NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Predicting Future Soiling Losses Using Environmental Data

Preprint

Leonardo Micheli and Michael G. Deceglie
National Renewable Energy Laboratory

Presented at the 35th European PV Solar Energy Conference & Exhibition (EU PVSEC)
Brussels, Belgium
September 23–28, 2018

Suggested Citation

Micheli, L., M.G. Deceglie. 2018. "Predicting Future Soiling Losses Using Environmental Data: Preprint." Golden, CO: National Renewable Energy Laboratory. NREL/CP-5K00-71127. https://www.nrel.gov/docs/fy19osti/71127.pdf.

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

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

Conference Paper
NREL/CP-5K00-71127
October 2018

Contract No. DE-AC36-08GO28308

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. 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.

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

NREL prints on paper that contains recycled content.

PREDICTING FUTURE SOILING LOSSES USING ENVIRONMENTAL DATA

Leonardo Micheli, Michael G. Deceglie
National Renewable Energy Laboratory, Golden, Colorado, United States

ABSTRACT: The correlations between soiling and parameters describing the atmospheric, pollution and geometric characteristic of a site have already been investigated in the literature. In particular, particulate matter and rainfall data have been shown to be the best parameters to estimate soiling losses occurring at photovoltaic sites in the United States. However, previous investigations only considered soiling and environmental data collected over the same time periods. The aim of the present research is, instead, to understand if future soiling losses can be predicted using environmental parameters recorded during previous time periods. To do this, we compared the soiling occurring at 41 U.S. locations with particulate matter and rainfall data recorded 1 to 30 years earlier.

Keywords: Soiling, Reliability, System Performance, Monitoring

1 INTRODUCTION

Soiling consists of accumulated dust and dirt on the surfaces of photovoltaic (PV) modules that reduces energy conversion [1]. This loss leads to a lower and more uncertain energy yield, which means reduced revenues, higher operation and maintenance costs, and more expensive financing.

Investors generally consider an average monthly yearly soiling loss to estimate the energy yield and the economic revenue of a site for a new PV installation. The yearly soiling losses can be estimated by using soiling data available from nearby sites [2] or by considering environmental data available at the site [3–6]. In the first case, the soiling losses can be directly measured using soiling stations [7,8] or by using soiling extraction algorithms to identify soiling trends in PV performance data [9,10]. In the second case, widely available parameters, describing the weather and pollution patterns, can be used to estimate the soiling losses occurring at different sites. In particular, our previous works showed how average particulate matter concentrations and rainfall data correlate with the annualized soiling losses registered at a number of sites in the United States [4,5].

In our previous works, we compared only soiling and environmental data collected over the same time periods. However, when a new PV site is selected, historical environmental data must be used to predict future losses. For this reason, the present work focuses on understanding the correlations between historical pollution and rainfall trends and future soiling losses.

2 METHODOLOGY

2.1 Soiling data

The soiling data used in this work have been extracted from 41 soiling stations installed in the United States between 2013 and 2016 for time periods ranging between 7 and 40 months [11]. We analyzed the same data of Ref. [5] and compared them with environmental parameters that occurred during the soiling data collection period. A soiling station is a system consisting of two PV devices—two cells, two modules or a combination of both—that are mounted outdoor with the same tilt, azimuth, and height. One of the two devices is cleaned regularly (reference device), whereas the second one is left to soil naturally (soiled device). Soiling is quantified by comparing the electrical output of the two devices using a "soiling ratio" metric that is the ratio of the short-circuit current of the soiled device to the short-circuit current of the reference device [12]. The soiling ratio has a value of 1 if both the devices are cleaned, and it decreases as soiling accumulates on the soiled device. The same filters described in [2,4] have been used: we consider only data points occurring between 12 pm and 2 pm and for irradiances > 500 W/m^2.

2.2 Environmental data

Our present work considers only particulate matter and rainfall data as environmental parameters. Indeed, these have been found to be the most relevant parameters for predicting PV losses in the United States [4,5].

The particulate matter consists of the mixture of soiling particles and liquid droplets suspended in the air. It is commonly expressed as PM_{10} and $PM_{2.5}$, which describe the mass concentration of particles less than 10 microns and 2.5 microns in diameter, respectively, suspended in 1 m^3 of air. The values of the annual PM_{10} and $PM_{2.5}$ concentrations are from the U.S. Environmental Protection Agency (EPA) database [13], which records annual average concentrations from monitoring stations installed in the USA. These data have been analyzed using two standard air quality interpolation techniques [5,14]:

- Nearest neighbor: The particulate concentration at a soiling station site is obtained as the mean of the annual values recorded by the closest EPA monitoring station over the data collection period;
- Declustered distance estimation: The particulate concentration at a soiling station site is calculated as the weighted mean of the annual values recorded by the EPA monitoring stations located within a set distance of the soiling site (either 30, 50, 100 or 250 km). As described in [15], we used the inverse of the distance between each EPA monitoring station and the soiling site as well as a parameter describing the distances among the EPA monitoring stations as a weight to give more influence to the closest monitoring stations and to reduce the impact of spatially clustered monitoring stations.

