PV Plant Equipment Labels and Layouts Can Be Validated by Analyzing Cloud Motion in Existing Plant Measurements

Joseph Ranalli, *Member, IEEE*, and William B. Hobbs, *Senior Member, IEEE*

***Abstract*—Large-scale photovoltaic plants collect monitoring and operational data at various spatial scales within the plant (e.g., strings, combiners, and inverters). Manual validation of the spatial position of these plant segments relative to the plant design requires on-site observations that may be prohibitively costly or labor intensive. This article presents a methodology for validating plant segment position based on operational data from the plant. By observing the delay between segment responses to cloud motion, predictions of their relative positions within the plant can be made. The method was demonstrated on combiner-level data from a 20-MW, operational photovoltaic plant in the United States. Several instances of apparently mislabeled combiners were identified from the analysis. A partial validation of 20 combiners was conducted by inspecting the plant, with results showing complete agreement between observation and predictions. Predictions derived from this methodology can serve as the basis for further plant inspection and corrective maintenance.**

***Index Terms*—Cloud advection model, irradiance variability, photovoltaics (PV), solar energy, transfer function.**

I. Introduction

CONTINUED impacts of climate change [1] provide the motivation for a global transition to carbon-free sources of electricity generation. Renewable sources of generation, including wind and solar photovoltaics (PV), are expected to represent a major share of a future clean energy portfolio. Least cost scenario analyses indicate that wind and solar should provide 60%–80% of electricity generation in the United States to meet the goal of transitioning to a carbon-free electricity system by 2035 [2]. Solar alone makes up around 25% of the overall generation in these scenarios. Meeting these targets would require deployment at rates as high as four times current levels [2].

Achieving that level of deployment will require continued growth in installation and operation of large-scale PV facilities. Operators of large-scale PV plants conduct extensive monitoring of plant operation and performance. These data are used to observe equipment failures, damage or off-design operation that can be used to guide operations and maintenance activities. In addition to monitoring at the overall plant level, data are typically collected at various segmented levels within the plant, such as strings, combiners, or inverters. This degree of monitoring leads to a significant number of labeled data collection entities. For example, a 20-MW plant may consist of thousands of strings, hundreds of combiners, and dozens of inverters. The potential for mistakes during the construction and commissioning process in labeling these entities reduces the confidence in plant operational data and represents a financial risk to plant operators. Due to the potential occurrence of mislabeled entities, monitoring that indicates malfunction in a section of the plant could lead to workers being deployed in the wrong region of the plant, or even an inability to determine where remediation is needed. Unfortunately, validating the labeling based on inspection or manual testing represents a potentially laborious and time consuming task that may itself be cost-prohibitive.

As an alternative, this article introduces a method to remotely validate the position of plant segments by observing cloud motion over the plant, derived from time-resolved generation data from the individual segments. To our knowledge, this concept was first proposed in a conference article by Ranalli [3] upon which this work is built.

The rest of this article is organized as follows. The principle of the method will be described in Section III, while a concrete application on combiner-level data from a utility-scale PV plant will be described in Section IV, including validation results in Section IV-G. Finally, Section V concludes this article.

II. Background

Broadly speaking, describing the relationship between generation in different segmented areas within a plant falls into the topic of spatiotemporal variability of the irradiance resource [4], [5]. Many previous studies have characterized spatiotemporal variability and its effect on irradiance forecasting. Studies have commonly described variability in terms of the correlation between spatially disparate measurements [6], [7]. In particular, results have shown that spatial correlations in the signal result from advected cloud cover [8]. As a result, it is well known that correlation decreases with increasing separation distance [4] due

Manuscript received 23 October 2023; revised 31 January 2024; accepted 8 February 2024. Date of publication 28 February 2024; date of current version 19 April 2024. The work of Joseph Ranalli was supported by the Penn State School of Engineering Design and Innovation and Penn State Hazleton. *(Corresponding author: Joseph Ranalli.)*

Joseph Ranalli is with Penn State Hazleton, Hazleton, PA 18202 USA (e-mail: jar339@psu.edu).

William B. Hobbs is with Southern Company, Birmingham, AL 35203 USA (e-mail: whobbs@southernco.com).

Python source code and a subset of data are available on-line at https://github.com/jranalli/solartoolbox.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JPHOTOV.2024.3366666.

Digital Object Identifier 10.1109/JPHOTOV.2024.3366666

to the reduced likelihood of identical cloud patterns overhead with greater spatial separation. Further highlighting the relationship between spatial distribution of irradiance and cloud cover, many studies have observed anisotropy in the correlation in the along- and across-wind directions relative to cloud motion [9], [10] and studies have been conducted to attempt to model these effects [11], [12].

When considering generation from a PV facility, the spatiotemporal variability in the irradiance is aggregated, leading to an effective smoothing of the variable irradiance across the plant's spatial extent. Studies have attempted to describe this aggregation through its relationship to correlation between sensors within the plant [13]. Studies have also utilized the principle of advected cloud features over the plant as a method of forecasting aggregate plant generation [14], [15]. One previous model explicitly described the relationship between distributed areas responding to frozen cloud motion using transfer functions to represent the plant [16].

III. Theoretical Principle of the Method

The method proposed in this article utilizes delay between spatially distributed plant segments subject to observed cloud motion to compute a prediction of segment positions relative to those clouds. A simplified example will be used to describe the mathematics behind the method this study proposes. Consider two sensors, P_1 and P_2, believed to be separated by a distance d. Assuming advection of frozen clouds across the two points at a velocity $\overrightarrow{V}$, the response of the the sensors will exhibit a relative delay, τ, in their irradiance signals [16]. The delay-based separation distance, $\overrightarrow{d'}$, between these points can be computed using the simple equation

$$\overrightarrow{d'} = \overrightarrow{V} \cdot \tau. \tag{1}$$

Ideally, d and d' would be equal, but in practice, differences between the two distances d and d' indicate a mismatch between the expected positions of the sensors and the positions inferred from the cloud motion. Such a discrepancy could be caused either by uncertainty in the measurement of τ or due to and incorrect expectation about the actual separation distance between the points (i.e., their true locations are unknown).

