Automating Methods for Validating PV Plant Equipment Labels

Joseph Ranalli[1] and William B. Hobbs[2]

[1] Penn State Hazleton, Hazleton, PA 18202, USA
[2] Southern Company, Birmingham, AL, 35203, USA

***Abstract* — Operators of utility scale photovoltaic plants can face challenges in accurately matching operational data at various levels with their physical locations in the plant. This may have financial implications due to misdirected maintenance efforts. A method to predict equipment locations using time series data analysis was developed, but previously relied on some manual processing. This paper details efforts to automate the technique for improved efficiency and to facilitate operational deployment. The automated methodology relies on selecting ideal cloud motion vectors by targeting time periods based on their variability score metric. Ten vectors with well-spaced angular directions were selected after filtering based on quality control parameters. After the analysis methodology was run for an entire plant, mislabeled equipment was detected by implementing optimization code for solving the well-known Assignment Problem. Because predicted equipment positions have a circular dependence on their original expected positions, the process was applied iteratively as mislabeled entities were identified to converge to a final assignment of predicted position for each plant component. The results from this algorithmic methodology agree with the validated results from the original manual method. These improvements further the goal of offering the method as a ready-to-use tool for validating the physical locations of equipment within a utility-scale photovoltaic plant.**

I. Introduction

Operators of utility scale photovoltaic plants monitor the performance of the plant at various levels of equipment aggregation; besides output data from the entire plant, diagnostic generation data may also be monitored at the level of the strings, combiners or inverters. A single plant may comprise thousands of these equipment entities for which data is collected. Plant operators have expressed concerns with their ability to confidently ensure that all equipment labels in the data acquisition system match with their expected physical locations from plant design documents.

A lack of confidence in the remote sensing data may have real financial costs. Conducting physical repairs on a plant is time and labor intensive and requires extensive planning. When equipment problems are identified from the data, low confidence in equipment locations could lead to delays as maintenance teams are dispatched to the wrong part of the plant, or in the worst case, the true physical location of the problem could remain impossible to identify.

A field analysis method has previously been developed by the authors to validate expected plant equipment locations using operational data [1]. The method analyzes spatial variability in time series data for plant subcomponents to identify cloud motion over the plant. Cloud motion allows determination of the apparent relative positions of the plant components, which can be compared with their expected design locations. The method was first introduced in a paper that includes successful validation of location predictions made by the method against equipment mislabeling in an actual PV plant.

One potential shortcoming of the published work is that initial implementations of the method relied on some manual interpretation of data. In the initial study, time periods with suitable variability and cloud motion were identified by a manual process. Additionally, identification of potentially mislabeled components within the plant and assignment of their corrected locations was also based on manual inspection of their predicted relative positions. In this paper, we detail progress toward fully automating the technique so that it can be applied more generally and without manual intervention.

II. Methodology

A full description of the field analysis method is provided in the original paper [1], but a summary is reproduced here. To apply the method to compute the effective position of a single plant component (in this case, a combiner), P_1, we implement the following steps:

1. Find two time periods, A and B, with stationary, non-parallel cloud motion vectors (CMVs).
2. Compute the relative time delay between P_1 and neighboring combiners, P_i, during each CMV window.
3. Perform quality control and averaging across i to compute an averaged error vector representing the difference between the expected and apparent position of P_1 along each CMV axis.
4. Triangulate the apparent position, P'_1, in two-dimensional space from the pair of error vectors.
5. Repeat and average for multiple A & B pairs to reduce uncertainty, yielding the final predicted position.

In the initial study, after computing the predicted positions for each combiner according to these steps, the results were inspected and an attempt was made to reassign each combiner to the closest neighboring position. An example of this process from the initial publication is shown in Fig 1. It is important to note that in the initial study, these reconfigurations were based on manual assignment and interpretation, with a focus on validating the capability of the method.

As described, two steps within the initial application of the method required some manual intervention: 1) identification of time periods with suitable CMVs, and 2) reassignment of combiners based on the predictions. In this paper, we report on efforts to automate these two stages of processing.

