Estimation of Photovoltaic System Location and Orientation from Power Signals

Alejandro Londono-Hurtado
Applied Energy Division
SLAC National Accelerator Laboratory
Menlo Park, CA, USA

Bennet Meyers
Applied Energy Division
SLAC National Accelerator Laboratory
Menlo Park, CA, USA

Elpiniki Apostolaki
Applied Energy Division
SLAC National Accelerator Laboratory
Menlo Park, CA, USA

Robert Flottemesch
Analytics & Technology
Brookfield Renewable U.S.
Baltimore, MD, USA

***Abstract*—We present a methodology and associated software implementation for estimating the latitude, longitude, tilt, and azimuth of a fixed-tilt photovoltaic (PV) array from a measured power signal. Our major contribution to this line of research is developing a robust estimation procedure for the local sunrise and sunset time from the power data and creating reusable, open-source code implementations. We validated the proposed algorithms on a labeled data set comprised of a fleet of more than 4400 inverter power measurements at 400 unique sites.**

***Index Terms*—photovoltaic systems, parameter estimation, systems modeling, optimization, statistical learning.**

I. Introduction

Fleet-scale digital management of photovoltaic (PV) systems requires the development of automated algorithms for validating data, detecting operational issues, and analyzing system performance. System configuration information (sometimes called, "system metadata") are often associated with the recorded time-series. These descriptive metadata are used in performance analysis processes to build performance indices [1], among other uses. Validating this metadata and estimating it when it is not available is important for digital operations and maintenance at scale.

We present a methodology for estimating the location and orientation of a fixed-tilt solar array, based only on the measured power signal. First, we present a novel statistical learning approach to estimating sunrise and sunset times from historical power data, and show that this is the fundamental task required to correctly estimate array latitude and longitude. Second, we develop a simple statistical learning algorithm to replace a complicated site model for the task of normalizing the data to estimate array tilt and azimuth. We implement the entire procedure in reusable, open-source Python code, available here: https://github.com/slacgismo/pv-system-profiler.

II. Background

Previous work on this topic has typically employed simulations of PV system output, requiring partial system metadata including panel technology, system size, and wiring configuration. Williams *et al.* proposed a methodology for estimating PV array location and orientation, using a simulation-based approach for the location estimation, and an astronomical and simulation-based approach for the orientation estimation [2]. This methodology allows to estimate either location or orientation provided that one of these two is known *a priori*, but does not allow to estimate all parameters at once. Haghdadi *et al.* developed a methodology to estimate a given system's parameters based on the PV system's output power [3]. In their work, the authors employ the measured PV data to extract sunrise, sunset and noon to then use the Equation of Time (EOT) to estimate longitude. Then, they proceed to fit the measured PV output power to obtain a time series of clear sky data. This clear sky model is combined with the Perez model [4] and the PVwatt model [5] to estimate latitude, tilt and azimuth. The code used in the validation of [3] has not been made public.

This work is focused on developing a methodology for estimating latitude, longitude, tilt and azimuth form a measured power signal. Our contributions include: a) A novel statistical learning approach for the sunrise and sunset estimation with improved accuracy, b) A robust methodology for the estimation of latitude, longitude, tilt, and azimuth based solely on the PV array power signal, and c) The development of open-source algorithms made available to the public.

III. Methods

In this section, we discuss the methods used to estimate longitude, latitude, tilt, and azimuth. Critical to all of these is first estimating sunrise and sunset times, which we discuss first.

A. Estimation of Sunrise and Sunset

The power signal from an array is linked to its geographic location through the local solar noon and number of daylight hours, as described in more detail in §-III-B and §-III-C. Both these quantities are directly linked to the local sunrise and sunset times. The local solar noon time is the average of the

Supported by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 34911 ("PVInsight")

sunrise and sunset times, and the number of daylight hours is the difference between the sunset and sunrise times.

Previous work, such as [3], used a simple thresholding scheme to detect the "system on" condition. We find that significant improvements can be made by a improving the accuracy and robustness of this estimation. We leverage two observations regarding solar PV data sets. One, different local conditions—including weather patterns, array orientation, and inverter technology—cause significant differences in the "low-power" behavior of different PV systems. Therefore, the optimal threshold differs across data sets and must be selected based on the observed data. Second, there is significant information to leverage regarding how the estimates of sunrise and sunset should change day to day. Therefore, a statistical estimation procedure should be used that leverages the fact that sunrise and sunset times change slowly day-to-day and repeat on a yearly period.