A. In Two Dimensions

Extending the example to two dimensions, we can consider two discrete time periods (A and B) which exhibit cloud motion along two distinct vectors, $\overrightarrow{V_A}$ and $\overrightarrow{V_B}$. For a target sensor believed to be located at P_1, consider two other sensors as reference points, $P_{2\,A}$ and P_{2B}, each aligned relative to P_1 along their respective cloud motion vectors (CMVs), $\overrightarrow{V_A}$ and $\overrightarrow{V_B}$. The measured time series delays, τ_A and τ_B, between P_1 and each of the reference points P_2 result in predicted distances, d'_A and d'_B, according to (1). These distances result in implied positions, $P'_{1\,A}$ and P'_{1B}, along the respective cloud motion axes. Stated another way, $P'_{1\,A}$ and P'_{1B} represent the apparent position of P_1 based on the signal time delay relative to each of the reference points, P_2. A depiction of this geometric interpretation can be seen in Fig. 1.

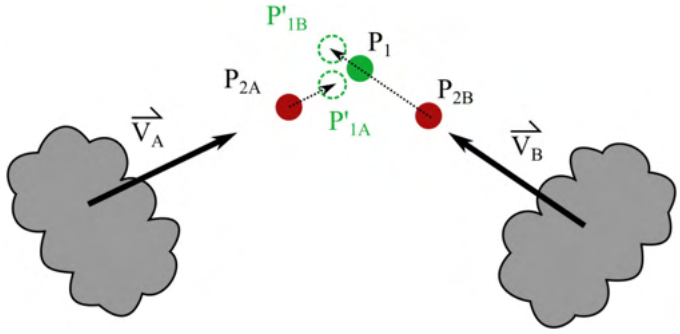

Fig. 1. Example of predicting position of P_1 based upon two periods with different cloud motion. Points $P_{2\,A}$ and P_{2B} are projected along vectors $\overrightarrow{V_A}$ and $\overrightarrow{V_B}$ by their respective delays, τ_A and τ_B, relative to the signal of P_1. The resulting predicted positions of P_1 along the A and B axes are termed $P'_{1\,A}$ and P'_{1B}. In the depicted case, the measured delay between $P_{2\,A}$ and P_1 is less than expected, while the opposite is true for P_{2B}.

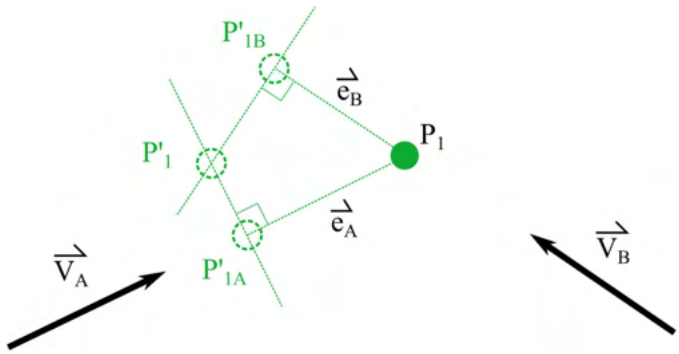

Fig. 2. Geometric relationship for obtaining the predicted position, P'_1, relative to the supposed position, P_1, based upon predictions $P'_{1\,A}$ and P'_{1B} along the CMV axes and the associated error vectors $\overrightarrow{e_A}$ and $\overrightarrow{e_B}$. Positions $P'_{1\,A}$ and P'_{1B} are consistent with those shown in Fig. 1.

As before, mismatch between the locations $P'_{1\,A}$ and P'_{1B}, and the supposed position of P_1 indicates a discrepancy between the expected position of P_1 and that inferred from the cloud motion. We will term these mismatches $\overrightarrow{e}$ as in the following equation:

$$\overrightarrow{e_A} = \overrightarrow{d'_A} - \overrightarrow{d_A}. \tag{2}$$

The predicted position of the sensor, P'_1, derived from the time delays, can be calculated as the intersection of the lines perpendicular to each cloud motion vector, passing through $P'_{1\,A}$ and P'_{1B}. This is graphically depicted in Fig. 2. Defining a cartesian coordinate system with P_1 as the origin, the offset of position P'_1 from P_1 can be calculated from the x and y components of vectors $\overrightarrow{e_A}$ and $\overrightarrow{e_B}$ as follows (a derivation may be found in the appendix):

$$P'_{1,x} = \frac{e_{By} \cdot |\overrightarrow{e_A}|^2 - e_{Ay} \cdot |\overrightarrow{e_B}|^2}{e_{Ax} \cdot e_{By} - e_{Ay} \cdot e_{Bx}} \tag{3}$$

$$P'_{1,y} = \frac{e_{Ax} \cdot |\overrightarrow{e_B}|^2 - e_{Bx} \cdot |\overrightarrow{e_A}|^2}{e_{Ax} \cdot e_{By} - e_{Ay} \cdot e_{Bx}}. \tag{4}$$

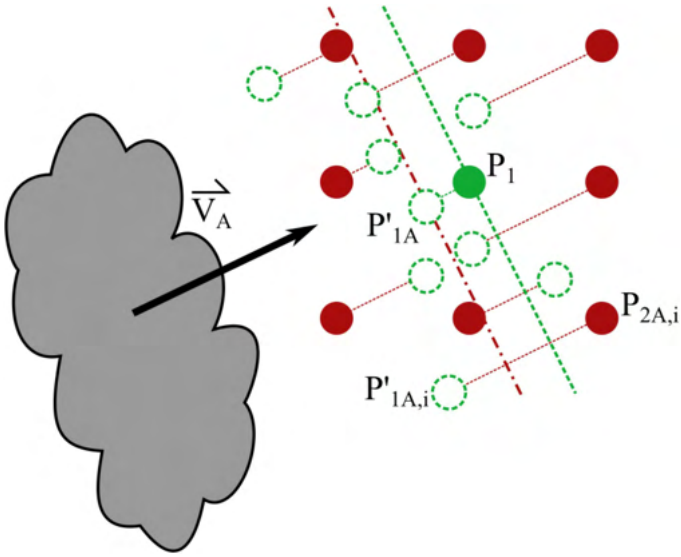

Fig. 3. Visual representation of a group of potential reference points, $P_{2\,A,i}$, each of which have a unique $P'_{1\,A,i}$ (green circles) based on delay between $P_{2\,A,i}$ and P_1. Averaging those predicted positions along the direction of the CMV, $\overrightarrow{V_A}$, yields an averaged position $P'_{1\,A}$ (at the distance of the green dashed line), which can be used to compute $\overrightarrow{e_A}$ for subsequent calculations.

Inspection of these equations highlights the importance of using cloud motion vectors that are nearly perpendicular to produce well conditioned results. As $\overrightarrow{V_A}$ and $\overrightarrow{V_B}$ approach parallel (or antiparallel) orientation, the denominator of (3) and (4) tends toward zero, magnifying any uncertainty in the measurement of τ, which serves as the basis for the vectors $\overrightarrow{e}$.

