{10} and $PM_{2.5}$, registered for each monitoring station deployed in the United States from 1990 through 2015 [33]. For this reason, PM data for soiling stations active in 2016 have not been included. Only data from monitoring stations located within a determined distance from the investigated soiling stations have been used for this analysis: distances of 10 km, 30 km, 50 km, and 100 km have been considered. Data influenced by unusual or natural non-controllable events, marked as “Exceptional events” by the EPA, have been included, because these events are also expected to impact the soiling of PV. The particulate matter data have been processed using the procedures described in [34], where different statistical methods for mapping particulate matter data are described. In the present work, three methods are considered:

- **Mean with no weighting (Arithm. Mean)**: data are obtained as arithmetic mean of all monitoring stations located within a set distance from the PV soiling station.

- **Distance-weighted mean (DW Mean)**: monitoring stations closest to the soiling station are weighted more heavily than those at greater distances from the soiling station.
- **Declustered-configuration mean (Decl. Mean)**: data are obtained by taking into account the distance between the monitoring and the PV soiling stations, as well as the average distances among the monitoring sites.

Declustering is a method that is mathematically designed to mitigate bias that occurs when multiple stations are located in close proximity to each other, and other stations are separated by greater distances (within the larger area under consideration). Indeed, in the case where a PV site is on the edge of an urban area, it is likely that there will be a larger number of stations clustered in the urban area and a smaller number near the site. The simple mean and the distance-weighted mean will give equal weight to all those stations that are clustered together. The declustering method will instead reduce the weight of spatially-clustered sites (based on distance between the clustered stations), so that together they have approximately the weight of one station at the given distance from the PV site.

Beyond considering different distances and weighting methods for PM stations, alternate indicators of airborne pollutants were examined, such as the 2014 satellite-derived $PM_{2.5}$ levels [35,36], the total number of $PM_{2.5}$ and PM_{10} emission sources located within a set distance of each site, and the estimated total amount of PM emitted by those sources [37]. A complete description of all parameters considered is reported in Appendix.

Land cover

The characteristics of the soil surface of each site have been determined by using the land survey made available from the United States Department of Agriculture (USDA) [38]. The USDA classifies the land cover in different categories: the ground cover at each of the 20 sites, shown in Table 2, has been obtained by using the exact coordinates of each soiling station, and then confirmed through a visual inspection of the satellite imagery of the area surrounding the soiling station. Categories have then been grouped into both numerical and binary classes.

Wildfire Hazard Potential

Wildfire emissions have an impact on air quality, and regions that are prone to wildfires are also often more arid; both have a potential relationship with soiling on PV panels. According to the EPA [39], between 2002 and 2014, wildfires were responsible for more than 16% and 5% of the total $PM_{2.5}$ and PM_{10} emissions in the US, respectively.

The wildfire hazard potential of each site has been identified through a map developed by the US Forest Service [40]. The fire regime is described by two categories: severity and frequency of wildfires. The sites considered in this work can be grouped in four

levels of risk, summarized in Table 3, which have been converted into four numeric categories (0 to 3, increasing with wildfire risk and severity), and into two binary categories (0 for non-frequency and 1 for frequent). The site in Hawaii has been included in the “35 to 100+ years; Mixed Severity” regime.

Presence of roadways

It is known that the level of pollutants is enhanced in the proximity of roadways [41]. Even if dependent on other factors, such as the traffic volume, the weather conditions or the season, pollutant concentrations have been found to decrease with the distance from the road, until they decay to surrounding background levels. Analyzing the results of 41 previous studies, Karner *et al.* [42] found that most pollutants reached background levels at about 0.5 km of the roadway, whereas the $PM_{2.5}$ did not reach the background level until beyond 1 km. The EPA assumes a decay to background levels within 180 m downwind from the roadways, and reports higher transport distances when low-speed winds are present [43].

Two types of roads have been considered in the present study: highways and unpaved roads. The distances in km between the site and each of these two categories of road have been used for the present study.

Distance from the ocean

The distance of each site from the seashore has been considered as well. Indeed, it is known that the chloride deposition rate on an inland surface decreases with increasing the distance from the sea. Different works [44–46] proposed an inverse exponential relationship between the distance from the sea and the mass deposition flux of sea-salt aerosols. Meira and his colleagues [46] collected experimental data on the Brazilian coast and compared the results with previous works. They reported that most of the reduction in chloride deposition takes place in the first 500 m from the coastline, even if the spatial deposition profile is strongly influenced by different factors, such as the wind speed and the deposition velocity [47,48].

A second major source of chloride aerosols is road salt, used to deice roads in different states (called the sea belt states) [49]. Indeed, depending on the wind conditions, road salt spray has been found as far as 150 m from the roadway [50]. In this case, a binary classification has been considered, to distinguish if road salt is used for the de-icing in the state (1) or not (0).

Results

102 independent variables were considered for predicting the variation in both the SRatio and SRate at the 20 sites (see the Appendix for a complete list and definitions of each). The impact of each parameter on the soiling losses was determined by

calculating the coefficient of determination (R^2). Of the 102 variables considered, only the measurements of PM_{10} and $PM_{2.5}$ levels, the number of nearby $PM_{2.5}$ and PM_{10} sources, and the variables used to represent information about rainfall statistics showed R^2 near or above 0.5. As discussed previously, many different metrics were considered to quantify both PM level and rainfall statistics, because there was no clear indication in the literature of the best way to quantify these parameters for predicting soiling losses. The proposed approach was to let the R^2 determine which metric was best.

In the case of the PM_{10} and $PM_{2.5}$ metrics, this evaluation did not prove so simple. As proposed, $PM_{2.5}$ and PM_{10} measurements were calculated using monitoring stations within 10, 30, 50, and 100 km of the soiling site, and considered a “mean with no weighting”, a “distance-weighted mean”, and a “declustered-configuration mean” (as described in the “Pollution” section). The results of this step showed that none of the 20 sites had PM_{10} stations within 10 km, and only four sites had $PM_{2.5}$ stations within 10 km. At 30 km, there were 11 soiling stations that allowed for PM_{10} and $PM_{2.5}$ calculations. Considering this, a reduced comparison was completed to compare the different metrics using only the 11 sites at 30, 50, and 100 km. The results of this comparison are given in Table 4.

Both 30 km and 50 km distances show a reasonable correlation between the various PM_{10} and $PM_{2.5}$ calculations and the soiling losses, but the correlations drop significantly in some cases for the 100 km distance. In order to test PM correlations for all 20 sites, a combination of highest-correlating measurements from the 11 sites (underlined in Table 4) and the next best alternative were combined to create the “best available monitoring stations”. For $PM_{2.5}$, the arithmetic mean at 30 km performed best, so these values were retained and the values for four additional sites were calculated using the 50 km arithmetic mean, and the values for the final five sites were taken from a single monitoring station measurement that was closest to the PV site. In the case of PM_{10}, the arithmetic mean at 50 km performed best, so these values were retained, as well as four additional sites available for the 100 km arithmetic mean. The values for the final five sites were calculated using a single PM_{10} monitoring station measurement that was closest to the PV site. These measurements together are considered the “average PM of best available monitoring stations” and allow the comparison of all 20 sites while considering particulate matter, alternate indicators of airborne pollution, rainfall, land characteristics and other site parameters. “Best available” should not be interpreted as “most accurate” or “true” for the site; rather, these are estimates for PM at the site considering the proximity of available PM measurement stations.