First, we define the "system on" threshold to be τ, a value between 0 and 1, corresponding to the minimum and maximum recorded power values respectively. Given a value for τ, two daily signals are generated from the data, corresponding to the time of the first power value above τ and the time of the last power value, called the "measurements" of sunrise and sunset, represented as the green dots in Figure 1. We then apply a *signal decomposition* framework [6] to these measurements, that models the measurements as a sum of two unobserved components: a smooth and seasonally periodic component and a non-Gaussian, asymmetric noise component. The signal decomposition framework implements a randomized holdout validation scheme [7] to allow for model selection. So, we treat τ as a model hyperparameter and use holdout validation to find the smallest possible value of τ that approximately minimizes the holdout error in the signal decomposition model. Having optimized the threshold parameter τ, we solve the signal decomposition problem one last time without holding out any data, generating the "estimates" of the sunrise and sunset times, represented by the blue dashed lines in Figure 1. By the constraints imposed in the signal decomposition problem, this estimate is guaranteed to be smoothly changing from day to day and periodic on a yearly time scale. An example of this estimation process, after τ has been optimized, is shown in Figure 1. The red lines represented the true values for this particular location, which we wish to estimate. Our proposed methodology provides the estimates in blue.

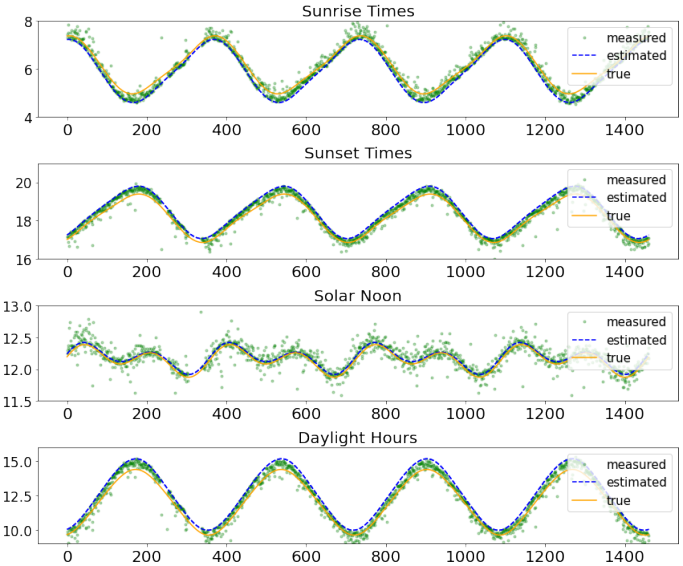

Fig. 1. The estimation of sunrise and sunset times, and the derived values of solar noon and daylight hours, using a *signal decomposition* methodology. These plots show the measurements (green dots) and estimates (blue dashed lines) after the selection of the "turn-on" threshold, τ, for an example system. The x-axes of the plots is the day index and y-axes are in hours. The green dots are the measured signal that is the input to the signal decomposition problem. The blue dashed line is the smooth, seasonal component from the signal model. Particularly in the bottom two plots, note how much closer the blue dashed line is to the true values in red, as compared to the green dots.

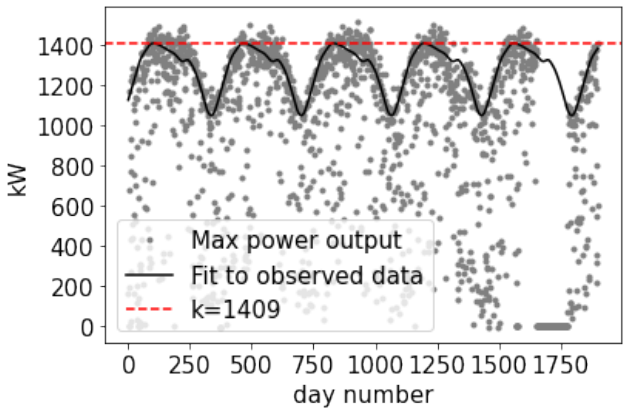

Fig. 2. Visualization of the method for estimating k, the parameter that relates measured power and angle of incidence. The black line is estimated using signal decomposition.