B. Averaging Multiple Reference Points

Expanding to a hypothetical plant, we use P_1 to represent a single target plant segment. Each other individual segment (i.e., string, combiner, inverter, etc.) represents a choice for the reference points, $P_{2\,A}$ and P_{2B}, removing the restriction that they lie exactly along the CMV relative to P_1. These numerous references provide an opportunity to compute many candidate positions $P'_{1\,A}$ and P'_{1B} and to average to reduce the effects of noise and uncertainty in computation of the τ values. A graphical example of this process in the $\overrightarrow{V_A}$ direction is shown in Fig. 3. Each reference point, $P_{2\,A,i}$ yields a predicted position $P'_{1\,A,i}$ projected parallel to the cloud motion vector, $\overrightarrow{V_A}$. The average predicted position, $P'_{1\,A}$, is then computed by averaging each of these distances to yield an average projection from P_1 along the cloud motion vector. Mathematically, this is equivalent to averaging the individual vectors, $\overrightarrow{e_{A,i}}$ according to

$$\overrightarrow{e_A} = \sum_{i=1}^{N} \overrightarrow{e_{A,i}}/N. \tag{5}$$

Any resultant discrepancy between the inferred P'_1 and the position P_1 (which is based on plant design drawings) represents the difference between the expected position of the plant segment and the apparent position of the segment as inferred from the measured signal time delays. As previously stated, any such discrepancy suggests a potential mislabeling and further investigation is warranted.

IV. Case Study and Discussion

In order to test the principles of the proposed method and to develop a concrete implementation thereof, a case study was performed based upon data from an operational PV plant, with the goal of of cross-checking the expected segment locations.

The plant studied was an approximately 20-MW PV plant located in the United States. The plant was analyzed at the level of its 221 combiners, with individual capacities ranging from approximately 75 to 120 kW each. Data from the plant were extracted for each combiner for the entire year of 2022. Data consisted of generation current timeseries, in amps, at the combiner level. The data source's sampling interval was irregular to save on storage space, with data recorded only when a sufficiently large value change occurred. The fastest sampling occurred at intervals of 10 s, so irregularly sampled data were interpolated to even sampling at a 10-s rate for analysis.

The following steps were used to apply the proposed method to data from the plant, and a graphical flowchart representing the overall process is shown in Fig. 4.

1) Digitize plant layout to obtain expected locations of combiners.
2) Select a target combiner, P_1, leaving remaining combiners as potential references, $P_{2,i}$.
3) Identify time periods with suitable CMVs.
4) Compute average predicted positions, $P'_{1\,A}$ and P'_{1B}, of the target along a pair of CMV axes.
5) Compute predicted target position, P'_1, for given CMV pair.
6) Repeat 4) and 5) for each possible CMV pair from 3).
7) Repeat 2)–5) for each desired target combiner.

A. Digitizing the Plant and Selecting Targets

Combiner footprints were identified from the plant design drawings. In this case, combiner footprints were all rectangular in shape and aligned along the cardinal direction axes. To represent these footprints as a single "average" spatial position, the geometric centroid of each combiner footprint was computed [16]. The layout of the plant use for the case study is shown in Fig. 5. Each individual combiner was analyzed sequentially to produce predictions for the entire plant.

B. Calculating Cloud Motion Vectors

A variety of methods exist for determining cloud motion vectors (CMVs). Some techniques allow the identification of cloud motion using All Sky Imagers [17], [18], satellite data [19], [20], or numerical weather prediction [13]. Other techniques obtain the CMV by observing spatiotemporal correlations between members of a distributed measurement network [21], [22]. In the present case study, we applied the method of Jamaly and Kleissl [22] on the effective measurement network made up of the individual combiners. Simply described, Jamaly and

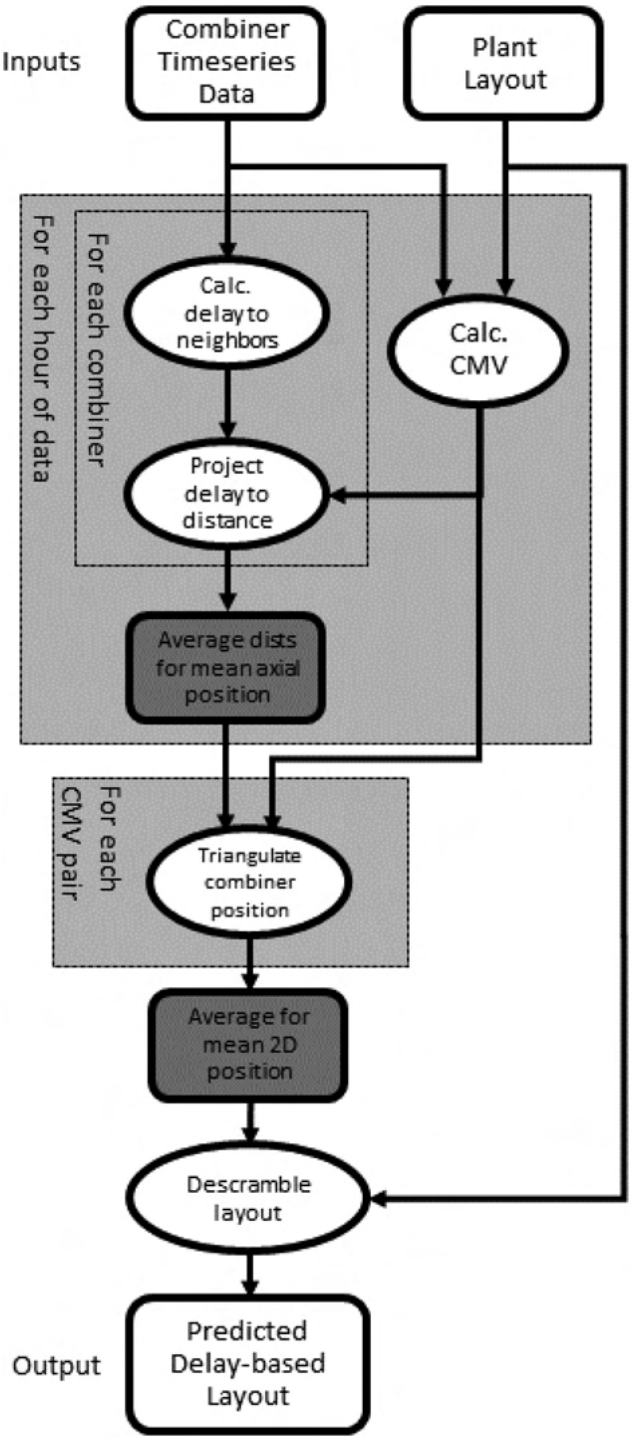

Fig. 4. Flowchart representing the basic steps in the process of applying the methodology.