For all 20 sites, Fig. 1 shows the parameters whose correlations with the Soiling Ratio and/or Soiling Rate have the highest R^2 values. In order to discard the parameters with no statistical significance, any regression where at least one variable had p-value greater than 0.05 has not been shown.

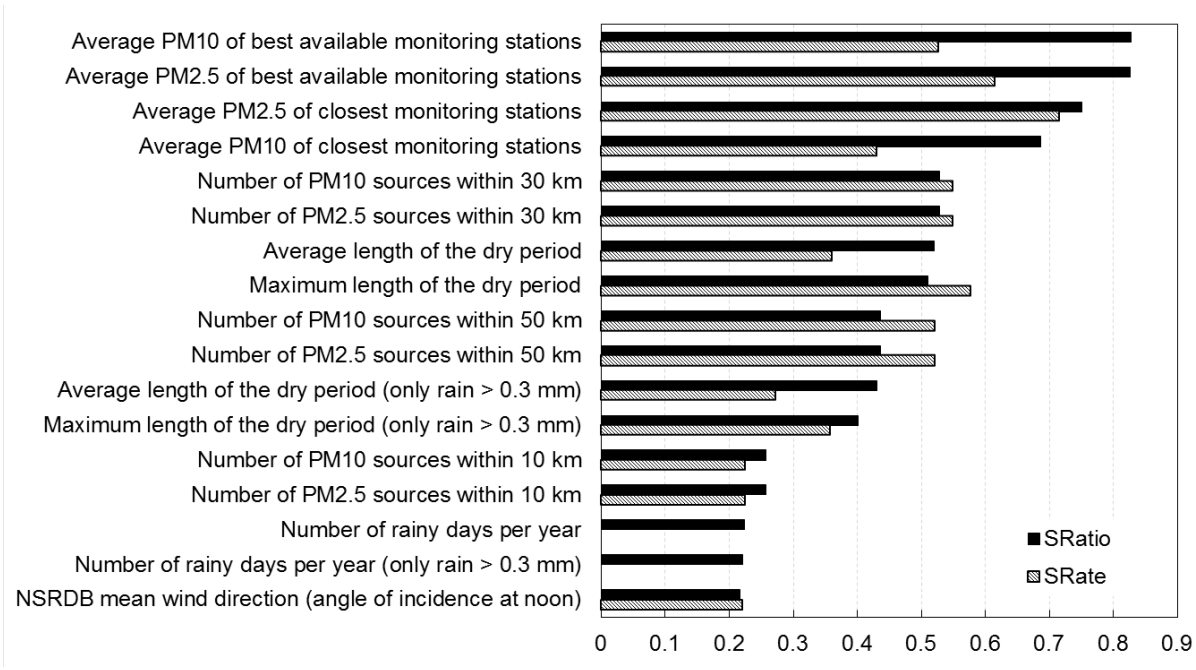


Fig. 1 - Parameters with the highest coefficient of determination when related to soiling ratios and soiling rates

Among all the investigated parameters, only direct measurements of airborne particulates, the number of PM_{10} and $PM_{2.5}$ pollution sources, and some metrics associated with precipitation showed significant correlations with the soiling ratio and the soiling rate. Both the PM_{10} and the $PM_{2.5}$ derived from the best available monitoring station combinations show an R^2 of 0.82 when related to the SRatio. The R^2 drops to about 0.5 for both the average length of the dry period and the maximum length of the dry period. All other predictive variables tested (see the Appendix) showed no significant correlation with either the soiling ratio or soiling rate for the 20 sites under test, with the only exception of the wind angle of incidence at noon, calculated using the NSRDB data [23–25].

The above results are for single-variable regressions, but significance of both particulates and rainfall in predicting soiling ratios and soiling rates suggests considering a two-variable regression. Table 5 therefore provides the "adjusted R^2" ($adjR^2$, which accounts for the bias in adding an additional variable) for running a two-variable multilinear regression with all combinations of the variables that had significance in the single-variable regression. Twenty observations of the independent variable (which corresponds to the number of sites that were investigated) is the statistical minimum for running a two-variable regression, and therefore these results should be considered preliminary, with the expectation that the results could change with the addition of more sites or observations. Only two combinations have been found able to predict the SRatio with accuracies higher than that found for a single-variable

regression. No combination was able to enhance the maximum R^2 registered for the single-variable SRate regressions.

Combinations of the "average PM_{10} of best available monitoring stations" with the "average $PM_{2.5}$ of best available monitoring stations" ($adjR^2$ = 0.88) or the "average PM_{10} of closest monitoring stations" ($adjR^2$ = 0.87) are those able to return an adjusted R^2 higher than the maximum R^2 of 0.82 for a single-variable approach. Interpretation of the R^2 for multiple linear regression results typically assumes that there are no strong cross-correlations between the independent variables. In the case of each of these two variable combinations, this assumption is violated, and therefore it is not clear that any improvement is made with a two-variable approach.

From the results shown in Table 5, it can be seen that a basic two-variable regression did not exceed an adjusted coefficient of determination of 0.73, below the maximum obtained so far with a single-variable regression (0.82). Moreover, a combination between pollution and rainfall parameters would be limited to R^2 values up to 0.68. In an additional attempt to validate the hypothesis that both particulate levels and rainfall data should have significance, the data were examined for clustering patterns associated with rainfall parameters. Plots were created of the soiling ratio against measurements of PM, and then color coding against the different metrics of rainfall. Fig. 2 shows a plot of the soiling ratio versus $PM_{2.5}$, which is then color coded for the length of dry periods. This plot shows that the sites with the longest average dry periods are also some of the sites with the lowest soiling ratios (most impacted by soiling). With this in mind, a binary variable was created that grouped sites based on the average length of the dry period. Sites with an average length of dry period less than 20 days were identified with a 0, and sites with an average dry period length greater than or equal to 20 days were identified with a 1. Two additional two-variable regressions were run pairing this binary variable with $PM_{2.5}$ or PM_{10} from the best available monitoring stations. The run with $PM_{2.5}$ showed that both variables were significant, and the adjusted R^2 was 0.90. Again, this should be considered a preliminary result, but it confirms that both a PM metric and a rainfall metric are significant as predictors of soiling loss ratio.