B. Longitude Estimation

The longitude of a system on a given day is related to the difference between solar time and standard time by

$$\text{solar time} - \text{standard time} = 4\,(L_{ST} - L_{LOC}) + E, \quad (1)$$

where L_{ST} is the standard meridian, L_{LOC} is the longitude and parameter E is the equation of time. The equation of time is related to parameter B by

$$E = 229.2(0.000075 + 0.001868\cos B - 0.032077\sin B - 0.014615\cos 2B - 0.04089\sin 2B), \quad (2)$$

where B is defined as

$$B = \frac{n-1}{360-365}, \quad (3)$$

where n is the day of the year [8]. We use (1) to solve for the longitude in terms of the daily solar noon. Given an estimate of solar noon on each day, provided by the method described in §-III-A, an estimate of the system longitude is generated for each day. The final estimate is taken as the median of the daily longitude estimates.

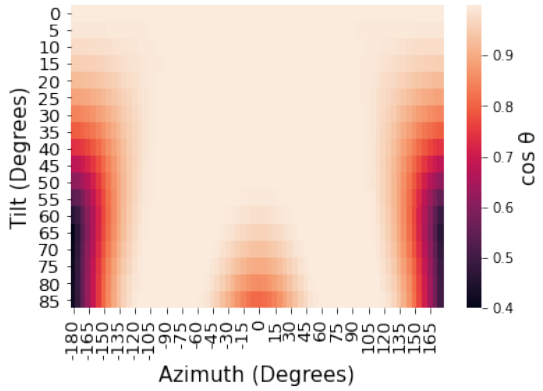

Fig. 3. Maximum value pf cosine of angle of incidence ($\cos\theta$) as a function of tilt and azimuth for a PV system located in Miami, FL.

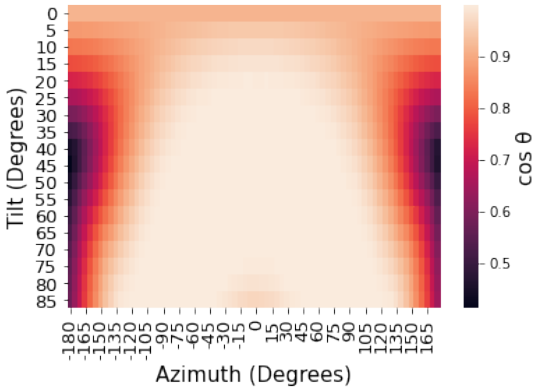

Fig. 4. Maximum value pf cosine of angle of incidence ($\cos\theta$) as a function of tilt and azimuth for a PV system located in Olympia, WA.

C. Latitude Estimation

The latitude of a system on a given day is related to the number of daylight hours and the declination angle by

$$N = \frac{2}{15}\cos^{-1}\left(-\tan\phi\tan\delta\right), \tag{4}$$

where N is the number of daylight hours, ϕ is the latitude, and δ is the declination angle [8]. The declination angle is a function of the day of the year [9], [10] and is therefore known. Similar to how longitude is estimated, As estimate of N is generated for each day in the data set according to the procedure described in §-III-A. The final estimate of the system's latitude is taken as the median of the daily latitude estimates.

Latitude is also a parameter in (5), suggesting the possibility of estimating that parameter in parallel along with tilt and azimuth. However, we find that this equation is underdetermined for three free parameters. By exploiting the specific relationship between the number of daylight hours and latitude given in (4), we are able to decouple the estimation of latitude from the orientation parameter and remove a degree of freedom from (5).

D. Tilt and Azimuth Estimation

The angle of incidence of a system is related to its latitude, tilt, azimuth, declination angle, and hour angle by

$$\begin{aligned}\cos\theta = \sin\delta\sin\phi\cos\beta - \sin\delta\cos\phi\sin\beta\cos\gamma + \\ \cos\delta\cos\phi\cos\beta\cos\omega + \cos\delta\sin\phi\sin\beta\cos\gamma\cos\omega + \\ \cos\delta\sin\beta\sin\gamma\sin\omega,\end{aligned} \tag{5}$$

where θ is the angle of incidence, β is the tilt, γ is the azimuth, ϕ is the latitude, and ω is the hour angle (defined as $15°\times$ number of hours before/after noon) [8]. The unknowns in this equation are therefore θ, β and γ. We provide a method for estimating $\cos(\theta)$ from the observed power, then allowing us to estimate β and γ through non-linear least squares.