Kleissl's method computes the delay between every pair of combiners. This method relies on computing the lag associated with the maximum value of cross correlation between the individual signals. The CMV orientation is determined as the direction that minimizes the variance in the corresponding velocity readings for all possible sensor pairings, and the CMV magnitude is computed as the median velocity at this direction.

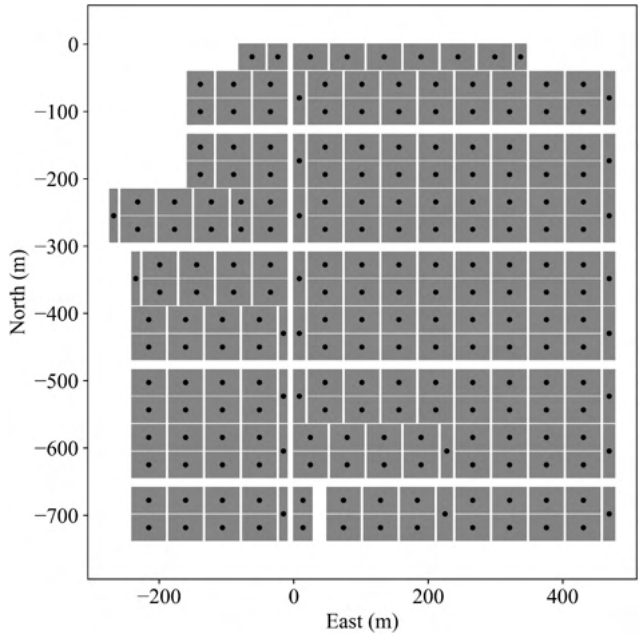

Fig. 5. Layout of the case study plant. Rectangles indicate combiner footprints. Dots indicate centroids used to represent the effective combiner position.

As discussed in Section III-A, it is necessary to analyze a minimum of two, nonparallel CMVs in order to predict the 2-D location of a target point. Practically speaking, these should each represent time periods during which the CMV is stationary and not varying in time. Additionally, as noted, the angle between CMVs must not be close to 0° or 180° to avoid singularities in the calculation.

For the purposes of this case study, CMVs were identified manually by visually comparing the output of the Jamaly and Kleissl CMV algorithm with animations of the time resolved intensity of the plant generation field. An example of the visualization using two snapshots in time is shown in Fig. 6. Suitable CMVs were identified for a total of 14 time periods, with durations ranging from 30 min to 1 h, between the months of June and November. CMVs were found with directions in all four quadrants of the cartesian plane. When defining "perpendicular" CMV pairs for the analysis, pairings were limited to those with relative angular separations between 45° and 135°. This resulted in 35 possible CMV pairings that were used for the analysis.

C. Computing the Position Along Each CMV

Predicting a combiner position according to (1) and (2) requires calculation of the delay, τ, between the target combiner and reference combiners. Several well known signal processing techniques could be employed for this purpose. Many studies have considered the lag at which peak cross-correlation occurs between the signals as a way to measure their relative lag, and as mentioned, that method is used by the CMV calculation [22].

For the purposes of this study, the delay was calculated from the phase of a transfer function between each pair of two points. The transfer functions were calculated by first normalizing individual signals to their mean, then applying a Hamming

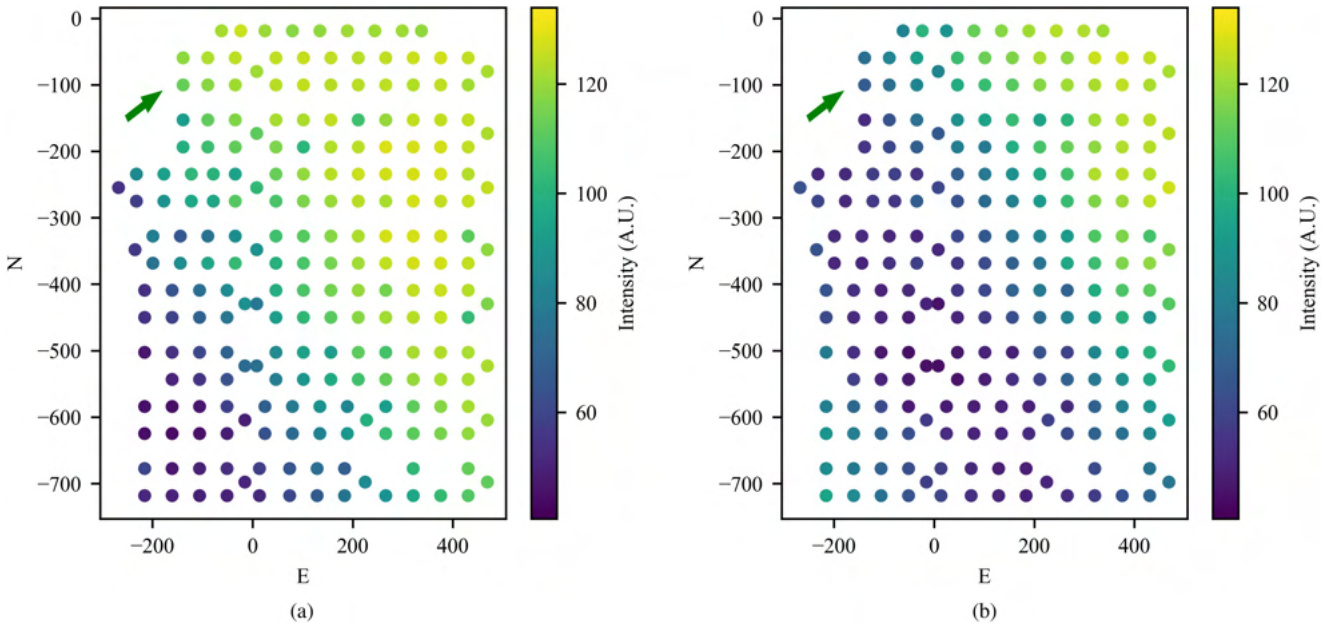

Fig. 6. Example visualizing cloud motion from the time resolved intensity of the plant. Figures show successive time instances of combiner generation from the data series. The arrow indicates the predicted cloud motion vector direction. A cloud front is visibly advancing across the plant in a northeasterly direction between the two frames. (a) Initial Time (b) Later Time.

