PV ARRAY SOILAGE DETECTION ON CYPRUS SOLAR DATASET

Joshua Martin, Kristen Jaskie, and Andreas Spanias
SenSIP Center, School of ECEE, Arizona State University

***Abstract* – Solar panel soilage detection is an important problem as soiled panels produce significantly less energy than clean ones. In this paper, we present two new methods for identifying soiling in residential solar installations. The first method aims to calculate a daily energy-lost-due-to-soilage value by comparing two calculated power curves: the expected best case scenario curve (after taking degradation into account) and a weather corrected curve, which estimates what the day's power curve would be in the absence of cloud cover. Our second method takes a different approach and compares sites in the same weather region to each other using a multi-level k-means clustering strategy. The key takeaway being that these new methods do not need feature rich datasets, which are often unavailable, rather they operate solely on time-series power values.**

***Index Terms*— residential solar, k-means, soilage**

1. INTRODUCTION

Solar panels encounter degradation in their power output from many sources. One of which is soiling or dirtying of the panels from various sources (dust, calcium from rain, pollen, etc.). For grid scale installations, this soilage can be removed by regular cleaning schedules. However, this soilage poses a greater issue for residential installations from the point of view of the company managing them because cleaning is left up to the property owner. If left unchecked, this soilage can lead to significant degradation in power output.

In this study, we set out to develop a method to identify soilage in rooftop-mounted residential solar installations using a new dataset from the island country of Cyprus in the eastern Mediterranean. Our aim is to utilize time series power data from customer sites to either classify sites as soiled or unsoiled, or to rate sites on their level of soilage. This information can then be used to inform customers that they would benefit from cleaning their panels.

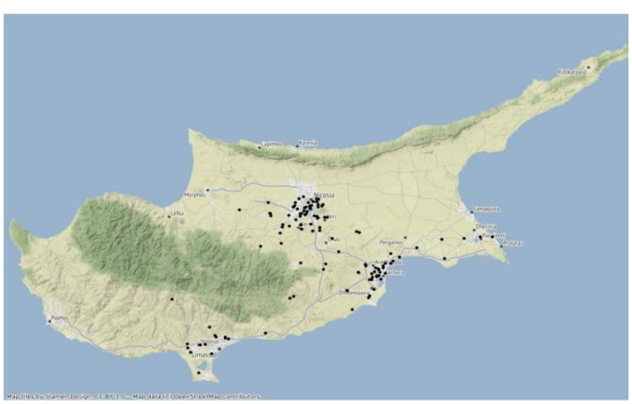


Figure 1: Geographical location of solar sites in Cyprus dataset

We approached this task in two different ways and ended up with two different classifiers. First, looking at high resolution data collected every 5 minutes, we estimated the degradation for each site on each day by correcting for weather and degradation over time. As part of this, we created a "sunny day" classifier. Next, we used geographical information to cluster PV arrays into groups with similar weather conditions. We normalized the data and used a machine learning clustering algorithm to classify soiled or degraded solar panels. We were able to get feet on the ground in Cyprus to provide feedback for our algorithm performance. These approaches will be described in more detail below.

This paper is broken into the following sections. We start by describing our dataset and feature space in depth in section 2. We then present our approach to estimating the levels of soilage degradation using a sunny day classifier we created in section 3. In section 4, we go in depth to describe our solar soilage classifier and describe how soilage can be identified over time. Results for both classifiers and feet on the ground feedback are provided in section 5. Finally, concluding discussion is provided in section 6. Acknowledgements and references are provided in sections 7 and 8.

2. CYPRUS SOLAR DATASET

The solar data used in this research is from the Republic of Cyprus and was provided by SolarEdge, a solar local solar company. This dataset contained 120 different homeowner rooftop solar sites, scattered across the island as shown in figure 1. Most solar arrays are in or around the main cities of Nicosia and Larnaca in the center of the island and the southeast, respectively. Data on solar power production in kW is collected every 5 minutes. This time-series data is provided for each site as it joined the grid. The oldest site has more than five years' worth of data while some of the newer sites only have a few months' worth of data. The data was live and being generated daily as we performed our analysis and soilage classification.

In addition to the power produced at each site, site metadata information including the GPS coordinates, the number, model, and orientation of the solar panels at the site, and the original site install date.