To estimate $\cos\theta$ from the measured power data, we assume a very simple model relating angle-of-incidence and power

$$p = k\cos\theta, \tag{6}$$

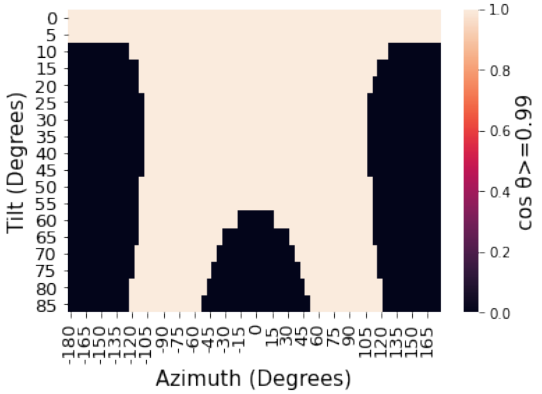

Fig. 5. Cosine of angle of incidence, ($\cos\theta$),=1 as a function of tilt and azimuth for a PV system located in Miami, FL.

where p is the system power and k is a parameter we'd like to estimate. Given an estimate for k, we can plug (6) into (5), and solve for the remaining parameters. We estimate k as follows:

1) calculate the daily maximum power output
2) utilize signal decomposition [6] to generate a smoothed, robust estimate of daily maximum power
3) the maximum value of the smoothed estimate is taken to be k

This processes is visualized for a representative site in Figure 2.

The explicit assumption in this process is that the angle-of-incidence is equal to zero ($\cos\theta = 1$) at least one time over the course of an entire year. This is a reasonable assumption for many PV systems located in North America. To support this assertion, we show here an analysis of two representative locations, Miami, FL and Olympia, WA. Figures 3 and 4 show the maximum value of $\cos\theta$ for various tilts and azimuths in Miami and Olympia respectively. In Figures 5 and 6, the tilts and azimuths corresponding to experiencing $\cos\theta \geq 0.99$ are marked in white, representing the space of tilts and azimuths at this location for which the assumption is valid.

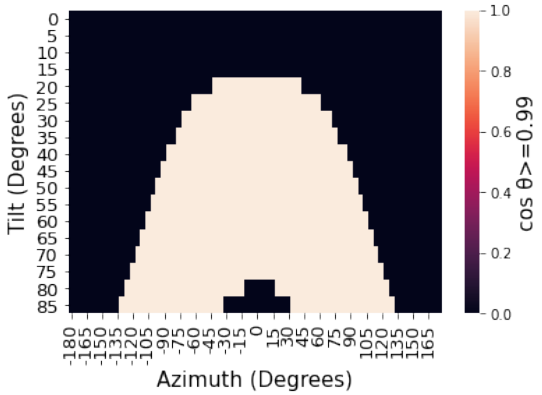

Fig. 6. Cosine of angle of incidence, ($\cos\theta$),=1 as a function of tilt and azimuth for a PV system located in Olympia, WA.

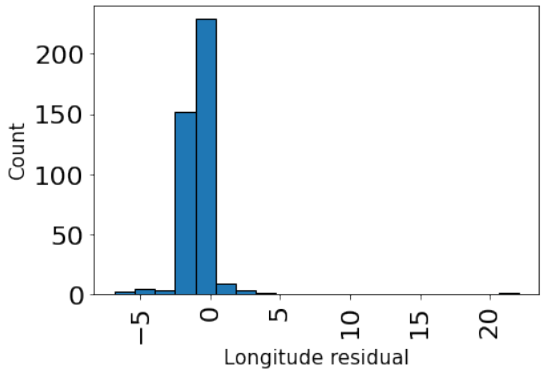

Fig. 7. Histogram of longitude residual for a sample of 407 system groups.