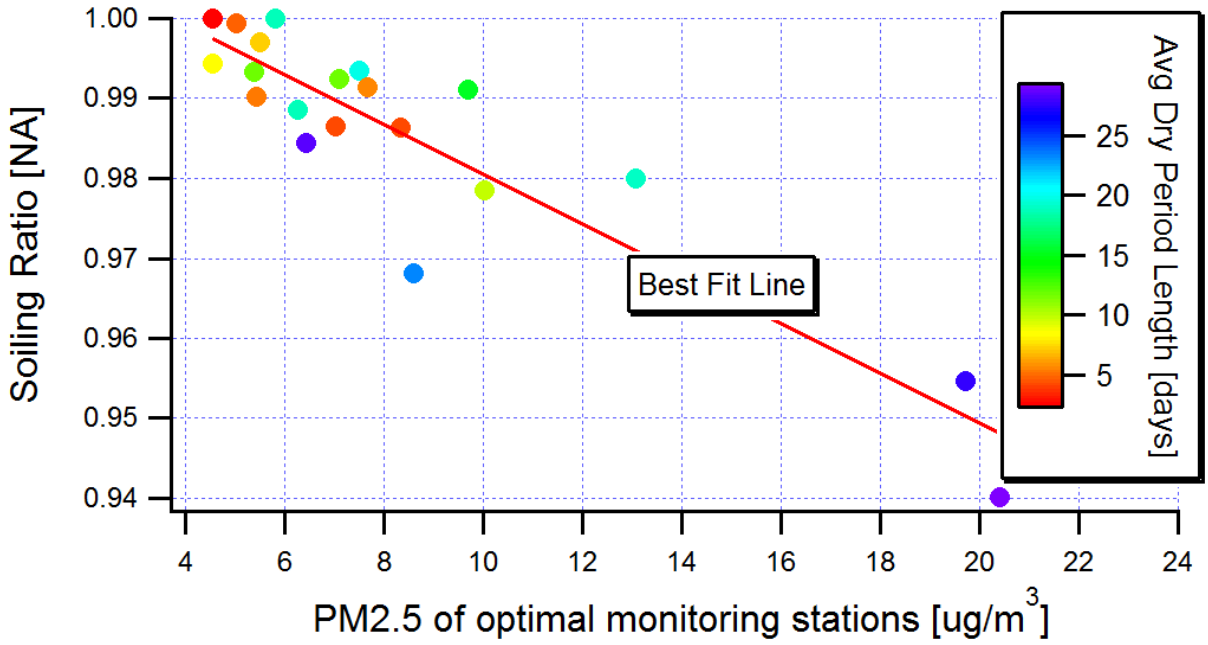


Fig. 2 - Soiling Ratio of each station plotted against the $PM_{2.5}$ obtained by the best available monitoring stations.

Discussion

An R^2 of 0.82 was achieved when using either "average PM_{10} of best available monitoring stations" or "average $PM_{2.5}$ of best available monitoring stations" to predict the site soiling ratio, as shown in Fig. 1. The correlation between particulate matter and the soiling rate can be explained, because particulate matter is generally measured through a gravimetric method, and airborne particles can be expected to deposit onto the PV cells and modules similarly to how they are collected by the monitoring stations. We believe that the R^2 of 0.82 for the 20 sites could be increased if more accurate data were available for PM values within 30 to 50 km of the test site. In other words, both $PM_{2.5}$ and PM_{10} measurements show potential to be highly predictive of the site soiling ratio, but available measurements of PM_{10} and $PM_{2.5}$ are limited for some of the sites. In some cases, the closest monitoring stations are located 80 or more km away from the soiling stations investigated in this study, and the readings of these stations can be expected to be less representative than closer stations.

The results show no clear conclusion as to whether PM_{10} or $PM_{2.5}$ is a better predictor of the soiling ratio. The $PM_{2.5}$ values from the closest monitoring station show a higher correlation with the soiling rate than the PM_{10} values from the closest monitoring station, but because not all sites had stations within the same proximity, caution should be taken in interpreting this. The smaller $PM_{2.5}$ particulates are indeed able to travel further distances before settling from the air, which suggests that for distant stations, $PM_{2.5}$

could be a better choice. On the other hand, Ref. [34] discusses PM transport being blocked by boundaries where there are changes in elevation. For example, if there is a PM source in a valley, the PM does not easily leave the valley. Therefore, only taking distance into account neglects the elevation factor, where stations within short distances of each other can have very different PM levels. Efforts to establish an a-priori method for estimating PV site PM levels (from nearby PM stations) were limited by the fact that only four PV sites had $PM_{2.5}$ measurements within 10 km, eleven had $PM_{2.5}$ and PM_{10} stations within 30 km and in some cases the nearest station was 80 km away. Although an a-priori method was not determined, $PM_{2.5}$ and PM_{10} station data ranging from 10 to 80 km was able to predict 82% of the variation in the average SRatio for the 20 sites in this study. This could be considered a surprising result for several reasons. First, PM values can vary by large amounts over distances of 80 km. Although this is true, this study does not use the distant PM stations to determine absolute PM levels at the PV site but rather to provide a long term relative comparison between the 20 PV sites. In other words PM stations as far 80 km away are able to help distinguish if that one PV site is dirtier than another on long term basis. Second, Ref. [5] showed poor correlation between adjacent measurements of the daily PM_{10} and the daily soiling rate and Ref. [51] showed that on-site PM_{10} measurements were only able to predict 9% of the variation of the mass loading rate on PV glass. Both of these studies were examining variation in PM_{10} and PV soiling metrics over a very short time period. In these same studies it was indicated that these metrics are highly variable and depend on many factors such as wind speed, PM size distributions, relative humidity and a number of other local factors. The results of the present study use instead average PM values over longer periods. The high correlations achieved between nearby PM values and the site soiling ratio suggest that long term average PM measurements can be used to predict the variation in soiling losses between locations.

It is important to highlight that the satellite-based $PM_{2.5}$ did not show any significant correlation with soiling losses (p-value was found to be higher than 0.10). This might be due to the different time scales between satellite-derived $PM_{2.5}$ (based on the 2013 and 2014 data for this analysis) and the data collection times for the soiling stations (ranging between 2013 and 2016). Moreover, the numerous monitoring stations installed in the USA and used in this study could have provided more accurate values than the 0.1° × 0.1° grid available on the satellite map. Lastly, the methodology used to draw this satellite-based dataset is different from the gravimetric one generally employed in the monitoring stations. The satellite-based dataset is built by using aerosol optical depth data available and validated using nearby ground stations [35]. This means that this dataset directly estimates the concentration of suspended particulate and this, even if expected to be related to the deposition rate, might differ from the amount of particles deposited on a surface. Despite that, the utility of satellite-based data should be still

investigated, both because of the potential for improvement of satellite-based models and the lack of particulate matter data available in many regions of the world.

It is interesting to note that the number of PM sources reported in the 2011 National Emission Inventory (NEI) has been found to be a useful metric for the prediction of the soiling losses. In particular, the number of PM sources within 30 km resulted in a coefficient of determination on the order of 0.5 for both the soiling ratio and the soiling rate. On the other hand, the total amount of PM emitted by these sources did not show any significance when related to the soiling ratio or the soiling rate. This might be due to the different times at which the NEI was released and at which the soiling stations have been operating, and to the fact that the amount of emitted particulates is based on estimations.

It is not surprising that, among the precipitation characteristics, the time between consecutive rainfalls (the average length of the dry period) has shown the strongest correlation with the soiling ratio. Assuming the same soiling rate, longer times between consecutive rainfalls should lead to a greater reduction in soiling ratio than for shorter time periods. Although daily rainfall of less than 0.3 mm appeared to have no cleaning effect of the reference cells, modifying the PRISM dataset to exclude these rain events resulted in a lower correlation than the unfiltered PRISM dataset. It is also important to highlight the different impacts between the length of the dry periods and the percentage of rainy days. Soiling losses are expected to take place in the dry days between consecutive rainfalls. The number of rainy days alone does not differentiate between consecutive and non-consecutive rainy days and, for this reason, does not appear to be a reliable soiling predictor.

The average length of the dry period takes into account the precipitation profile of the whole year and may be the best metric for predicting annual soiling ratios, but the maximum length between consecutive rainfalls gives an idea of the seasonal trend. During these longest non-rainy time periods, the SRatio is expected to reach its minimum, and these are times when planned cleaning events may be most important. For the determination of annual or long-term losses, therefore, the average length of the dry period should be considered, whereas the maximum dry period might become the dominant factor in seasonal analysis and cleaning schedules.