The average and the maximum lengths of the dry periods, which describe the mean and maximum number of days between two consecutive rain events, have been calculated by using the daily rainfall data downloaded from the Oregon State University's PRISM database [16].

2.3 Statistical metrics

The soiling ratios of each site, calculated as the mean of the daily ratios measured over the data collection periods, were compared with the average value of the environmental parameters. These parameters describe the pollution and precipitation trends that occurred at each site for the year and for the 3, 5, 10 and 30 years before the soiling stations were installed. This means that, for example, if a station was installed in 2013, environmental data from 2010 to 2012 were considered for the "prior 3 years" analysis, and from 1983 to 2012 were considered for the "prior 30 years" analysis.

It is important to mention that, while the PM_{10} data have been available since 1982, whereas the first $PM_{2.5}$ concentrations were reported in 1997. Therefore, even if labelled as "30 years before", $PM_{2.5}$ values have been calculated as the average of the concentrations measured between 15 and 17 years before the start of the soiling data collection.

We used a linear single-variable regression analysis to study the correlations between soiling and each parameter. The adjusted coefficient of determination (adjR^2) has been employed to determine the quality of each correlation [4,5]. In addition, we discarded any correlation showing a p-value higher than 0.05.

3 RESULTS

In Ref. [5], we found that PM_{10} and $PM_{2.5}$ concentrations interpolated from the EPA dataset were the parameters with the best correlations to soiling. In particular, we found that the best correlation of PM_{10} occurred for distances between 30 km and 50 km, with R^2 up to 69%, independent of the interpolation techniques used. On the other hand, the correlations between $PM_{2.5}$ and soiling were more robust to the distance, with R^2 ranging between 70% and 63% for distances between 30 km and 250 km, if we used the declustered distance estimation technique.

The same analyses have been repeated in the present work, considering the particulate matter measured over different time periods. The results are shown in Figure 1 and Figure 2, with maximum adjusted coefficients of determination between 68% and 71% for both of the parameters. Overall, we found the same trends of the previous work [5]: PM_{10} performs better at shorter distances, whereas $PM_{2.5}$ seems to be more robust, having limited drops in R^2 at greater distances. In both cases, the nearest neighbor interpolation returns results similar to those obtained with the declustered distance estimation at larger distances (≥ 100 km).

For both indexes, the best results are obtained if data are collected during the same time that soiling was measured. $PM_{2.5}$ returns more consistent correlations to soiling than PM_{10} when different time periods are considered, with adjusted R^2 always above 40%, independent of the method and distance. $PM_{2.5}$ returns higher R^2 values, with better results obtained when we use data collected for long time periods (≥ 10 years).

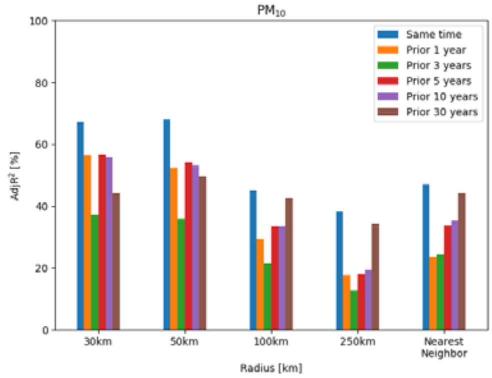

Figure 1: Coefficients of determinations, in %, obtained by interpolating the PM_{10} data using the declustered distance estimation with different radii and the nearest neighbor methods. The bars are colored according to the time the data were collected compared to the soiling data monitoring period.

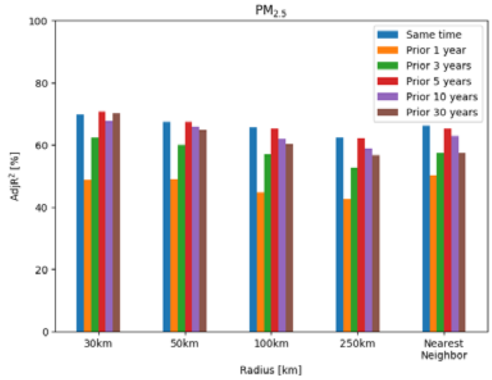

Figure 2: Coefficients of determinations, in %, obtained by interpolating the $PM_{2.5}$ data using the declustered distance estimation with different radii and the nearest neighbor methods. The bars are colored according to the time the data were collected compared to the soiling data monitoring period.