3. ESTIMATING SOILAGE DEGRADATION

Our initial approach to PV soilage classification was to look at the high-resolution data produced every 5 minutes. We found that some days were clearly sunny due to the nice Gaussian power curve over the course of a single day. Days with large amounts of cloud cover did not fit a Gaussian curve but instead were random and spiky as illustrated by the black shape in figure 2. From this, we were able to construct a "sunny day" classifier by measuring the mean squared error (MSE) between

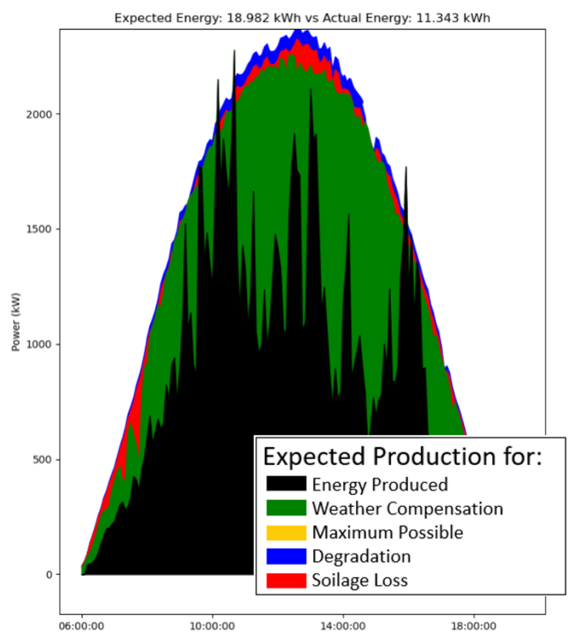


Figure 2: Energy production on a cloudy day with estimated weather compensation, degradation, and soilage loss.

the actual data and its closest Gaussian fit. Sunny days would have a small MSE while cloudy days had a high MSE.

To estimate the potential power that would be produced on a sunny day, we took the maximum energy produced at each 5-minute time period over all the sunny days in a 10- and 20-day window around the day of interest. As seasonal variations are not significant during this time and since Cyprus is a relatively sunny country, by taking the maximum for each time period, we were able to calculate the amount of energy that could be produced if the day were sunny. We could then identify the amount of energy production lost to weather (shown in green in figure 2).

Using the maximum values from sunny days in the same time window over the multi-year lifetime of the site, we were able to estimate a "maximum possible" energy production. This is the outer outline of the figure shown in figure 2. Solar panels have an annual degradation rate of 0.5-1% per year [1], [2]. To be on the safe side, we used 1% as an expected annual degradation rate. Calculating the expected degradation, we subtracted it from the maximum possible energy. This expected degradation is shown in blue in figure 2.

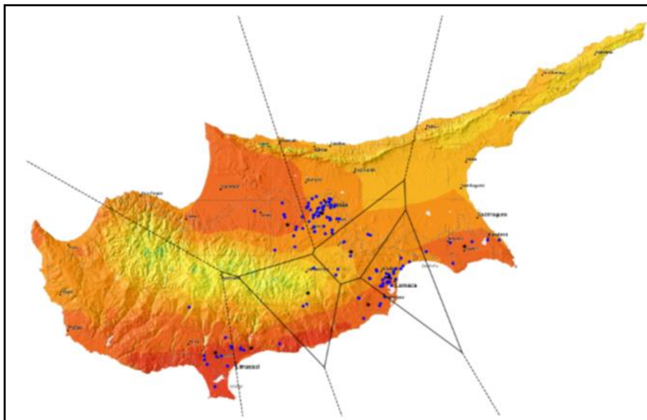
Finally, we estimated the soilage rate to be the difference between the expected maximum taking annual degradation into account and the weather corrected "sunny day" estimate. This estimated soilage is shown in red in figure 2.

While this method shows a lot of potential, we found that the estimated soilage levels were erratic and didn't behave in the expected pattern of a panel slowly becoming soiled before suddenly getting cleaned, whether manually or through a heavy rainstorm. Instead, we found that our predictions appeared random and pattern-less. Because of this, we moved on from a high-resolution data to a lower resolution daily soilage classifier.

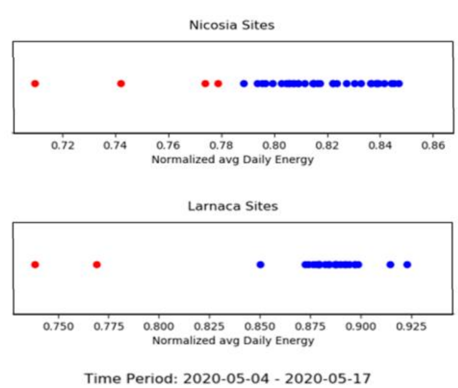
4. SOILAGE CLASSIFICATION

Our approach to the problem utilized multilevel unsupervised clustering, namely k-means. We first ran k-means on the GPS coordinates of the individual solar sites. The purpose of this first round of clustering was to group the sites by weather region to normalize for weather conditions. We continued to increase the number of clusters until we felt this was achieved. Our final "optimal" clustering is shown in figure 3(a). This decision was made manually after observing the decision boundaries between clusters. Our goal was to isolate the two main metropolitan areas on Cyprus - Nicosia and Larnaca.