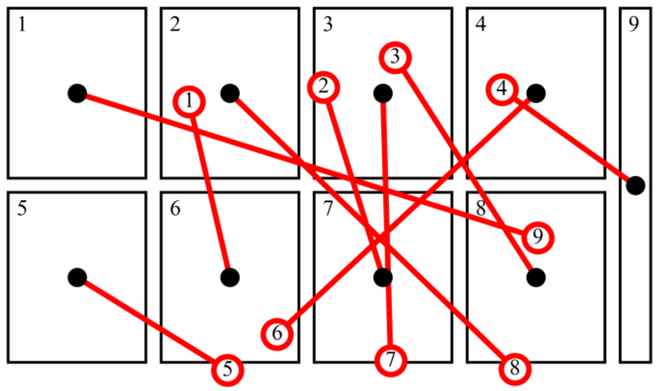


Fig. 1. Example of manual position reassignment for a single inverter. Circles show predicted positions, while numbers in the circles indicate manually reassigned combiner footprint. Copied from [1].

A. Cloud Motion Vector Identification

The methodology requires, at minimum, two time periods that experience cloud motion vectors that span the horizontal plane. Ideally, multiple such pairs would be identified to allow averaging that reduces the impacts of noise and uncertainty in the predictions. When approaching the task of applying the method to a plant, investigators are likely to initially be faced with the problem of too much data. For example, initial development of the method began with a full year of 10s resolution combiner current time series for the entire plant. Simply computing results for all possible time periods presents a computationally prohibitive task, so we aimed to develop an approach that can target suitable time periods that are likely to produce high quality CMV calculations and corresponding combiner position predictions.

Several considerations impact the choice of suitable CMVs. First, both the field analysis approach and the CMV identification method of Jamaly and Kleissl [2] are based upon computing relative delays between plant constituent elements. The ability to compute consistent time delays between combiners requires that the CMV be uniform and stationary across the plant during the time period (in this case, targeting whole hours). This requires that we identify hour periods during which clouds are present and during which the cloud motion is consistent.

Additionally, besides the mathematical requirement that the vectors span the plane, it is also desirable to constrain the relative angle of the CMVs to those with orientations that have sufficiently large perpendicular components. The triangulation of expected combiner position is performed by finding the intersection of the lines, perpendicular to the CMV axes, that pass through the corresponding error vector along each CMV axis. For pairs of axes that are nearly parallel or anti-parallel, small uncertainties in the error vector length may be amplified to large variations in the predicted combiner position. To reduce this effect, the original study globally required that CMV axes have an angular separation of at least 45 degrees (i.e. vectors are relatively angled from 45 degrees to 135 degrees). For the purposes of automating the process, it would be desirable to ensure that the CMVs calculated yield a large number of nearly-perpendicular pairs so that averaging can be performed to reduce error in the final position predictions.

We developed several steps as part of a process that were applied in this study to ensure that the most favorable CMVs were computed and chosen for further processing. Despite some degree of optimization that was performed on the codes being utilized, computational efficiency concerns still promote a focused approach. For the sample plant, calculation of the CMV for a 1-hour period takes around 5-10 seconds on a laptop computer. This would scale to around 12 hours of processing time for CMV's for a full year of data, a timespan that would be unacceptable for revision and iteration of the method.

To identify periods that were likely to have a consistent CMV, we first filtered to 1-hour periods with the highest values of the variability score (VS) [3]. CMVs were then computed for a subset of hours with the highest VS value. While a definitive size of this subset may depend on case-by-case details, choosing 50 was found to be sufficient and computationally tractable for this case. From those 50 time periods with computed CMVs, we then eliminated those of low-quality based on metrics computed during the CMV calculation. First, the Jamaly and Kleissl CMV method contains both pairwise and holistic quality control values, so we eliminated any CMVs that failed the overall quality control. We also eliminated those for which only a very small number of combiner pairs passed the method's pairwise quality control (using a threshold of a minimum of 100 pairs). This ensured that each retained CMV was based on a sufficiently large subset of the plant. Finally, we computed the correlation coefficient between the pairwise separation distance and time delay for all the combiner pairs that passed quality control. We excluded any CMVs for which this coefficient had a correlation coefficient value less than 0.8, as this would indicate that different combiner pairs experience inconsistent cloud speeds. For two different sample plants, this filtering from 50 CMVs resulted in 37 and 46 CMVs remaining.

The final step was to select a subset of 10 CMVs that represented a variety of motion directions to ensure that multiple perpendicular pairs were available. In order to do so, we formed an optimization problem, whereby we defined a set of equally spaced vectors with orientations ranging from 0 – 180 degrees. We rotated all CMVs into this same half-circle, motivated by the fact that both parallel and anti-parallel vectors are equivalently undesirable. Since the choice of zero degrees was arbitrary, we allowed the optimization to consider rotation of the equally spaced set by up to 18 degrees (i.e. the spacing between two optimally spaced vectors). We used an existing optimization code from the scipy library [4] to perform this calculation and select the 10 CMV vectors that most closely matched the best-case equally spaced set. A sample of the outcome of this process is shown in Fig. 2.