Other than the pollution parameters, the average length of the dry periods and the maximum length of the dry periods were the only significant variables showing correlations to soiling among those investigated in [4,5]. These same parameters were analyzed in this study: their correlations with soiling, depending on the data collection periods, are shown in Figure 3 and Figure 4. Overall, both these parameters return R^2 lower than the pollution concentrations. In particular, the correlation of the average length of the dry period with soiling drops rapidly, with R^2 equal to or lower than 20% if data are collected more than 1 year before soiling occurs. The maximum dry periods are more consistent to the variation in time periods. The best results are generally obtained if 3 years are averaged, and the correlations are slightly worse for longer time period (≥ 5 years). In both cases, the correlations are steady for thresholds ≤ 1mm, whereas they tend to decrease, especially for the maximum length of the dry period, for thresholds of 5 mm.

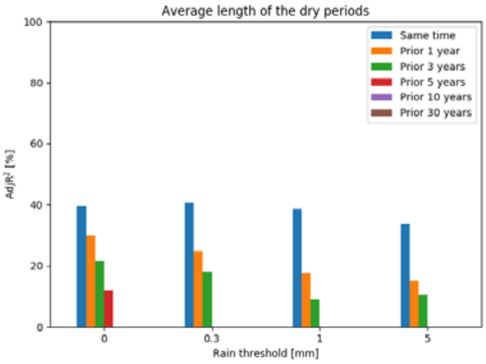

Figure 3: Coefficients of determinations, in %, obtained by comparing the average length of the dry periods with the soiling losses. The bars are colored according to the time the data were collected compared to the soiling data monitoring period.

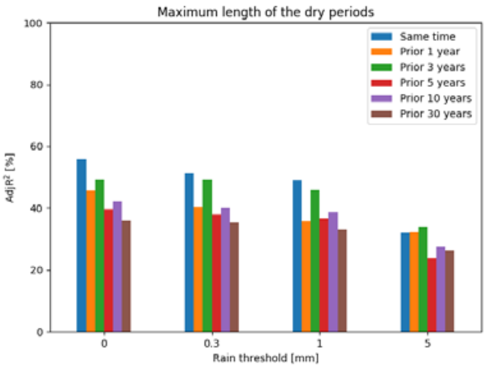

Figure 4: Coefficients of determinations, in %, obtained by comparing the maximum length of the dry periods with the soiling losses. The bars are colored according to the time the data were collected compared to the soiling data monitoring period.

4 DISCUSSION

Among the four parameters analyzed in this study, PM2.5 seems to be the most consistent to the different time periods considered. R^2 values consistently above 50% are obtained if data are averaged over long time periods (≥10 years), independently of the distance considered. This result is because the $PM_{2.5}$ concentrations of the most soiled sites are clearly higher than the rest of the sites (Figure 5). The same distinction cannot be done by instead considering PM_{10} concentrations in Figure 6.

Among rainfall data, the maximum length of the dry periods seems to perform better than the average length of the dry periods. This is probably because, over a 30 years period, the most soiled sites are more likely to experience long summer dry periods, where most of the soiling losses occur (Figure 7). On the other hand, the average dry period of the most soiled locations does not appear to be necessarily longer than those of the cleaner sites.

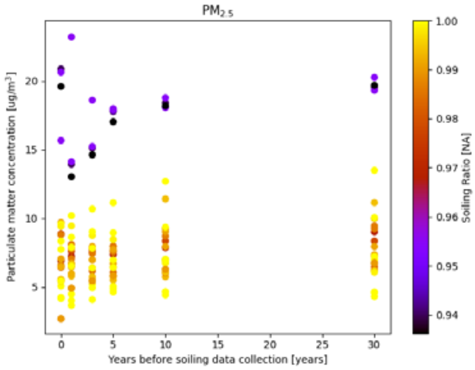

Figure 5: $PM_{2.5}$ concentrations at the 41 sites during the data collections and 1, 3, 5, 10 and 30 years before. The data are colored according to the average soiling ratio of the site.

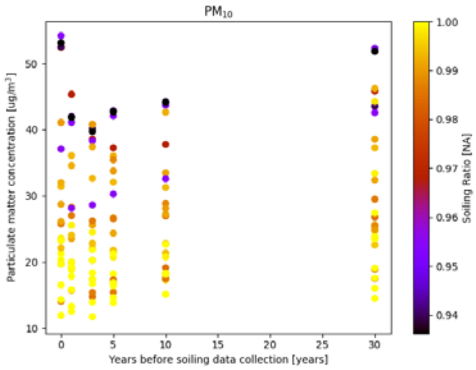

Figure 6: PM_{10} concentrations at the 41 sites during the data collections and 1, 3, 5, 10 and 30 years before. The data are colored according to the average soiling ratio of the site.