Several interesting points of discussion can be noted from Fig. 2. First, only five of the twenty sites under study had average soiling ratios less than 0.98. This indicates that this study would benefit by increasing the number of sites at some of the dirtiest locations, but it also points out that a wide range of geographic locations still cluster with soiling ratios between 0.98 and 1.00. As both PM measurements and the average length between rainfalls showed a correlation with the average soiling ratios, it may be that with data from enough sites, a set of boundary conditions could be established to

distinguish between low soiling sites (SRatio > 0.98) and more problematic sites. For example, if the annual $PM_{2.5}$ is below *X* μg/m^3 *or* the annual average length of the dry period is less than *Y* days, then the annual SRatio will be 0.98 or greater. It should also be noted that this work does not include some of the most extreme soiling locations such as the Middle East, India, and China. It is possible that in these extreme climates those correlations between long term average PM values and the long term soiling ratio will not hold true.

The results in this work should not be considered exhaustive. Soiling loss data from 20 soiling sites have provided valuable statistical results, but the analysis was limited to just a two-variable regression. In some cases, the limited number of sites did not allow an accurate discussion. For example, it is worth nothing that the variable describing the angle of incidence between the cell's orientation and the mean wind direction, derived from the NSRDB, showed correlations both with the SRatio and the SRate, with R^2 in the order of 0.22. In particular, the soiling metrics have been found to decrease when the angle increases. Despite that, because of the low value of the coefficient of correlation, it has not been possible to conduct an in-depth analysis of this result. Moreover, some parameters, such as the height at which modules are installed, the tilt angle, or the tracking system, have been previously proven to have an impact on soiling losses when varied at the same site, and therefore are still expected to have an impact on soiling at sites with otherwise similar conditions. In the present investigation, only ground-mounted soiling stations have been considered, predominantly installed at one or two meters from the soil surface. Twenty-five percent of the soiling stations had a single-axis tracker, whereas the others had cells oriented at a fixed tilt angle. Of the fixed stations, eleven were at 20° or 25°, two were at 45°, one was at 35°, and one was at 10°. In order to get reliable statistics on tilt, large enough clusters are needed to distinguish each group. The sites analyzed in this study did not meet this requirement for tilt angle, height and other secondary variables. In order to extend the current investigation, data from a larger number of sites should be analyzed. This would enhance the ability to distinguish the level of impact of parameters that are secondary to PM levels and rainfall statistics.

Conclusions

The present work reports the results of a systematic analysis of 20 soiling stations installed in the USA. The main scope of the investigation has been the identification of those local parameters that can best predict the soiling losses at any site. For this reason, a correlation analysis has been run between more than one hundred parameters describing the weather profile, pollution, and the land characteristic of each site, and the ratio between the short-circuit currents of a soiled cell and a regularly-

cleaned cell. The data collection process has been explained, and the description of all the parameters has been reported in an Appendix.

The results of the investigation, which quantify the strength of correlations between soiling and environmental factors (summarized in Fig. 1), showed that only metrics of particulate matter and some parameters describing the precipitation pattern had significant correlations with the soiling metrics. In particular, the annual average of the daily mean values of PM_{10} and $PM_{2.5}$ have been found to be the best parameters to predict the soiling losses at the investigated sites, both having a R^2 of 0.82. Considering, instead, the number of particulate matter emission sources located within a 30 to 50 km radius of the soiling site lowers the R^2 to a value between 0.45 and 0.55. Among the precipitation metrics, the average number of days between consecutive rainfalls showed the highest correlation with the soiling ratio. Investigations were completed to determine optimal distances and statistical techniques for determining the best metrics for PM_{10} and $PM_{2.5}$ for a PV site, but limitations in available stations prevented a clear conclusion. The correlation results that were achieved were primarily for PM_{10} and $PM_{2.5}$ stations within 50 km of the PV site, and therefore PM measurements within 50 km of a site can be said to provide meaningful data towards predicting PV soiling losses.

The number of soiling stations analyzed here (20) was considered the minimum acceptable for a preliminary investigation of a two-variable multilinear regression for the prediction of PV soiling losses. Among all the possible permutations, only two of them, obtained by the combination of two pollution parameters, have been able to enhance the correlation with the SRatio compared to the best single-variable regression available. None of the initial two-variable regressions conducted which included both a pollution and a rainfall parameter showed an R^2 higher than 0.68 and a combination of PM10 and mean wind direction resulted in a R^2 of 0.73. By introducing instead a binary description of the average length of the dry period, a two-variable regression with the $PM_{2.5}$ resulted in an adjusted R^2 as high as 0.90. The analysis of a larger number of soiling stations, also installed in various countries and with high expected soiling losses, would allow extending the present study to investigate the effect of secondary parameters that are known to impact soiling losses.

Acknowledgments

This work was completed under Contract No. DE-AC36-08GO28308 with the U.S. Department of Energy. Leonardo Micheli has been supported by Sapienza – University of Rome (Italy) through the 2014 International Visiting Fellowship program.

The authors wish to thank Sarah Kurtz and Michael Deceglie for their valuable comments; Billy Roberts, Nicholas Gilroy and Jonathan Duckworth for their help in the

data collection process; Prof. Govindasamy TamizhMani and his group at Arizona State University for sharing the performance of one of their stations; and Pat Kline for editing the text.

References

1. Kimber A, Mitchell L, Nogradi S, Wenger H. The Effect of Soiling on Large Grid-Connected Photovoltaic Systems in California and the Southwest Region of the United States. *Photovoltaic Energy Conversion, Conference Record of the 2006 IEEE 4th World Conference on*, 2006.

2. Canada S. Impacts of Soiling on Utility-Scale PV System Performance. *SolarPro Magazine* 2016. http://solarprofessional.com/articles/operations-maintenance/impacts-of-soiling-on-utility-scale-pv-system-performance [accessed November 21, 2016].

3. Caron JR, Littmann B. Direct monitoring of energy lost due to soiling on first solar modules in California. *IEEE Journal of Photovoltaics* 2013; **3**(1): 336–340. DOI: 10.1109/JPHOTOV.2012.2216859.

4. Mejia FA, Kleissl J. Soiling losses for solar photovoltaic systems in California. *Solar Energy* 2013; **95**: 357–363. DOI: 10.1016/j.solener.2013.06.028.

5. Guo B, Javed W, Figgis BW, Mirza T. Effect of dust and weather conditions on photovoltaic performance in Doha, Qatar. *1st Workshop on Smart Grid and Renewable Energy (SGRE 2015)* 2015. DOI: 10.1109/SGRE.2015.7208718.

6. AlBusair HA, Möller HJ. Performance Evaluation of CdTe PV Modules Under Natural Outdoor Conditions in Kuwait. *Eu PVSEC 25*, vol. 10, 2010. DOI: 10.4229/25thEUPVSEC2010-3BV.2.111.

7. Jones RK, Baras A, Saeeri A Al, Al Qahtani A, Al Amoudi AO, Al Shaya Y, *et al.* Optimized Cleaning Cost and Schedule Based on Observed Soiling Conditions for Photovoltaic Plants in Central Saudi Arabia. *IEEE Journal of Photovoltaics* 2016; **6**(3): 730–738. DOI: 10.1109/JPHOTOV.2016.2535308.