Next, we calculated the total energy produced each day for every site. We normalized each site's values by dividing each by the site's max energy ever produced. We then had a weather normalized, magnitude normalized data set on which we performed k-means again with two clusters, shown in figure 3(b). The sites belonging to the cluster with the larger centroid we labeled clean and those belonging to the other we labeled dirty. This second level of clustering could be performed on a



(a) First, geographical clustering using k-means algorithm



(b) Second region specific normalized performance clustering

Figure 3: A nested two-step k-means clustering for soilage classification

per-day basis, or by averaging energy values over a period of time.

As this clustered soilage information is calculated daily, we can construct a heat map to illustrate the soilage patterns over time by site. Data from the sites in Larnaca is shown classified at both the daily level (a) and the bi-weekly level over the course of a year (b) in figure 4. Clean sites are marked in green while soiled or dirty sites are shown in red.

5. RESULTS

Evaluating our soilage classifiers was difficult, as true labels were not known. Despite this, by working in partnership with the researchers in Cyprus who provided the dataset, we were able to close the loop and get feedback on some sites, though only for a single day in time as our colleagues physically visited the sites and talked to the owners.

Despite this limitation, the feedback we received suggests that this second approach to soilage classification was effective. Of the two clean and two dirty sites in each of Nicosia and Larnaca that were evaluated, feet on the ground feedback indicated that the sites we classified as clean were clean, and those classified as dirty were either dirty or unexpectedly shaded by trees or buildings. This was unexpected as we had originally been told that there were no shaded sites. This confirmation feedback indicates that this soilage classification approach has a great amount of potential.

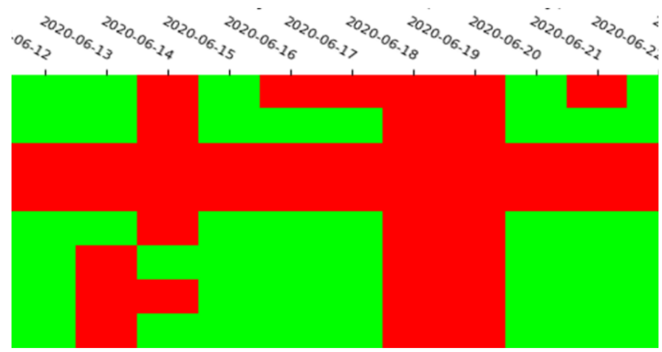
6. CONCLUSION

In conclusion, this study collected and utilized a new solar energy data set to test and develop new machine learning techniques for identifying soilage in household solar installations. We ran into issues with our initial method, which looked for evidence of degradation over time due to soiling by looking at daily power curves, but it is an interesting area for future research. Our second approach yielded promising results which were verified by our colleagues on the group in Cyprus. After normalizing for both magnitude and weather effects, we used 2-centroid k-means clustering to perform daily soiled/non-soiled classification. By plotting these classifications over time as a heatmap, we then identified the worst offenders and were able to verify in the field that these sites were in fact dirty. We believe that with further work this system could be developed into an autonomous system which alerts homeowners when their system has become too dirty.

In addition to the papers listed above, we were strongly influenced by the following papers on solar panel soilage and solar fault classification: [3], [4], [13]–[22], [5], [23]–[26], [6]–[12].

7. ACKNOWLEDGEMENTS

This work was supported in part by NSF Grant No. 1854273. The work was also supported in part by the ASU SenSIP Center. Special thanks to Sunil Rao for his help in understanding the solar aspects of this project.



(a) Daily soilage classification

(b) Bi-weekly soilage classification over the course of one year

Figure 4: Heatmap view of soilage classification. Each row is a single site and each column represents either one day (a) or two weeks (b). Classified soiled days are red, clean days are green.

8. REFERENCES

[1] S. Gambone, "Solar Panel Degradation and The Lifespan of Solar Panels," *Paradise Energy Solutions*. [Online]. Available: https://www.paradisesolarenergy.com/blog/solar-panel-degradation-and-the-lifespan-of-solar-panels.

[2] D. C. Jordan and S. R. Kurtz, "Photovoltaic degradation rates - An Analytical Review," *Prog. Photovoltaics Res. Appl.*, vol. 21, no. 1, pp. 12–29, 2013.

[3] L. L. Jiang and D. L. Maskell, "Automatic fault detection and diagnosis for photovoltaic systems using combined artificial neural network and analytical based methods," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2015-Septe, no. March, 2015.

[4] L. Bonsignore, M. Davarifar, A. Rabhi, G. M. Tina, and A. Elhajjaji, "Neuro-Fuzzy fault detection method for photovoltaic systems," *Energy Procedia*, vol. 62, pp. 431–441, 2014.

[5] H. Braun *et al.*, *Signal processing for solar array monitoring, fault detection, and optimization*, vol. 4. 2012.