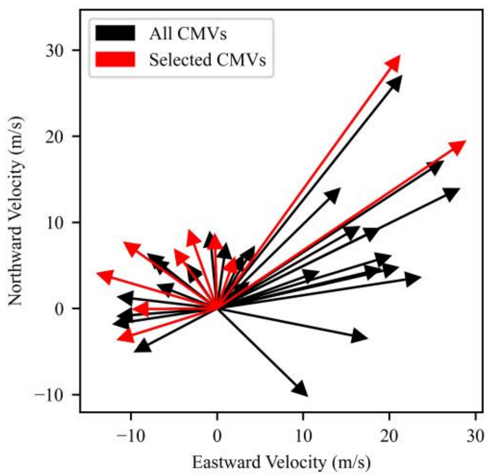


Fig. 2. Example of downselecting CMV pairs from an initial set of 50 hours. All valid pairs shown in black, and the chosen subset is shown in red. Vectors shown have their true direction, but the downselection process shifts all into the 0-180 degree range to eliminate both parallel and anti-parallel vectors.

In summary, the automation of the CMV identification used the following steps:

1. Compute VS for each combiner on an hourly basis.
2. Choose 50 hours with the highest median VS for the plant. Compute CMVs for those hours.
3. Downselect to CMVs that meet the following metrics:
    a. The CMV calculations pass holistic QC from the Jamaly and Kleissl method.
    b. At least 100 combiner pairs pass pairwise QC from the Jamaly and Kleissl method.
    c. The correlation coefficient between pairwise separation and time delay for valid combiner pairs exceeds 0.8.
4. Select approximately ten one-hour periods with CMVs representing a variety of orientations, using these steps:
    a. Temporarily rotate south-directed vectors by 180 such that all lie in quadrants I & II.
    b. Define an ideal, equally spaced subset of 10 target vectors
    c. Perform optimization to select the 10 CMVs that form the closest alignment to that set.

This yields a subset of 10 CMVs, from which valid pairings can be identified, allowing the subsequent steps of the field analysis methodology to be applied. The method will yield predictions of the effective, delay-based position of each combiner within the entire plant.

B. Reassigning Combiner Locations

After applying the field analysis method to determine predicted positions of each combiner, it is necessary to determine whether any mislabeling has occurred. In the test cases considered here, based on the physical configuration of the plant's combiner connections, it is unlikely for combiner labeling errors to extend beyond a shared inverter. This is because the location of these connections is easier to confirm. That is, combiner wire outputs physically converge at the location of the connection to a single inverter and it is unlikely that they could be incorrectly connected to an inverter at a different location. Thus, the method for automated identification developed here focused on labeling errors that occurred within a given inverter only. In principle, however, the processes developed should be applicable to cases with different limitations or assumptions about the occurrence of labeling errors.

The remapping computation was performed using a solver for the well-known Assignment Problem in optimization, as implemented in the *scipy* python package [4], [5]. The cost function for the solver was minimization of the total distance between the inferred combiner positions from the methodology and possible initial positions from the site plans. For each inverter, this resulted in a prediction of which combiner was most likely located in each location based on aggregate distances.

It is important to note that there is a circular dependency in the computation of the predicted combiner positions. Since all computations are based upon relative time series delay between a combiner and its neighbors, a cluster of mislabeled combiners could still appear to have "correct" relative time delays compared to each other, while being in error relative to the plant as a whole. To handle this type of issue, the descrambling was performed iteratively: relative positions for combiners within an inverter footprint are computed, which are used to generate a remapped set of combiner positions, which then are used to recompute the relative positions. This process is repeated until either the recomputed positions converge (i.e. do not result in a change of the map), until the predicted configuration repeats (indicating a loop), or until an iteration limit is reached. This ensures that the final predictions were updated to reflect the information gained during previous remapping iterations.