8. Salim A, Huraib F, Eugenio N. PV power-study of system options and optimization. *Proceedings of the 8th European PV Solar Energy Conference*, 1988.

9. Rahoma U, Hassan A, Elminir H, Fathy A. Effect of airborne dust concentration on the performance of PV modules. *Journal of Astronomical Society Egypt* 2005; **13**(1): 24–38.

10. B N, Saed S. Effects of dust on the performance of thermal and photovoltaic flat plat collectors in Saudi Arabia: Preliminary results. *Alternative Energy Sources* 1979; **2**: 145–152.

11. IIT Bombay. SERIIUS Site Soiling Rate of the World.

http://www.ncpre.iitb.ac.in/pages/SERIIUS_Soiling_rate_of_the_World.html [accessed November 21, 2016].

12. Micheli L, Muller M, Kurtz S. Determining the effects of environment and atmospheric parameters on PV field performance. *2016 IEEE 43rd Photovoltaic Specialist Conference (PVSC)*, Portland, OR: IEEE; 2016.

13. International PV Quality Assurance Task Force (PVQAT). Soiling and Dust: Task Group 12. http://www.pvqat.org/project-status/task-group-12.html [accessed November 21, 2016].

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

15. Lave M, Hayes W, Pohl A, Hansen CW. Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the United States. *IEEE Journal of Photovoltaics* 2015; **5**(2): 597–606. DOI: 10.1109/JPHOTOV.2015.2392938.

16. Hansen C, Pohl A, Jordan D. *Uncertainty and Sensitivity Analysis for Photovoltaic System Modeling*. Albuquerque, NM: 2013.

17. Hay JE, Davies JA. Calculation of the Solar Radiation Incident on an Inclined Surface. *Proceedings of the First Canadian Solar Radiation Workshop*, Ministry of Supply and Services, Canada; 1980. DOI: 10.1016/0960-1481(93)90104-O.

18. Gostein M, Caron JR, Littmann B. Measuring soiling losses at utility-scale PV power plants. *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)* 2014: 0885–0890. DOI: 10.1109/PVSC.2014.6925056.

19. Deceglie MG, Muller M, Defreitas Z, Kurtz S. A Scalable Method for Extracting Soiling Rates from PV Production Data. *2016 IEEE 43rd Photovoltaic Specialist Conference (PVSC)*, 2016.

20. Di Luzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology* 2008; **47**(2): 475–497. DOI: 10.1175/2007JAMC1356.1.

21. Daly C, Smith JI, Olson K V. *Mapping atmospheric moisture climatologies across the conterminous United States*. vol. 10. 2015. DOI: 10.1371/journal.pone.0141140.

22. PRISM Climate Group - Oregon State University. PRISM Climate Group - Oregon State University. http://www.prism.oregonstate.edu/explorer/ [accessed November 21, 2016].

23. National Renewable Energy Laboratory. National Solar Radiation Data Base (NSRDB). https://nsrdb.nrel.gov/ [accessed November 21, 2016].

24. Sengupta M, Habte A, Gotseff P, Weekley A, Lopez A, Molling C, *et al. A Physics-Based GOES Satellite Product for Use in NREL's National Solar Radiation Database*. Golden, CO: 2014.

25. Sengupta M, Habte A, Gotseff P, Weekley A, Lopez A, Anderberg M, *et al. Physics-Based GOES Product for Use in NREL's National Solar Radiation Database*. Golden, CO: 2014.

26. Vandormael PJ, Couderc P. An algorithm to ensure spatial consistency in collaborative photo collections. In: Lee S, Narasimhan P, editors. *Software Technologies for Embedded and Ubiquitous Systems*, vol. 5860 LNCS, Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. DOI: 10.1007/978-3-642-10265-3_30.

27. US Environmental Protection Agency. What is PM, and how does it get into the air? 2016. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM [accessed November 21, 2016].

28. US Environmental Protection Agency. Report on the Environment: Particulate Matter Emissions 2011. https://cfpub.epa.gov/roe/indicator_pdf.cfm?i=19 [accessed November 21, 2016].

29. Goossens D, Van Kerschaever E. Aeolian dust deposition on photovoltaic solar cells: The effects of wind velocity and airborne dust concentration on cell performance. *Solar Energy* 1999; **66**(4): 277–289. DOI: 10.1016/S0038-092X(99)00028-6.

30. Kaldellis JK, Fragos P, Kapsali M. Systematic experimental study of the pollution deposition impact on the energy yield of photovoltaic installations. *Renewable Energy* 2011; **36**(10): 2717–2724. DOI: 10.1016/j.renene.2011.03.004.

31. Mani M, Pillai R. Impact of dust on solar photovoltaic (PV) performance: Research status, challenges and recommendations. *Renewable and Sustainable Energy Reviews* 2010; **14**(9): 3124–3131. DOI: 10.1016/j.rser.2010.07.065.

32. Boyle L, Flinchpaugh H, Hannigan M. Ambient airborne particle concentration and soiling of PV cover plates. *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)* 2014: 3171–3173. DOI: 10.1109/PVSC.2014.6925609.

33. US Environmental Protection Agency. Air Quality System Data Mart [internet database]. https://www.epa.gov/airdata [accessed November 21, 2016].

34. Falke S, Husar R, Schichtel B. Mapping Air Pollutant Concentrations from Point Monitoring Data I: Declustering and Temporal Variance. http://capita.wustl.edu/capita/capitareports/mappingairquality/mappingaqi.pdf [accessed November 21, 2016].

35. Van Donkelaar A, Martin R V., Brauer M, Hsu NC, Kahn RA, Levy RC, *et al.* Global Estimates of Fine Particulate Matter using a Combined Geophysical-

Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science and Technology* 2016; **50**(7): 3762–3772. DOI: 10.1021/acs.est.5b05833.

36. Atmospheric Composition Analysis Group - Dalhousie University. Satellite-Derived PM2.5 datasets. http://fizz.phys.dal.ca/~atmos/martin/?page_id=140 [accessed November 21, 2016].

37. US Environmental Protection Agency. 2011 National Emissions Inventory (NEI) Data 2011. https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-data [accessed November 21, 2016].

38. Soil Survey Staff, Natural Resources Conservation Service USD of A. Web Soil Survey. http://websoilsurvey.nrcs.usda.gov/ [accessed November 21, 2016].

39. US Environmental Protection Agency. Air Pollutant Emissions Trends Data 1970-2014 2016. https://www.epa.gov/sites/production/files/2015-07/national_tier1_caps.xlsx [accessed November 21, 2016].

40. US Forest Service. Wildfire Hazard Potential 2016. http://www.arcgis.com/home/item.html?id=fc0ccb504be142b59eb16a7ef44669a3 [accessed November 21, 2016].

41. Weinstock L, Watkins N, Wayland R, Baldauf R. *EPA's Emerging Near-Road Ambient Monitoring Network*. 2013.

42. Karner AA, Eisinger DS, Niemeier DA. Near-roadway air quality: Synthesizing the findings from real-world data. *Environmental Science and Technology* 2010; **44**(14): 5334–5344. DOI: 10.1021/es100008x.