[6] S. Rao *et al.*, "An 18 kW solar array research facility for fault detection experiments," *Proc. 18th Mediterr. Electrotech. Conf. Intell. Effic. Technol. Serv. Citizen, MELECON 2016*, no. April, pp. 18–20, 2016.

[7] D. Riley and J. Johnson, "Photovoltaic prognostics and heath management using learning algorithms," *Conf. Rec. IEEE Photovolt. Spec. Conf.*, pp. 1535–1539, 2012.

[8] T. R. S. Martz-oberlander, "The Dirt on Solar Energy : A study of Dutch solar panel efciency losses from soiling," no. May, 2017.

[9] T. Alquthami and K. Menoufi, "Soiling of photovoltaic modules: Comparing between two distinct locations within the framework of developing the photovoltaic soiling index (PVSI)," *Sustain.*, vol. 11, no. 17, 2019.

[10] G. P. Smestad *et al.*, "Modelling photovoltaic soiling losses through optical characterization," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020.

[11] S. Mehta, A. P. Azad, S. A. Chemmengath, V. Raykar, and S. Kalyanaraman, "DeepSolarEye: Power Loss Prediction and Weakly Supervised Soiling Localization via Fully Convolutional Networks for Solar Panels," *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018*, vol. 2018-Janua, pp. 333–342, 2018.

[12] K. Niazi, W. Akhtar, H. A. Khan, S. Sohaib, and A. K. Nasir, "Binary Classification of Defective Solar PV Modules Using Thermography," *WCPEC*, no. 1, pp. 753–757, 2018.

[13] X. Gao, L. Golab, and S. Keshav, "What's wrong with my solar panels: A data-driven approach," *CEUR Workshop Proc.*, vol. 1330, no. c, pp. 86–93, 2015.

[14] D. H. W. Li and S. Lou, "Review of solar irradiance and daylight illuminance modeling and sky classification," *Renew. Energy*, vol. 126, pp. 445–453, 2018.

[15] M. R. Maghami, H. Hizam, C. Gomes, M. A. Radzi, M. I. Rezadad, and S. Hajighorbani, "Power loss due to soiling on solar panel: A review," *Renew. Sustain. Energy Rev.*, vol. 59, pp. 1307–1316, 2016.

[16] S. K. Firth, K. J. Lomas, and S. J. Rees, "A simple model of PV system performance and its use in fault detection," *Sol. Energy*, vol. 84, no. 4, pp. 624–635, 2010.

[17] S. Daliento *et al.*, "Monitoring, diagnosis, and power forecasting for photovoltaic fields: A review," *Int. J. Photoenergy*, vol. 2017, 2017.

[18] L. S. Bruckman *et al.*, "Statistical and domain analytics applied to PV module lifetime and degradation science," *IEEE Access*, vol. 1, pp. 384–403, 2013.

[19] M. G. Deceglie, L. Micheli, and M. Muller, "Quantifying Year-to-Year Variations in Solar Panel Soiling from PV Energy-Production Data," pp. 2804–2807, 2018.

[20] R. R. Cordero *et al.*, "Effects of soiling on photovoltaic (PV) modules in the Atacama Desert," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, 2018.

[21] S. Rao, G. Muniraju, C. Tepedelenlioglu, S. Member, and A. S. Fellow, "Dropout and Pruned Neural Networks for Fault Classification in Photovoltaic Arrays," no. 1646542, pp. 1–3.

[22] S. Katoch *et al.*, "Shading prediction, fault detection, and consensus estimation for solar array control," *Proc. - 2018 IEEE Ind. Cyber-Physical Syst. ICPS 2018*, pp. 217–222, 2018.

[23] J. Fan, S. Rao, G. Muniraju, C. Tepedelenlioglu, and A. Spanias, "Fault Classification in Photovoltaic Arrays via Graph Signal Processing," in *ICPS*, Tampere, Finland, Jun. 2020.

[24] M. H. Ali, A. Rabhi, A. El Hajjaji, and G. M. Tina, "Real Time Fault Detection in Photovoltaic Systems," *Energy Procedia*, vol. 111, no. September 2016, pp. 914–923, 2017.

[25] Y. Zhao, F. Balboni, T. Arnaud, J. Mosesian, R. Ball, and B. Lehman, "Fault experiments in a commercial-scale PV laboratory and fault detection using local outlier factor," *2014 IEEE 40th Photovolt. Spec. Conf. PVSC 2014*, no. December 2015, pp. 3398–3403, 2014.

[26] Y. Zhao, B. Lehman, R. Ball, J. Mosesian, and J. F. De Palma, "Outlier detection rules for fault detection in solar photovoltaic arrays," *Conf. Proc. - IEEE Appl. Power Electron. Conf. Expo. - APEC*, no. March, pp. 2913–2920, 2013.