For each inverter in the plant, the automated steps used to compute the remapping for all combiners connected to that inverter were as follows:

1. Apply field analysis to compute the inferred combiner positions for all valid CMV pairs.

2. Average the positions over the CMV pairs to reduce error in the inferred position.
3. Perform optimization at the inverter level to assign each combiner to the best fit combiner position from the site plans.
4. Repeat steps 1 through 3 until the optimization does not result in a change to the assignments, or until 5 total iterations are performed.

The resulting mapping is recorded and assumed to reflect the corrected positions of each combiner.

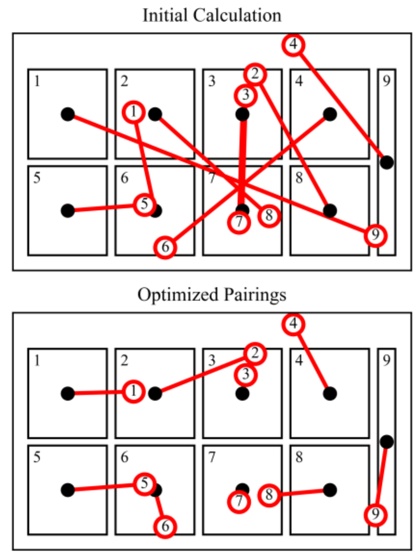
A full implementation of the automated formulation of the method, along with demonstration codes, is available in the open-source *solartoolbox* library available on GitHub [6].

III. RESULTS & DISCUSSION

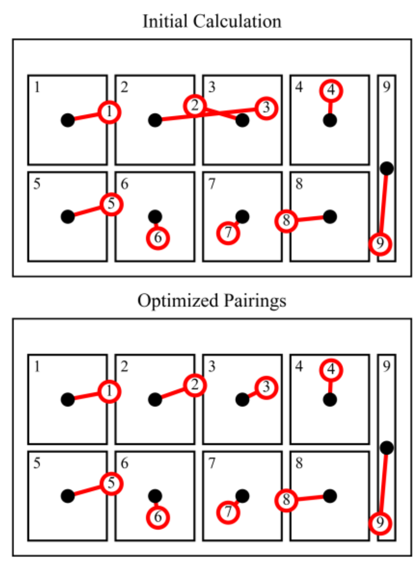
The field analysis methodology using these automated processing approaches was applied to data from two operational PV plants. The first plant was an approximately 20 MW (25 inverters, 221 combiners) plant, described in the original manuscript introducing the methodology. The second was a larger plant, approximately 30 MW in capacity, consisting of 48 inverters with 366 total combiners. The described automated workflow was applied to both plants on an inverter-by-inverter basis to yield predictions of corrected combiner positions within the plant.

A. Iterative reassignment

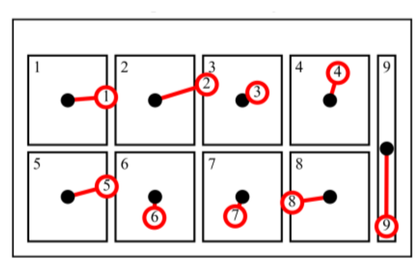
A sample of results from the iterative method for a single inverter that experienced mislabeling is shown in Fig. 3. As is evident, the initial calculations for the first iteration show a significant degree of disagreement between the combiner positions indicated in the design plans and those apparent from the relative time delay between the combiners. The descrambled (optimized) positions from iteration 1 were re-processed by the model, leading to changes in the predicted combiner positions, as reflected in the initial calculation for the second iteration. Due to these shifts, the second iteration's positions for combiners 2 and 3 were also remapped and an additional update was made. Finally, when optimizing the combiner mapping for the third iteration, no change in footprint assignment was observed, indicating that the re-mapping had converged. For the test cases plants, the most common result was no mislabeling, meaning only a single iteration was performed. For inverters where mislabeling was identified, convergence typically occurred in 2-3 iterations. A very small number of inverters failed to converge, which will be discussed subsequently in Section III.C.



(a) First iteration



(b) Second iteration



(c) Final mapping

Fig. 3. Three iterations of the reassignment output for one inverter of the sample plant. In the final mapping, the reassignment was complete because assignments were unchanged after additional assignment.

B. CMV Selection

In terms of evaluating the effectiveness of the CMV selection algorithm, our operational experience on one year of data for these two example plants indicates that an abundance of potential CMV periods were available and that those periods represented a sufficiently diverse set of cloud directions from which to form many suitable CMV pairs. In fact, this plethora of data is what lead to the need for the downselection methodology in the first place. Fig. 2 shows the diversity of time windows for which valid CMVs were computed and indicates the sufficiency of directional variation.