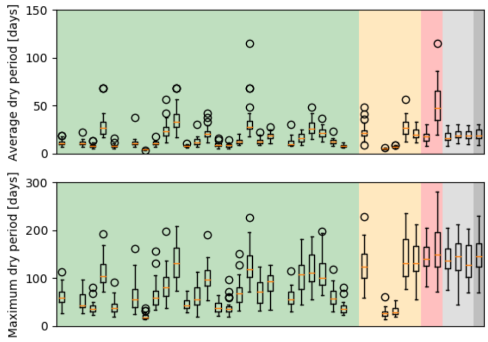

Figure 7: Boxplots representing the average and maximum lengths of the dry periods of each site over a 30 year periods. The background is colored according to the severity of soiling: green for soiling ratio ≥ 99%, orange for soiling ratio ≥ 98%, red for soiling ratio ≥ 96%, light grey for soiling ratio ≥ 94%, dark grey otherwise. Boxplots calculated for rain threshold of 1 mm.

5 CONCLUSIONS

This work analyzes the correlations between average historical environmental data and future soiling losses. We found the parameters describing the particulate matter concentrations to be the best predictors of soiling, even when historical data were used to estimate future soiling losses. In particular, $PM_{2.5}$ was found to be the most consistent predictor, independent of the time period considered, because of the steady distinction in $PM_{2.5}$ concentrations between high and low soiling sites. Among the rainfall parameters, the maximum length of the dry period returned the best results because of the consistent longer summer dry periods experienced by high soiling sites.

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] T. Sarver, A. Al-Qaraghuli, 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. 22 (2013) 698–733. doi:10.1016/j.rser.2012.12.065.

[2] L. Micheli, M.G. Deceglie, M. Muller, Mapping Photovoltaic Soiling Using Spatial Interpolation Techniques, IEEE J. Photovoltaics. (n.d.).

[3] L. Micheli, M. Muller, S. Kurtz, Determining the effects of environment and atmospheric parameters on PV field performance, in: 2016 IEEE 43rd Photovolt. Spec. Conf., IEEE, Portland, OR, 2016: pp. 1724 – 1729. doi:10.1109/PVSC.2016.7749919.

[4] L. Micheli, M. Muller, An investigation of the key parameters for predicting PV soiling losses, Prog. Photovoltaics Res. Appl. 25 (2017) 291–307. doi:10.1002/pip.2860.

[5] L. Micheli, M.G. Deceglie, M. Muller, Predicting photovoltaic soiling losses using environmental parameters: An update, Prog. Photovolt Res. Appl. (n.d.). doi:10.1002/adsc.201.

[6] W. Javed, B. Guo, B. Figgis, Modeling of photovoltaic soiling loss as a function of environmental variables, Sol. Energy. 157 (2017) 397–407. doi:10.1016/j.solener.2017.08.046.

[7] M. Gostein, J.R. Caron, B. Littmann, Measuring soiling losses at utility-scale PV power plants, 2014 IEEE 40th Photovolt. Spec. Conf. (2014) 0885–0890. doi:10.1109/PVSC.2014.6925056.

[8] M. Gostein, T. Duster, C. Thuman, Accurately Measuring PV Soiling Losses With Soiling Station Employing Module Power Measurements, in: IEEE 42nd Photovolt. Spec. Conf., 2015.

[9] A. Kimber, L. Mitchell, S. Nogradi, H. Wenger, The Effect of Soiling on Large Grid-Connected Photovoltaic Systems in California and the Southwest Region of the United States, in: Photovolt. Energy Conversion, Conf. Rec. 2006 IEEE 4th World Conf., 2006: pp. 2391–2395.

[10] M.G. Deceglie, L. Micheli, M. Muller, Quantifying Soiling Loss Directly from PV Yield, IEEE J. Photovoltaics. (2018). doi:10.1109/JPHOTOV.2017.2784682.

[11] National Renewable Energy Laboratory, Photovoltaic modules soiling map, (2018). https://www.nrel.gov/pv/soiling.html (accessed May 18, 2018).

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

[13] US Environmental Protection Agency, Air Quality System Data Mart [internet database], (n.d.). https://www.epa.gov/airdata (accessed May 15, 2018).

[14] D.W. Wong, L. Yuan, S.A. Perlin, Comparison of spatial interpolation methods for the estimation of air quality data, J. Expo. Anal. Environ. Epidemiol. 14 (2004) 404–415. doi:10.1038/sj.jea.7500338.

[15] S. Falke, R. Husar, B. Schichtel, Mapping Air Pollutant Concentrations from Point Monitoring Data I: Declustering and Temporal Variance, (n.d.). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.612.5742&rep=rep1&type=pdf (accessed May 18, 2018).

[16] PRISM Climate Group - Oregon State University, PRISM Climate Group - Oregon State University, (n.d.). http://www.prism.oregonstate.edu/explorer/ (accessed June 3, 2017).