window and computing 5 averages with a 50% overlap. The value of the delay was then obtained by fitting the unwrapped, low-frequency phase of the transfer function to the well known equation representing the linear phase shift associated with signals separated by a finite time delay, as given in (6). Only points with frequency below 0.02 Hz and for which the transfer function coherence was above 0.6 were used in the curve fitting

$$\phi = \text{angle}(\exp(-2\pi i f \tau)). \tag{6}$$

One benefit of obtaining the delay from the transfer function is that it enables noninteger values of the delay relative to the period of the data sampling rate to be computed. Additionally, and of specific importance to this study, delay can also be computed for cases whose value is less than the sampling time step. An example transfer function between two points, along with the curve fit for delay, is shown in Fig. 7. While the time series data used here had a sampling period of 10 s, in principle, the method could be used for data with a slower sampling rate (e.g. 30 s or 1 min data). For example, the magnitude of the delay indicated in Fig. 7 has a value of $\tau = 6.7$ s, which is less than the 10 s resolution of the time series data and suggests that delays can be observed that are smaller than the sampling period. However, the ability to use lower frequency data is likely to face practical limitations in the ability to accurately identify the relative delay between combiner pairs. As spatial separation and time are related by the actual speed of cloud motion, lower frequency data may favor the use of slower cloud speeds, but focusing on lower cloud speeds also increases the likelihood of the frozen clouds assumption being violated by clouds evolving during their transit of the plant. Future research may wish to

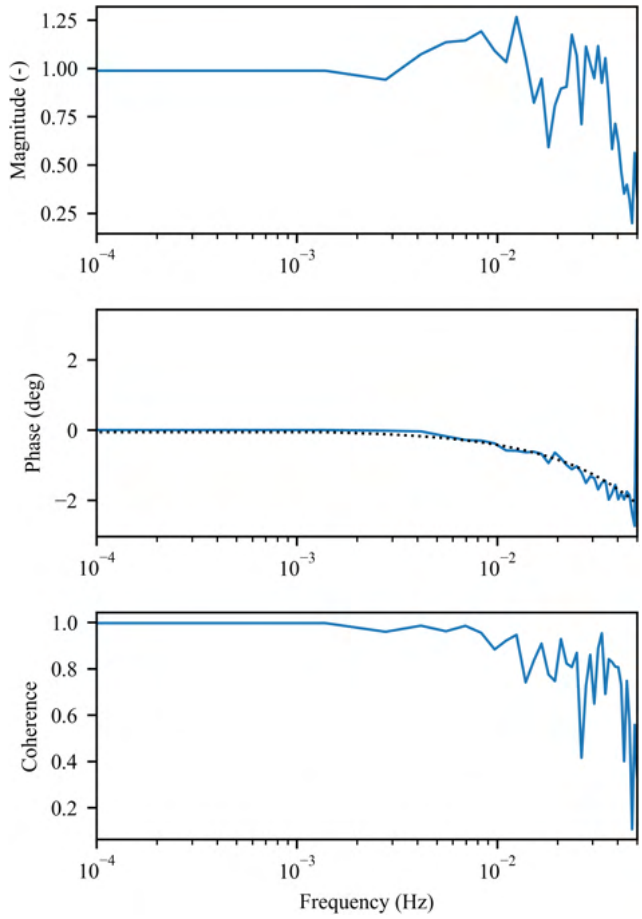

Fig. 7. Example transfer function between an arbitrary combiner pair. Phase plot shows the fit of the group delay, as indicated by a dotted line. Magnitude of delay is $\tau = 6.7$ s.

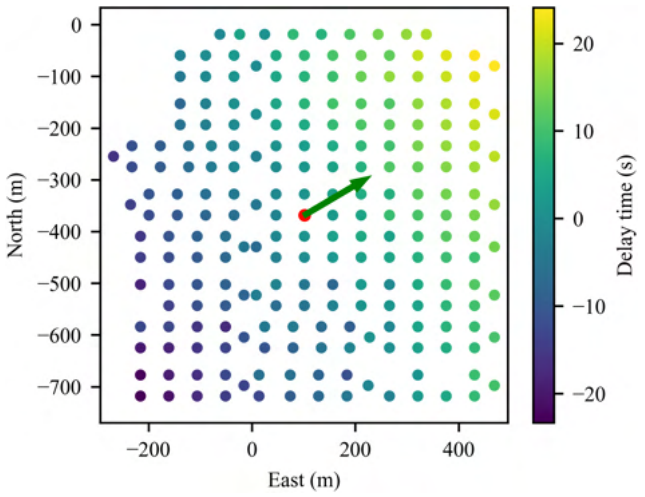

Fig. 8. Delay measured between each combiner and a central reference. Red dot indicates the reference combiner. The arrow indicates the CMV direction.

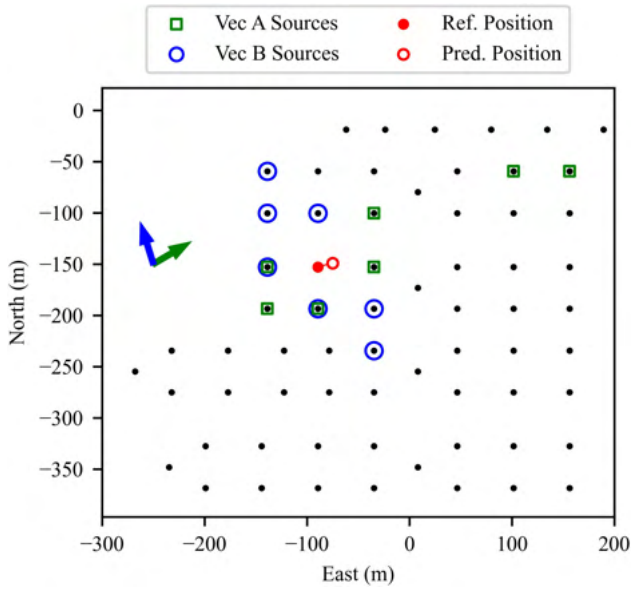

Fig. 9. Zoomed view on northwest region of plant showing predicted position for a single target combiner. Arrows show the two CMVs used. Colored markers show the location of nearby combiners that exhibited the strongest coherence for each CMV. These were averaged to obtain the position along the CMV as described in Section IV-C.

consider the degree to which the methodology is compatible with lower frequency time series data.