However, it is feasible to consider a case in which, for example, all high quality CMVs returned by the selection algorithm are nearly parallel or anti-parallel, resulting in a lack of access to sufficiently diverse CMVs to form perpendicular pairs. In such a case, we would recommend first attempting to tune the filtering parameters described in Section II.A, or if available, looking at a longer time period (i.e. more than one year of data). Additionally, it may be possible to produce optimal results by selecting different CMVs for different parts of the plant (e.g. due to localized quality control issues during different time windows), but this was not tested for the plants in question, due to the abundance of suitable data. It remains true that suitable CMV identification remains a limitation of the method. Sites with insufficient cloud cover to produce variability from which CMV can be identified, or with a singular CMV direction (i.e. with nearly parallel CMVs) are unsuitable for use with this method.

C. Plant-wide reassignments

We applied the methodology to reassign positions for two complete test plants and the results for reassignment for each plant are shown in Figs. 4 and 5. Three subsections of the first plant were validated in the initial study on the method [1], and results of the automated process agree with those validations. Further, on both plants, the majority of predicted combiner positions coincide with their expected positions from the design drawings. This is consistent with the assumption that labelling errors in the plant are isolated and uncommon. When predictions indicate a need for reassignment, most final reassigned positions fell approximately within inverter footprints throughout the plant during and after convergence of the calculation. These results indicate the overall suitability of the automation approach for the field analysis technique.

With regard to the descrambling methodology, we observed that the method appears to produce results that are reasonably consistent with operational experience on the plants with a few exceptions. Across the two plants, 3 out of 73 total inverters resulted in predictions that were unable to converge after continued iteration, all occurring in the first plant (shown in Fig. 4). These instances primarily led to cases where the overall descrambling of the positions was effective, but a subset of two or three combiners led to a closed loop cycle of predictions as the iterations were repeated. Further work at identifying uncertainty in the calculations will hopefully address that type of situation, but absent further study, users may consider inspecting the results on an individual CMV-pair basis (or running on an alternate CMV subset) to attempt to identify whether patterns in the predictions lead to the method's confusion. Even in the worst case, the method still provides a greater degree of clarity as to which parts of the plant experience ambiguity in apparent combiner locations relative to the expected design plans.

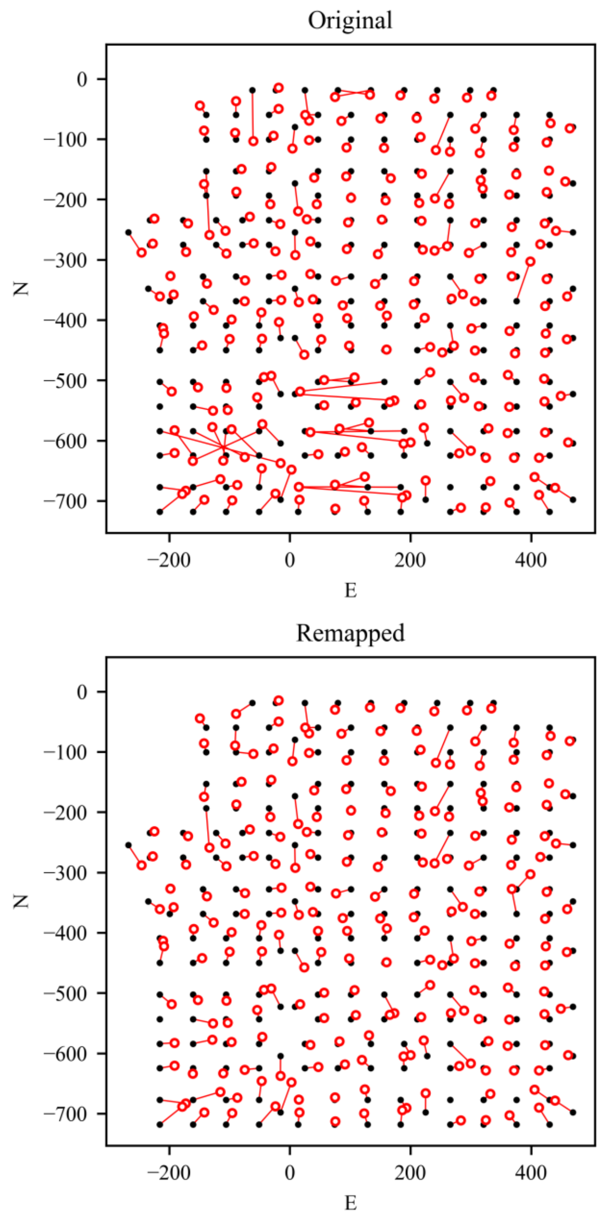


Fig. 4. Full reassignment for first sample plant. Upper shows results prior to remapping while lower shows remapped results.