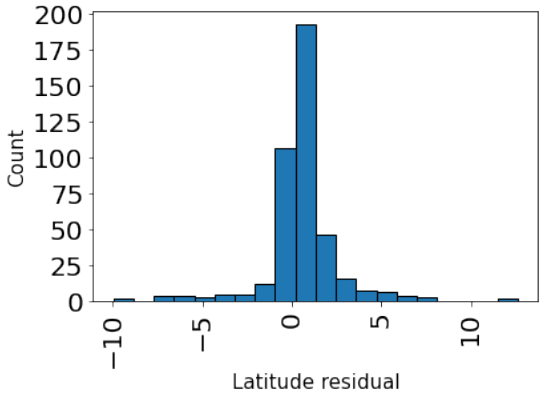

Fig. 8. Histogram of latitude residual for a sample of 407 system groups.

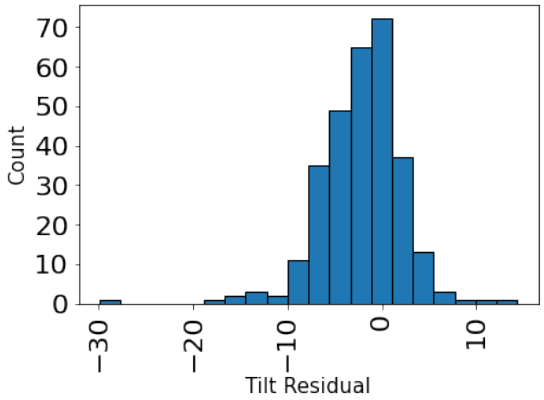

Fig. 9. Histogram of tilt residual for a sample of 296 system groups.

Finally, the parameters β and γ are estimated by applying non-linear least-squares to (5) and (6), using the least-squares function in Scipy's Optimize package [11].

IV. Results

The methodology presented in this work has been validated by performing a fleet scale analysis of currently installed PV systems. In total, we have a data set that contains 4380 unique inverter power signals. Most of these signals have ground truth values for location and about half of them have ground truth for orientation. Many of these inverters are part of larger homogeneous system groups, with nearly identical locations and configurations. Therefore, their signals are highly correlated. In this study, we analyze the latitude, longitude, tilt, and azimuth for each inverter signal. Then, the estimated parameters are averaged across the system group corresponding to the same physical location, and the final error statistics are given relative to the number of independent systems groups, resulting in 407 system groups for location estimation and 296 system groups for orientation estimation.

A. Longitude Estimation

Figure 7 presents a histogram of the longitude residual for the 407 system groups. It is observed that the majority of the values are distributed around 0° with some outliers. The Root Mean Square (RMS) error for the estimation is 1.73°. The outlier system group in the histogram with an RMS error higher than 20° corresponds to a set of inverters whose power signals are missing multiple months of data at different times of the year. Overall, we have found that the estimator is very robust and is able to handle missing data without problems. However, if the power data is significantly degraded, the parameter estimation algorithm loses accuracy. We also found that the estimation error is highly dependent on the method used to estimate solar noon, with the method described in III-A yielding the best results.

B. Latitude Estimation

Figure 8 presents a histogram of the longitude residual for the 407 system groups. It is observed that the majority of the values are distributed around 0° with some outliers. RMS error for the estimation is 1.97°. We found that the estimation error is less dependent on the method used to estimate solar noon than for the case of longitude. Similar to longitude estimation, there are some outlier sites. These are sites with degraded power signals. Overall, even when including outlier sites, the algorithms presented in this work are able to predict location of a system within 2°. In general, the validation presented in

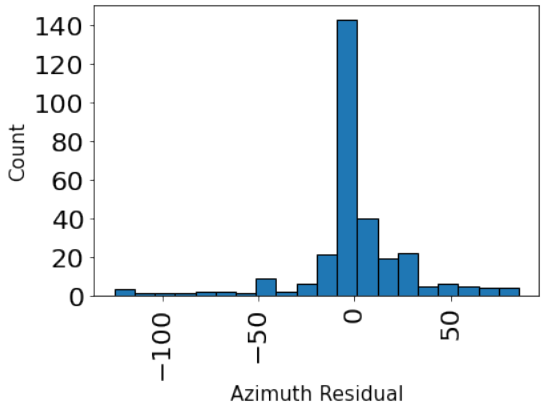

Fig. 10. Histogram of azimuth residual for a sample of 296 system groups.