43. US Environmental Protection Agency. Near Roadway Air Pollution and Health: Frequently Asked Questions 2014(August). https://www.epa.gov/sites/production/files/2015-11/documents/420f14044_0.pdf [accessed November 21, 2016].

44. Gustafsson MER, Franzen LG. Dry deposition and concentration of marine aerosols in a coastal area, SW Sweden. *Atmospheric Environment* 1996; **30**(6): 977–989. DOI: 10.1016/1352-2310(95)00355-X.

45. Feliu S, Morcillo M, Chico B. Effect of Distance from Sea on Atmospheric Corrosion Rate. *Corrosion* 1999; **55**(9): 883–891. DOI: 10.5006/1.3284045.

46. Meira GR, Andrade MC, Padaratz IJ, Alonso MC, Borba JC. Measurements and modelling of marine salt transportation and deposition in a tropical region in Brazil. *Atmospheric Environment* 2006; **40**(29): 5596–5607. DOI: 10.1016/j.atmosenv.2006.04.053.

47. Meira GR, Andrade C, Alonso C, Padaratz IJ, Borba JC. Modelling sea-salt transport and deposition in marine atmosphere zone - A tool for corrosion studies. *Corrosion Science* 2008; **50**(9): 2724–2731. DOI: 10.1016/j.corsci.2008.06.028.

48. Cole IS, Paterson D a., Ganther WD. Holistic model for atmospheric corrosion Part 1 - Theoretical framework for production, transportation and deposition of marine salts. *Corrosion Engineering, Science and Technology* 2003; **38**(2): 129–134. DOI: 10.1179/147842203767789203.

49. Salt Institute. The Snowfighters Handbook 2013. http://www.saltinstitute.org/wp-content/uploads/2013/07/Snowfighters_HB_2012.pdf [accessed November 21, 2016].

50. Transportation Research Board. Road Salt Impacts on the Environment. *Highway Deicing: Special Report 235* 1991: 69–82. http://onlinepubs.trb.org/onlinepubs/sr/sr235/069-082.pdf [accessed November 21, 2016].

51. Boyle L, Flinchpaugh H, Hannigan M. Assessment of PM Dry Deposition on Solar Energy Harvesting Systems: Measurement - Model Comparison. *Aerosol Science and Technology* 2016; **6826**(February): 0–0. DOI: 10.1080/02786826.2016.1153797.

52. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 2006; **15**(3): 259–263. DOI: 10.1127/0941-2948/2006/0130.

53. Multi-Resolution Land Characteristics Consortium. National Land Cover Database 2006 (NLCD2006): Product Legend 2006. http://www.mrlc.gov/nlcd06_leg.php [accessed November 21, 2016].

Table 1 - Description of location, land cover and characteristic weather for each soiling site. The land cover has been determined by using the USDA soil survey [38] and then checked by visually inspecting the area surrounding the soiling station using satellite imagery. The characteristic weather has been identified by using the Köppen-Geiger climate classification [52].

Site	Data collection period [mm/yy - mm/yy]	County	Land Cover	Characteristic weather
A	09/14 - 05/16	San Miguel, NM	Shrub/Scrub	Arid cold steppe
B	06/14 -12/15	Luna, NM	Shrub/Scrub	Cold arid desert
C	12/14 -12/15	Imperial, CA	Cultivated Crops	Hot arid desert
D	05/15 - 06/16	Madera, CA	Cultivated Crops	Arid cold steppe
E	02/13 - 12/15	Kern, CA	Shrub/Scrub	Warm temperate with dry summer
F	07/14 - 12/15	Kern, CA	Shrub/Scrub	Warm temperate with dry summer
G	01/14 - 06/16	Yuma, AZ	Shrub/Scrub	Hot arid desert
H	06/15 - 07/16	San Luis Obispo, CA	Shrub/Scrub	Warm temperate with dry summer
I	01/13 - 12/15	Pima, AZ	Shrub/Scrub	Arid hot steppe
J	05/15 - 12/15	Honolulu, HI	Hay/Pasture	Equatorial savannah with dry summer
K	02/13 - 06/14	Kern, CA	Cultivated Crops	Arid cold steppe
L	02/13 - 11/14	Kern, CA	Cultivated Crops	Arid cold steppe
M	09/14 - 08/15	Riverside, CA	Developed, Open Space	Warm temperate with hot dry summer
N	07/14 - 08/15	Pueblo, CO	Shrub/Scrub	Arid cold steppe
O	12/14 - 12/15	Winkler, TX	Shrub/Scrub	Arid cold steppe
P	04/14 - 04/15	Iron, UT	Shrub/Scrub	Warm temperate with dry summer
Q	12/13 - 01/15	Kern, CA	Shrub/Scrub	Warm temperate with dry summer
R	12/13 - 02/15	Polk, FL	Shrub/Scrub	Warm temperate, fully humid, with hot summer
S	05/13 - 05/16	Arapahoe, CO	Cultivated Crops	Arid cold steppe
T	12/14 - 09/16	Maricopa, AZ	Developed, Medium Intensity	Hot arid desert

Table 2 - Soil cover categories of the investigated sites and related numerical and binary classifications. The description of each category has been sourced from [53].

Classification	Description	Numerical class	Binary class
Developed, Medium Intensity	Areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover.	1	0
Developed, open space	Areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover.	1	0
Grassland/Herbaceous	Areas dominated by gramanoid or herbaceous vegetation, generally greater than 80% of total vegetation. Not subject to intensive management (i.e. tilling).	2	1
Hay/Pasture	Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle.	2	1
Shrub/Scrub	Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation.	3	1
Cultivated Crops	Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards.	3	1

Table 3 - The fire regime categories, their definitions and the numerical and binary classifications introduced for this study. The description of each fire regime category is reported by [25].

Fire regime	Description	Numeric category	Binary category
35-100+ years; Mixed Severity	Non-frequent mixed-severity fires replacing up to 75% of the dominant overstock vegetation.	0	0
35-100+ years; Stand Replacement	Non-frequent fires lethal to the most ground vegetation.	1	0
0-35 years; Low Severity	Frequent low-severity fires replacing less than 25% of the dominant overstock vegetation.	2	1
0-35 years; Stand Replacement	Frequent high-severity fires replacing more than 75% of the dominant overstock vegetation.	3	1

Table 4 - Comparison of metrics for reporting PM_{10} and $PM_{2.5}$ levels at the 11 sites with at least one monitoring station within 30 km. The best coefficients of determination are underlined. The averaging methods have been described in the "Pollution" section.

Distance	**30 km**			**50 km**			**100 km**		
Averaging method	**Decl. Mean**	**DW Mean**	**Arithm. Mean**	**Decl. Mean**	**DW Mean**	**Arithm. Mean**	**Decl. Mean**	**DW Mean**	**Arithm. Mean**
R^2 SRatio PM_{10}	0.76	0.79	0.81	0.77	0.79	<u>0.84</u>	0.02	0.70	0.39
R^2 SRatio $PM_{2.5}$	0.82	0.84	<u>0.84</u>	0.81	0.81	0.80	0.74	0.81	0.57

Table 5 - Adjusted R^2 for two-variable linear regressions using all the possible combinations among the parameters shown in Fig. 1. Cells colored in grey indicate that the independent variable is the SRate, whereas non-colored cells indicate that the independent variable is the SRatio. Empty cells mean that at least one of the variables had a p-value higher than 0.05.