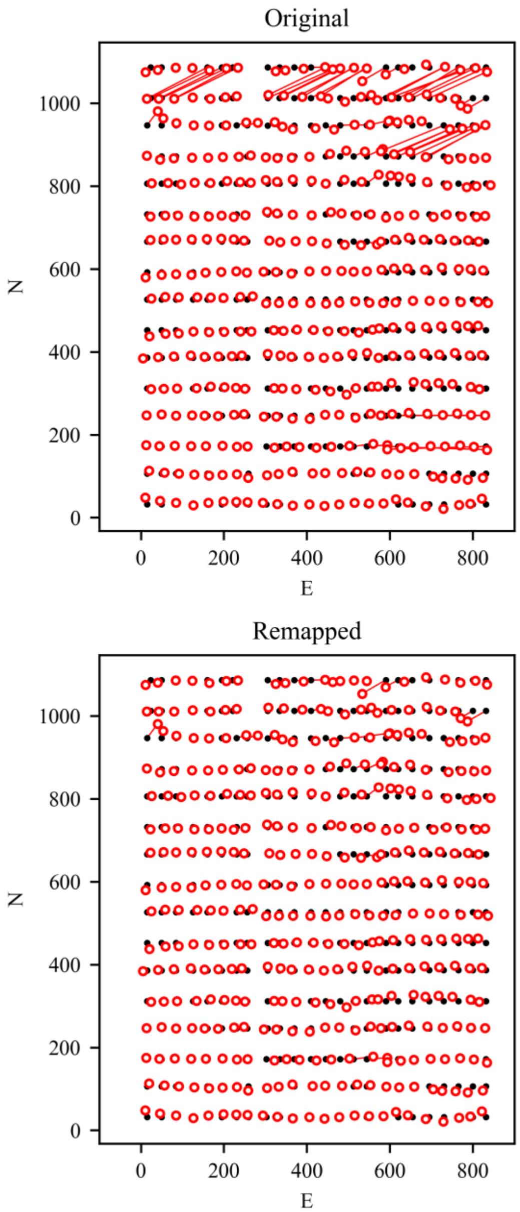


Fig. 6. Full reassignment for second sample plant. Upper shows results prior to remapping while lower shows remapped results.

III. CONCLUSION

The approaches described in this paper significantly improve the ease of implementing the field analysis methodology for validating plant layout. By eliminating the need for manual selection of cloud motion vectors and position reassignments, potential users could directly implement the methodology to validate a PV plant's as-built connections and improve confidence in the operational data. Case study implementation of the methodology on two operational plants demonstrates the functionality and capability of the approach, though a few areas for continued study and improvement were identified. Further study on the methodology is warranted to help address questions about the uncertainty of the position predictions, and help users resolve edge cases where the method fails to converge, or where issues occur due to poor underlying data quality. Ultimately, we believe the methodology as implemented here may be of significant interest to plant operators to help them identify locations where data confidence may be questionable.

REFERENCES

[1] Joseph Ranalli and William B. Hobbs, "PV Plant Equipment Labels and Layouts can be Validated by Analyzing Cloud Motion in Existing Plant Measurements," *J. Photovolt.*, Under Review.

[2] M. Jamaly and J. Kleissl, "Robust cloud motion estimation by spatio-temporal correlation analysis of irradiance data," *Sol. Energy*, vol. 159, pp. 306–317, Jan. 2018, doi: 10.1016/j.solener.2017.10.075.

[3] M. Lave, M. J. Reno, and R. J. Broderick, "Characterizing local high-frequency solar variability and its impact to distribution studies," *Sol. Energy*, vol. 118, pp. 327–337, 2015.

[4] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[5] David F. Crouse, "On implementing 2D rectangular assignment algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 4, pp. 1679–1696, Aug. 2016, doi: 10.1109/TAES.2016.140952.

[6] J. Ranalli, "solartoolbox." Nov. 04, 2023. Accessed: Jan. 18, 2024. [Online]. Available: https://github.com/jranalli/solartoolbox