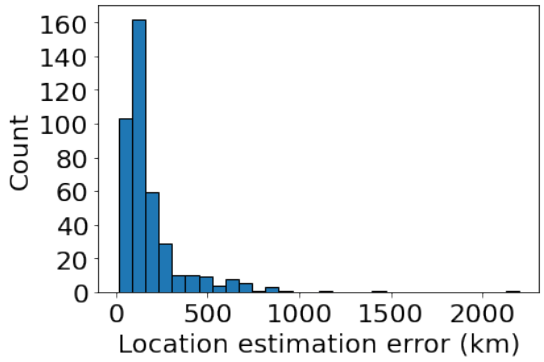

Fig. 11. Histogram of location estimation error for a sample of 407 system groups.

this work serves as confirmation that longitude and latitude can effectively be estimated from a system power signal.

C. Location Estimation Error

With the estimated values of longitude and latitude, we calculated the error associated with location parameter estimation. This error is calculated as the Euclidean distance between the estimated and the actual location of the system as provided by the ground truth values. Figure 11 presents a histogram of the location estimation error. This error is lower than 200 km for 74% of the sites. The outlier site groups in the histogram correspond to the ones previously identified as having degraded signals.

D. Tilt and Azimuth Estimation

The tilt and azimuth estimation algorithms were validated using the 296 system groups that had ground truth orientation values. While we designed the procedure to be used with any measured *power* signal, we found that in practice, much better results are obtained when using the *DC current* measurements instead of power, when available. There is a simple physical motivation for this observation. Our model relating the measured quantity to $\cos\theta$ (6) does not include a temperature correction because we do not assume access to module temperature data. DC maximum power point current

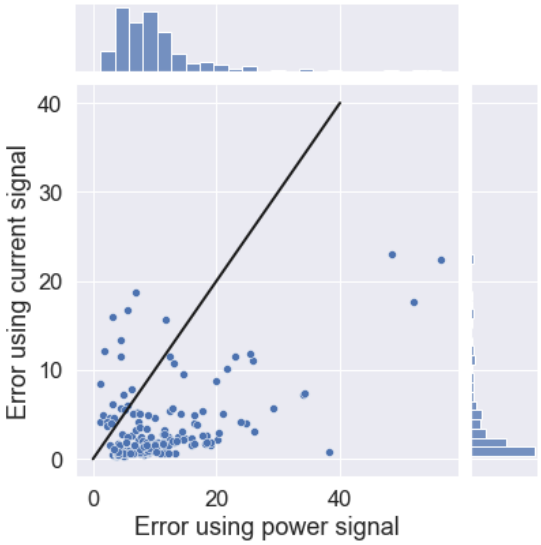

Fig. 12. Error of $\cos\theta$ estimated using DC power and AC current signals for 200 randomly selected systems. The line corresponding to $y = x$ is provided for visual reference.

has a much smaller temperature dependence than power, so (6) is a more accurate model of current than power.

Figure 12 presents the estimation error for $\cos\theta$, using DC current and AC power for 200 randomly selected systems. This error is calculated as the difference between $\cos\theta$ calculated analytically and its estimated value. Results reveal that using the DC current signal leads to a smaller error. For the 4400 inverter PV fleet used in this work, all sites had power signals and only 2% of the systems were missing a DC current signal.

We present the errors in tilt and azimuth estimation when utilizing DC current as the measurement of interest. Figures 9 and 10 present histograms of the tilt and azimuth residual, respectively. In the plots, the majority of the values are distributed around 0°, but the distribution is wider than for the case of longitude and latitude. The RMS errors for the estimation are 5° and 28° for tilt and azimuth, respectively. Similar to location parameter estimation, the outlier sites in the histogram are mostly systems that have degraded signals. A tilt RMS variation of 2° is within the range of adjustment and measurement limits for a PV system in the field. A tilt RMS error of 5° can be considered acceptable. The RMS error for azimuth is high. The histogram reveals that, although there is significant spread, the algorithm provides good estimations for the majority of the system groups. For the case of Azimuth estimation, the RMS error is less or equal to 10° for 76% of the system groups. Analysis of results suggests that azimuth estimation is the most sensitive to signal quality, while longitude, latitude and tilt are more permissive. As it will be discussed in the next section, accurate orientation estimation can be achieved provided a clean system power signal.