	Average PM_{10} of best available monitoring stations	Average $PM_{2.5}$ of best available monitoring stations	Average $PM_{2.5}$ of closest monitoring stations	Average PM_{10} of closest monitoring stations	Number of PM_{10} sources within 10 km	Number of $PM_{2.5}$ sources within 10 km	Number of PM_{10} sources within 30 km	Number of $PM_{2.5}$ sources within 30 km	Number of PM_{10} sources within 50 km	Number of $PM_{2.5}$ sources within 50 km	Average length of the dry period	Max length of the dry period	Average length of the dry period (only rain > 0.3mm)	Max length of the dry period (only rain > 0.3mm)	Number of rainy days per year	# of rainy days per year (only rain > 0.3 mm)	Mean wind direction (AOI at noon)
Average PM_{10} of best available monitoring stations		0.88	0.87														
Average $PM_{2.5}$ of best available monitoring stations																	
Average $PM_{2.5}$ of closest monitoring stations																	
Average PM_{10} of closest monitoring stations																	0.73
Number of PM_{10} sources within 10 km											0.59		0.54		0.40	0.38	
Number of $PM_{2.5}$ sources within 10 km											0.59		0.54		0.40	0.38	
Number of PM_{10} sources within 30 km											0.69	0.61	0.68	0.60	0.62	0.62	
Number of $PM_{2.5}$ sources within 30 km											0.69	0.61	0.68	0.60	0.62	0.62	
Number of PM_{10} sources within 50 km	0.61										0.64		0.60	0.52	0.51	0.51	
Number of $PM_{2.5}$ sources within 50 km	0.61										0.64		0.60	0.52	0.51	0.51	
Average length of the dry period									0.59	0.59							
Max length of the dry period							0.67	0.67	0.67	0.67							
Average length of the dry period (only rain > 0.3 mm)																	
Max length of the dry period (only rain > 0.3 mm)																	
Number of rainy days per year											0.47	0.70					
Number of rainy days per year (only rain > 0.3 mm)											0.48	0.70					
Mean wind direction (AOI at noon)																	

Appendix

Complete list of parameters considered in the present investigation.

Parameter	Description	Range	Source
Site characteristics			
Altitude [m]	Vertical distance between the site's ground and the average sea level	-40 to 1820	
Distance from highway [km]	Distance between the site and the closest highway	0.75 to 15.44	
Distance from dirt road [km]	Distance between the site and the closest dirt road	0.04 to 9.38	
Distance from ocean [km]	Distance between the site and the closest seashore	5 to 1222	
Salt belt state	Binary classification attributed to the soiling stations depending on the state's legislation on the use of de-icing road salt.	0 (if salt is not used) to 1	
Fire regime (Binary)	Binary classification attributed to the soiling stations depending on the wildfire risk.	0 (if low risk and frequency) to 1	[40]
Fire regime (Numerical)	Numerical classification attributed to the soiling stations depending on the wildfire risk.	0 (if low risk and frequency) to 3	[40]
Land cover (Numerical)	Numerical classification attributed to the soiling stations depending on the land cover.	1 (if low impact from land cover is expected) to 3	[38]
Land cover (Binary)	Binary classification attributed to the soiling stations depending on the land cover.	0 (if low impact from land cover is expected) to 1	[38]
Wind erosion index [tons]	Amount of soil yearly removed per acre due to wind erosion	48 to 220	[38]
Wind erosion group	Classes of susceptibility of ground surface layer to wind erosion	1 (if high) to 6	[38]
Percentage of clay in the soil [%]	Mineral particles less than 0.002mm in equivalent diameter as a weight percentage of the less than 2.0mm fraction	1.5 to 42.5	[38]
Soil pH	Relative acidity or alkalinity of the soil surface layer.	5.6 to 8.2	[38]
Soiling station characteristics			
Time period [days]	Number of days between the start and the end of the data acquisition	214 to 1096	

Height [m]	Vertical distance between the center of the module and the ground	0.7 to 2	
Tilt [deg]	Tilt angle of the modules on the soiling stations. For tracked system, a tilt of 45 degrees has been considered.	10 to 45	
Tilt (Numerical)	Numerical classification attributed to the soiling stations depending on the modules tilt angle.	0 (if tracked) to 5 (if above 40°)	
Tracking	Numerical classification attributed to the soiling stations depending on the presence or not of a tracking system.	1 (if tracked) to 2 (if fixed)	
Particulate matter			
Distance to nearest PM_{10} station [km]	Distance between the soiling station and the closest PM_{10} monitoring station operating during the investigated period.	4.5 to 86	[33]
PM_{10} recorded by the nearest PM_{10} monitoring station [$\mu g/m^3$]	Average of the annual PM_{10} values recorded each year by the closest monitoring stations.	10 to 60	[33]
Distance to nearest $PM_{2.5}$ station [km]	Distance between the soiling station and the closest $PM_{2.5}$ monitoring station operating during the investigated period.	4.8 to 147	[33]
$PM_{2.5}$ recorded by the nearest $PM_{2.5}$ monitoring station [$\mu g/m^3$]	Average of the annual $PM_{2.5}$ values recorded each year by the closest monitoring stations.	4.4 to 22	[33]
Number of PM_{10} sources within a set distance	Number of PM_{10} sources located within a set distance from the soiling station, as listed in the 2011 National Emissions Inventory. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	0 to 67 [10 km] 1 to 548 [30 km] 2 to 751 [50 km] 29 to 1736 [100 km]	[37]
PM_{10} produced by sources within a set distance [tons]	Amount of PM_{10} emitted by all the sources located within a set radius from the soiling stations, as listed in the 2011 National Emissions Inventory. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	0 to 591 [10 km] 1 to 548 [30 km] 0 to 4456 [50 km] 24 to 6944 [100 km]	[37]
Number of $PM_{2.5}$ sources within a set distance	Number of $PM_{2.5}$ sources located within a set distance from the soiling station, as listed in the 2011 National Emissions Inventory. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	0 to 67 [10 km] 1 to 548 [30 km] 2 to 751 [50 km] 29 to 1736 [100 km]	[37]

$PM_{2.5}$ produced by sources within a set distance [tons]	Amount of $PM_{2.5}$ emitted by all the sources located within a set radius from the soiling stations, as listed in the 2011 National Emissions Inventory. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	0 to 279 [10 km] 0 to 2016 [30 km] 2 to 2673 [50 km] 22 to 4385 [100 km]	[37]
Arithmetic mean of thePM_{10} recorded by monitoring stations located within a set distance [$\mu g/m^3$]	PM_{10} calculated as arithmetic mean of the annual average of the daily records of the stations located within a set distance from the PV soiling station. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	14.0 to 25.9 [10 km] 14.1 to 56.3 [30 km] 16.0 to 56.3 [50 km] 12.9 to 44.4 [100 km]	[33]
Distance weighted mean of the PM_{10} recorded by monitoring stations located within a set distance [$\mu g/m^3$]	PM_{10} calculated as mean of the annual average of the daily records of the stations located within a set distance, weighted considering each station's distance from the PV soiling station. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	14.0 to 25.9 [10 km] 14.0 to 56.3 [30 km] 14.3 to 56.3 [50 km] 12.9 to 53.4 [100 km]	[33]
Declustered configuration mean of the PM_{10} recorded by monitoring stations located within a set distance [$\mu g/m^3$]	PM_{10} calculated as mean of the annual average of the daily records of the stations located within a set distance, considering the distance from the PV soiling station as well as the average distances among the monitoring stations. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	14.0 to 25.9 [10 km] 14.0 to 56.3 [30 km] 14.1 to 56.4 [50 km] 12.9 to 47.5 [100 km]	[33]