According to (1) and (2), delays were calculated between the input target combiner and the reference combiner as an output, considering all other combiners in the plant as potential references. A graphical visualization of the delays obtained for the plant during a single CMV window are shown in Fig. 8. As seen, delay values show a relatively smooth gradient in the direction of the cloud motion, corresponding to the leading or lagging of the individual combiner signals relative to the reference and their geometric separation. As observed in other studies [7], [9], [13], [23], the correlation between points decreases at greater distance from the input, particularly perpendicular to the cloud motion. This manifests as a reduction in the coherence of the transfer function at greater separation distances.

The delays were transformed into positions along the CMV axis using the equations described previously. As in Section II-I-B, delays from multiple reference points were averaged together to produce a single value for the predicted position along a given CMV axis. Several methods for selecting a subset of combiners to average were tested, including the use of all references, or downselecting to the closest neighbor references or those with the strongest correlation to the target point. Good results were obtained when computing the predicted positions along the CMV axis using the eight reference points with the highest average value of the coherence within a low-frequency window ($f < 0.02$ Hz). This ensured that only the points with the strongest relationship to the target point influenced the position prediction. This filtering was particularly helpful for managing the potential loss of correlation perpendicular to the cloud motion [9], [10]. The same process was repeated for each suitable CMV to obtain the predicted positions along each axis.

D. Obtaining a Final Position Prediction

Given a unique CMV pair, the predictions along each CMV axis were combined into a single spatial prediction for the position of the target combiner using (3) and (4). An example of the prediction for a single combiner is shown in Fig. 9. The prediction shown in the figure is closely located to the expected combiner position, indicating that the relative delays observed are consistent with the site plan's location for that combiner, and that the error vectors $\overrightarrow{e_A}$ and $\overrightarrow{e_B}$ are small. The figure highlights the individual combiners that were used in computing the averaged position (those exhibiting the strongest coherence for each CMV axis, as described in the preceding section). The indicated combiners with the highest coherence tend to fall along the cloud motion vector relative to the reference, which is consistent with expected anisotropic correlation behavior observed in the literature [7], [9]. The actual and predicted distance along the CMV vectors for each of these reference combiners is shown in Fig. 10, from which the values for $\overrightarrow{e_A}$ and $\overrightarrow{e_B}$ could be obtained.

E. Repeating for Each Target Combiner

While Figs. 9 and 10 showed the details of the prediction for a single combiner, validation of the entire plant's layout requires that predictions to be made for each combiner within the plant. The method was applied sequentially, considering each combiner as a target, with results shown in Fig. 11. Generally, the predicted location of each combiner falls approximately within its own footprint, indicating general agreement between the positions inferred from cloud motion and the site design documents. A few areas exhibited especially poor agreement, including the southern central and southwestern regions within the plant.

As described previously, poor agreement between the design plan and these inferred positions could indicate either uncertainty in the computation of the time delay, or potential mislabeling. To attempt to differentiate these effects, calculations were

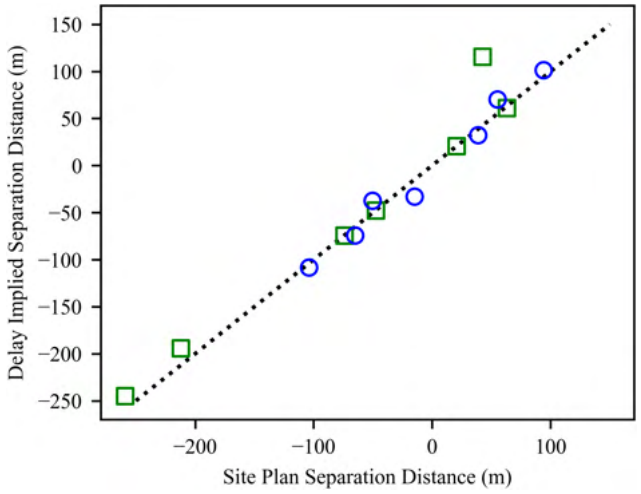

Fig. 10. Separation distances parallel to the CMV axis between the target and each reference. The two axes compare the expected distance as measured from the plant site plans and that computed from the delay. Each point indicates the distance for one of the combiners with strongest coherence for the given CMV axis, as shown in Fig. 9. In perfect agreement, points would all fall along the $x = y$ dashed line. Thus, the vertical separation between each point and the dashed line represents the error vectors $e_{A,i}$ or $e_{B,i}$.

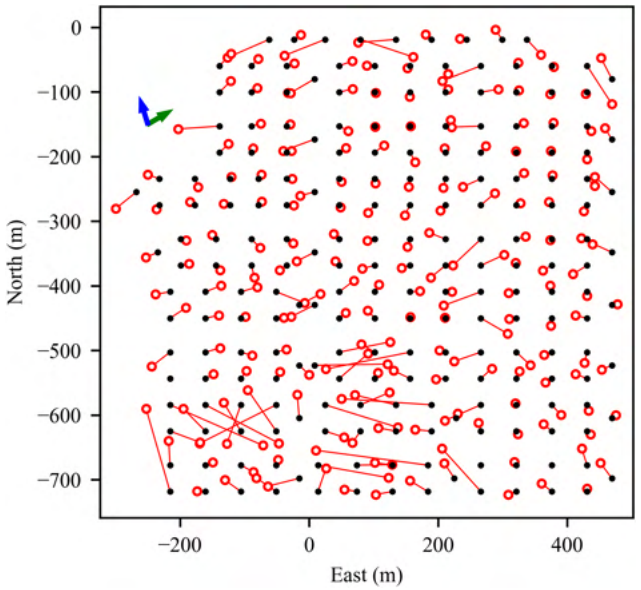

Fig. 11. Predictions for every combiner in the plant using a single CMV pair, as indicated by the arrows in the top left. Black dots show the position indicated by the site plan, while red lines and circles show the distance to the predicted position for the attached combiner.

repeated for each of the possible CMV pairings with relative angles between 45° and 135°, as described in Section IV-B. The final position prediction, P'_1, was found by averaging the x and y components of all individual P'_{1j} predictions, where the j subscript represents a single valid CMV pair. This provides an indication of the repeatability of the predictions, independent of the individual CMVs. In addition, the confidence of the predictions was estimated by computing the average separation between the individual P'_{1j} and P'_1 for each given target point. These results are shown in Fig. 12.

F. Identifying Potentially Mislabeled Combiners

Comparing the results in Figs. 11 and 12(a) shows that results for a single CMV pair are relatively similar to those averaged across multiple CMV pairs. The colorization of scatter across the CMV pairs presented in Fig. 12(b) and shows that most combiners experienced uncertainties less than 40 m (combiner row spacing is around 37 m). More details of the uncertainty are presented in Fig. 13, which shows a histogram and cumulative distribution function (CDF) of the combiner-by-combiner uncertainty. When compared to the typical combiner row spacing, Fig. 13 indicates that 85% of combiner position predictions are repeatable to within the size of a combiner footprint, which suggests that the localization of the combiner predictions is reasonable. In conjunction with the visual from Fig. 12(a), it is possible to conclude that most expected combiner positions from the plant design documents were validated by the approach.