E. Algorithm Validation using NIST experimental PV systems

In order to further validate our work, we tested the estimation algorithms on 15 fully instrumented and PV systems

installed by NIST their Maryland campus [12], [13]. All these systems have good quality measurements both for DC current and AC power. All sites also provide ground truth information for longitude, latitude, tilt and azimuth. In similar fashion to the 4400 inverter fleet used in this work, we estimated latitude, longitude tilt and azimuth for these systems and calculated the RMS errors based on the ground truth values. The RMS estimation errors for these sites were of 1° for longitude and latitude and 3° for tilt and azimuth. Similar to the 4400 inverter PV fleet used in this work, more accurate estimates of tilt and azimuth were obtained using DC current than AC power signals. These results suggest that, for fairly clean data, the algorithm estimations are within a couple of degrees. The authors recommend the use of Solar Data Tools for identifying operational issues and dirty data [14].

V. Conclusions

We present the theory and implementation of estimating latitude, longitude, tilt, and azimuth from PV system operational data. In contrast to previous work, we assume minimal data and metadata requirements, utilizing only measured operational signals (power and/or current), and no information about system configuration or technology. Additionally, we have built a publicly available software implementation of the methods so that future researchers may do their own validation and build on this work [15].

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office (SETO) Award Number 34368.

References

[1] T. Townsend, C. Whitaker, B. Farmer, and H. Wenger, "A new performance index for PV system analysis," in *Proc. of 1994 IEEE 1st World Conf. on Photovoltaic Energy Conversion*, vol. 1. IEEE, 1994, pp. 1036–1039.

[2] M. K. Williams, S. L. Kerrigan, A. Thornton, and L. Energy, "Automatic detection of PV system configuration," in *World Renewable Energy Forum*, 2012.

[3] N. Haghdadi, J. Copper, A. Bruce, and I. MacGill, "A method to estimate the location and orientation of distributed photovoltaic systems from their generation output data," *Renewable Energy*, vol. 108, pp. 390–400, 2017.

[4] R. Perez, R. Seals, P. Ineichen, R. Stewart, and D. Menicucci, "A new simplified version of the Perez diffuse irradiance model for tilted surfaces," *Solar energy*, vol. 39, no. 3, pp. 221–231, 1987.

[5] B. Marion, "Comparison of predictive models for photovoltaic module performance," in *2008 33rd IEEE Photovoltaic Specialists Conference*, 2008, pp. 1–6.

[6] B. E. Meyers and S. Boyd, "Signal decomposition via distributed optimization," Ph.D. dissertation, Stanford University Department of Electrical Engineering, 2021, unpublished thesis.

[7] T. Hastie, R. Tibshirani, and J. Friedman, "Ch. 7: Model assessment and selection," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2013, pp. 219—-257.

[8] J. A. Duffie and W. A. Beckman, *Solar engineering of thermal processes*. John Wiley & Sons, 2013.

[9] P. I. Cooper, "The Absorption of Solar Radiation in Solar Stills," *Solar Energy*, vol. 12, no. 3, 1969.

[10] J. Spencer, "Fourier Series Representation of the Position of the Sun," *Search*, vol. 2, no. 5, pp. 162–172, 1971.

[11] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[12] M. Boyd, T. Chen, and B. Doughert, *NIST Campus Photovoltaic (PV) Arrays and Weather Station Data Sets*. National Institute of Standards and Technology, 2017, [Data set]. https://doi.org/10.18434/M3S67G.

[13] M. Boyd, "High-speed monitoring of multiple grid-connected photovoltaic array configurations," *NIST Technical Note 1896*, 2015, http://dx.doi.org/10.6028/NIST.TN.1896.

[14] B. E. Meyers, E. Apostolaki-Iosifidou, and L. T. Schelhas, "Solar data tools: Automatic solar data processing pipeline," in *47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020, pp. 655–656.

[15] "`pv-system-profiler` on GitHub." [Online]. Available: https://github.com/slacgismo/pv-system-profiler/tree/development