Number of PM_{10} stations within a set distance	Number of PM_{10} monitoring stations installed within a set radius from the PV soiling station. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	0 to 1 [10 km] 0 to 9 [30 km] 0 to 19 [50 km] 1 to 28 [100 km]	[33]
Arithmetic mean of the $PM_{2.5}$ recorded by monitoring stations located within a set distance [$\mu g/m^3$]	$PM_{2.5}$ calculated as arithmetic mean of the annual average of the daily records of the stations located within a set distance from the PV soiling station. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	5.2 to 6.5 [10 km] 4.5 to 20.4 [30 km] 4.5 to 19.7 [50 km] 4.5 to 13.1 [100 km]	[33]
Distance weighted mean of the $PM_{2.5}$ recorded by monitoring stations located within a set distance [$\mu g/m^3$]	$PM_{2.5}$ calculated as mean of the annual average of the daily records of the stations located within a set distance, weighted considering each station's distance from the PV soiling station. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	5.2 to 6.5 [10 km] 5.0 to 20.7 [30 km] 5.0 to 19.8 [50 km] 4.5 to 19.4 [100 km]	[33]
Declustered configuration mean of the $PM_{2.5}$ recorded by monitoring stations located within a set distance [$\mu g/m^3$]	$PM_{2.5}$ calculated as mean of the annual average of the daily records of the stations located within a set distance, considering the distance from the PV soiling station as well as the average distances among the monitoring stations. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	5.2 to 6.5 [10 km] 5.0 to 21.8 [30 km] 5.0 to 21.6 [50 km] 4.5 to 21.5 [100 km]	[33]
Number of $PM_{2.5}$ stations within a set distance	Number of $PM_{2.5}$ monitoring stations installed within a set radius from the PV soiling station. Repeated for distances of 10 km, 30 km, 50 km and 100 km.	0 to 1 [10 km] 0 to 5 [30 km] 0 to 12 [50 km] 0 to 12 [100 km]	[33]
Average PM_{10} of best available monitoring stations [$\mu g/m^3$]	Combination of the arithmetical average of the PM_{10} recorded by stations within 50 km and the PM_{10} recorded by the nearest PM_{10} monitoring station for those sites with no PM_{10} monitoring stations within 50 km.	13.4 to 56.3	[33]
Average $PM_{2.5}$ of best available monitoring stations [$\mu g/m^3$]	Combination of the arithmetical average of the $PM_{2.5}$ recorded by stations within 30 km and the $PM_{2.5}$ recorded by the nearest $PM_{2.5}$ monitoring station for those sites with no $PM_{2.5}$ monitoring stations within 30 km.	4.5 to 20.4	[33]

Satellite-derived $PM_{2.5}$ [$\mu g/m^3$]	Global satellite-derived $PM_{2.5}$ averaged over 2001–2006	6.5 to 13	[35,36]
Precipitation			
Daily accumulated precipitation [mm]	Amount of rain accumulated during the investigated period per day	0.1 to 4.1	[20–22]
Daily accumulated precipitation for rain > 0.3 mm [mm]	Amount of rain accumulated during the investigated period per day considering rain higher than 0.3 mm only	0.1 to 4.1	[20–22]
Rainy days [%]	Number of days with a rain event	5.5 to 60.9	[20–22]
Rainy days for rain > 0.3 mm [%]	Number of days with a rain event considering rain higher than 0.3 mm only	4.2 to 50.7	[20–22]
Precipitation intensity [mm]	Average amount of rain accumulated on each rainy day	2.1 to 9.8	[20–22]
Precipitation intensity for rain > 0.3 mm [mm]	Average amount of rain accumulated on each rainy day considering rain higher than 0.3 mm only	2.5 to 11.1	[20–22]
Length of dry period [days]	Average number of non-rainy days between two rain events	2 to 27	[20–22]
Length of dry period for rain > 0.3 mm [days]	Average number of non-rainy days between two rain events considering rain higher than 0.3 mm only	2 to 29.1	[20–22]
Maximum length dry period [days]	Maximum number of consecutive days with no rain recorded	8 to 155	[20–22]
Maximum length dry period for rain > 0.3 mm [days]	Maximum number of consecutive days with no rain recorded considering rain higher than 0.3 mm only	10 to 161	[20–22]
Days with more than 12.5 mm [days]	Average number of days with an accumulated precipitation of at least 12.5 mm	1.5 to 27.4	[20–22]
Days with more than 25 mm [days]	Average number of days with an accumulated precipitation of at least 25 mm	0 to 3.7	[20–22]
Days with more than 50 mm [days]	Average number of days with an accumulated precipitation of at least 50 mm	0 to 1.4	[20–22]
Rain on the wettest day [mm]	Maximum amount of rain accumulated in one day considering rain higher than 0.3 mm only	11 to 125	[20–22]
Rain on the five wettest days [mm]	Rain accumulated in the five days with the highest precipitation intensity	31 to 385	[20–22]
Other meteorological parameters			
Number of days with dew cycles [%]	Percentage of days on which these conditions occur for at least one hour: RH ≥ 95% and wind speed ≤ 3.2 m/s and ambient temperature > 0°.	0 to 93	
Number of days with dew cycles (RH ≥ 99%) [%]	Percentage of days on which the relative humidity is equal or higher than 99% for at least one hour	0 to 90	
Average RH [%]	Average relative humidity during the investigated period.	55.0 to 99.5	

Average wind speed [m/s]	Average wind speed recorded on the site during the investigated period.	0.4 to 4.9	
Average peak gust [m/s]	Average of the hourly peak gusts recorded on the site during the investigated period.	1.0 to 5.7	
Onsite mean wind direction (0 to 360°) [°]	Mean wind direction recorded on the site during the investigated period (0° is no wind and 360° is wind blowing from north).	25 to 325	
Onsite angle of incident at noon [°]	Absolute value of the angle difference between the mean wind direction derived (by the data onsite) and the azimuth angle of the cells at noon (if 0°, mean wind direction is blowing from south).	17 to 154	
Days with peak above 5m/s [%]	Number of days with a maximum hourly peak gusts above 5 m/s.	3.4 to 67.2	
Days with peak above 10m/s [%]	Number of days with a maximum hourly peak gusts above 10 m/s.	0 to 24.8	
Days with peak gust above average [%]	Number of days with a maximum hourly peak gusts above the average peak gusts.	15.8 to 55.7	
Peak gust above 2*average	Number of days with a maximum hourly peak gusts twice above the average peak gusts.	1.3 to 43.5	
NSRDB Average wind speed [m/s]	Average wind speed recorded on the NSRDB database during the investigated period.	1.7 to 3.9	[23–25]
NSRDB mean wind direction (0 to 360°) [°]	Mean wind direction recorded on the NSRDB database during the investigated period (0° is no wind and 360° is wind blowing from north).	55 to 331	[23–25]
NSRDB angle of incident at noon [°]	Absolute value of the angle difference between the mean wind direction derived by the NSRDB database and the azimuth angle of the cells at noon (if 0°, mean wind is blowing from south).	7 to 151	[23–25]