The several sections of the plant for which combiner predictions were not consistent with the site layout bear some further scrutiny and discussion. Two centrally located combiners in the north-most row appear to have their positions swapped with relatively high confidence [labeled rectangle A in Fig. 12(a)]. Manual inspection of the transfer functions for that combiner pair showed that when clouds moved from west-to-east, the phase delay between the two combiners exhibited an opposite trend from neighbors, which is consistent with the implication that their positions have been swapped.

The apparent inconsistencies in the southern portion of the plant were also investigated for potential mislabeling. Though few of the combiners have predicted positions near their expected locations, the average scatter across predictions is relatively low, indicating that the predicted result is repeatable independent of the CMV direction. Notably, all combiner positions predicted with high confidence to be incorrectly located still fell within the footprint of the connected inverter, indicating that labeling errors tended to occur within the shared inverter connection. This is interesting, as it is consistent with observations of the plant operations team, which suggest that errors spanning multiple inverters are unlikely based on the wiring configuration of the plant.

To determine whether these predictions of incorrectly labeled combiners could be useful for identifying the combiners' true locations, the predictions in Fig. 12(a) were used to attempt a manual descrambling of the combiner labels in search of a pattern by which mislabeling may have occurred. A set of updated label predictions for combiners within the inverter area labeled C in Fig 12(a) are given in Table I. Updated label predictions were also computed for the southern central region of the plant. The predicted descrambled labels were then used in a reprocessing of the position prediction calculation, leading to Fig. 14. As is evident, the combiners that previously showed significant mismatch between expected and predicted position appear to have been corrected, i.e., mapped to a condition in which the delay-inferred and expected positions agree.

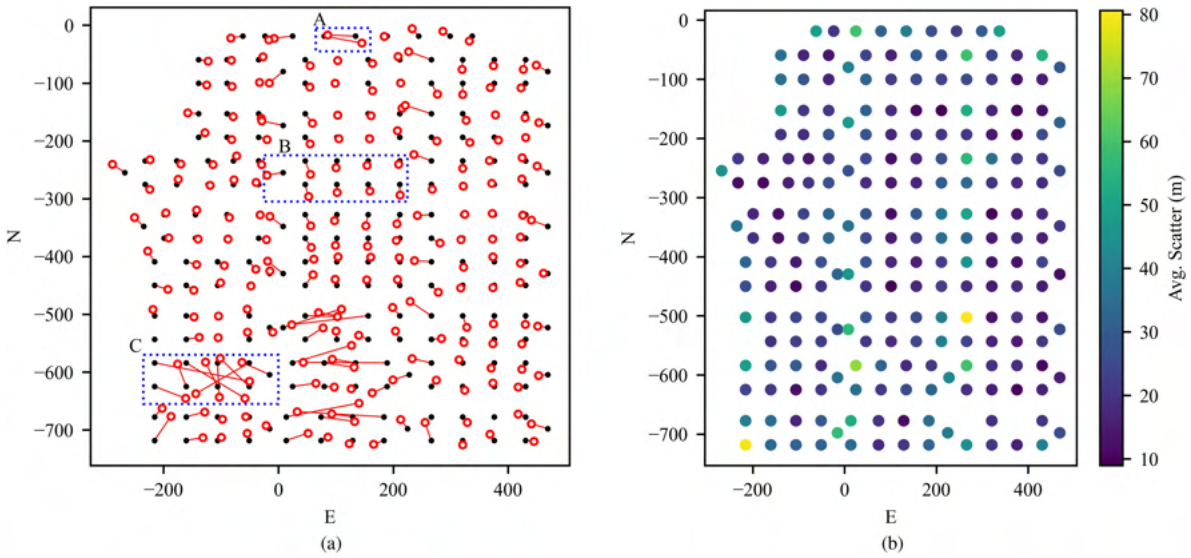

Fig. 12. Aggregated results across all possible CMV pairs. (a) Shows the average predicted position for each combiner, while (b) shows the average scatter of the predictions for each individual combiner across all CMVs. Dashed rectangles indicate areas that were used subsequently for manual verification (see Section IV-G). (a) Positions (b) Distances.

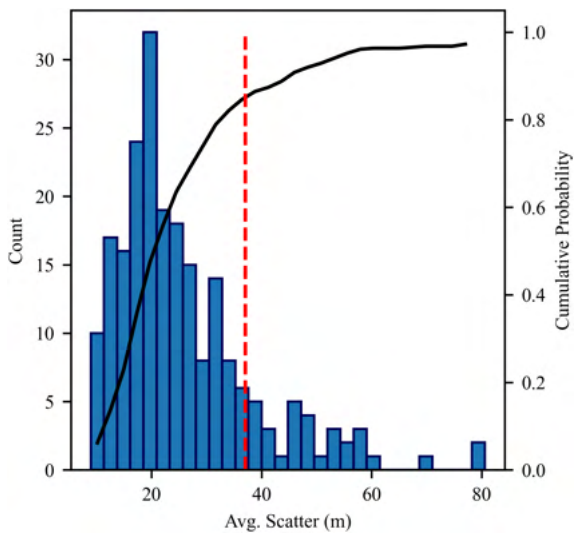

Fig. 13. Histogram (bars) and CDF (line) showing the mean scatter associated with each combiner in Fig. 14. Red dashed line indicates the row spacing of 37 m. CDF indicates that 85% of combiners experience less scatter than the typical row spacing.

TABLE I
VALIDATED COMBINER POSITIONS

Plant drawing	Descrambled location (predicted)	Observed location
1	9	9
2	8	8
3	7	7
4	6	6
5	5	5
6	1	1
7	2	2
8	3	3
9	4	4

It is interesting to note that when comparing Figs. 12(a) and 14(a), predicted positions changed for combiners whose labels had been unchanged in the descrambling process. This can be observed by comparing the position of combiner 5 between the two parts of Fig. 15. This observation highlights the somewhat circular nature of the approach developed in this study. That is, the expected combiner positions are used to calculate delays, delays are used to calculate predicted combiner positions, and predicted positions are compared to the expected positions to determine label accuracy. Any inconsistency in the original expected positions therefore impacts the quality of the delay calculations. In the worst case event that labels were randomly scrambled throughout the entire plant, it is unlikely that delay calculations would be able to yield sensible position predictions for the combiners. Stated another way, since delay is measured relative to a reference, in the absence of any confident reference position, no relative distance predictions are possible. In such an event, we hypothesize that other references (e.g., irradiance monitoring stations or aggregate level production data, such as inverter-level power) may be able to serve as the reference from which delays are calculated. In the present study, this is supported by the fact that all predicted location corrections used to produce Fig. 14 occurred by relabeling combiners that shared a common inverter.

G. *Validating Predictions Against Operational Plant*

As stated in the preceding section, application of the method allowed prediction of descrambled positions of plant combiners. The apparently mislabeled inverter in the southwest of the plant (rectangle C in Figs. 12 and 14) was targeted for validation of the

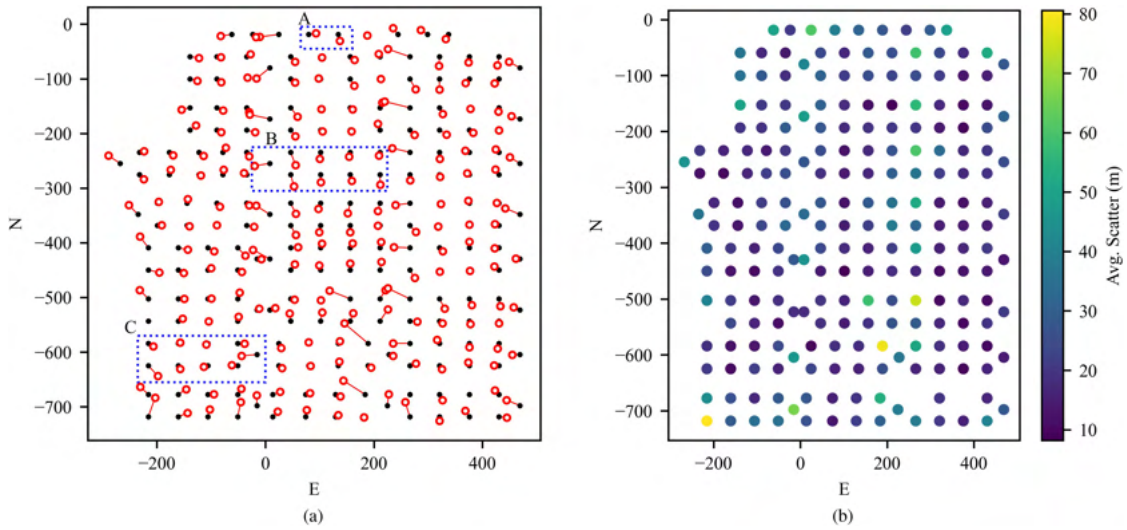

Fig. 14. Aggregated results across all possible CMVs after swapping combiners based on the results of Fig. 12. (a) shows the average predicted position for each combiner, while (b) shows the average scatter of the predictions across all CMVs for each combiner. Dashed rectangles indicate areas that were used in manual verification (see Section IV-G). (a) Positions (b) Distances.

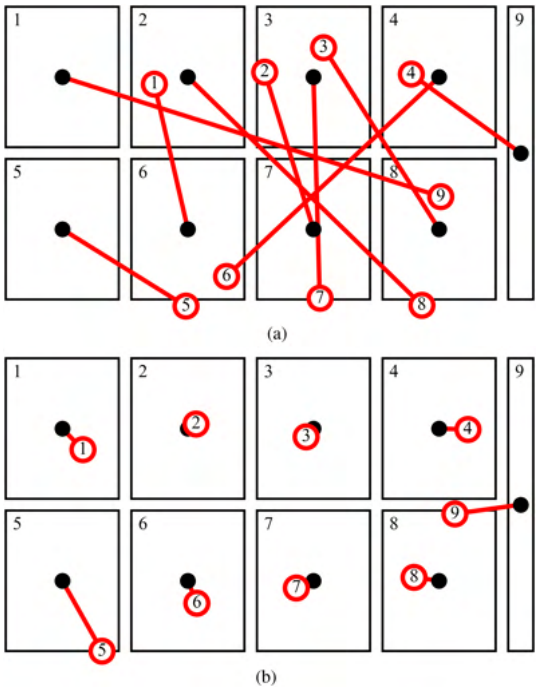

Fig. 15. Map of the southwest inverter in rectangle C, where rectangles indicate combiner footprints with numbers indicating the plant drawing labels. Red lines and circles show the distance to the predicted position, and numbers in the circles indicate the predicted location of that combiner. (a) Arrangement before descrambling: note that predicted locations are not necessarily based on the predicted position falling inside of a combiner's footprint – they are assigned to the nearest combiner such that all combiners get a predicted location within an inverter. (b) Arrangement after reprocessing data following descrambling, predicted positions move due to the dependence of predictions on initial positions. (a) Before descrambling (b) Re-processed after descrambling.

predictions against the actual plant. Maps of that inverter initially and after the predicted descrambling are shown in Fig. 15. In order to validate the predictions, combiners within that inverter footprint were individually disconnected in sequence for a period of at least 20 s prior to reconnecting, resulting in clearly visible zero readings for each combiner on the plant's data monitoring system. The "true" combiner labels were obtained by correlating the sequence of zero readings with the sequence in which the combiners were disconnected. The results of this validation, compared with the predictions made by the model, are shown in Table I. As the table indicates, perfect agreement occurred between predictions and observations. This agreement is also shown graphically in Fig. 15(b).

Additional manual inspections were conducted on a centrally located inverter (9 combiners) whose locations were predicted to agree with the site design documents, and on the two north-central combiners whose positions were predicted to be swapped. These plant locations are indicated by rectangles A and B in Figs. 12 and 14. As in the case of area C, the predictions made by the model agreed with observations for these additional combiners. Thus, the model exhibited a perfect validation for predictions on all 20 combiners that were manually cross-checked. The process of systematically traversing and individually disconnecting the combiners required several hours of technician time investment, not including test plan preparation time and time required for travel to the plant site. This represents the potential of this analytical approach to avoid a significant labor cost.

V. Conclusion

This article introduced a novel methodology for predicting the positions of individual segments within a photovoltaic plant by analysis of cloud motion over the plant from operational data. The principle of the methodology, introduced in Section III, could be applied in multiple ways depending on the data source, but one concrete implementation was demonstrated on combiner-level data for an actual plant in Section IV. The method is based on the concept of calculating the relative delay

between the plant segments as clouds advect over the plant. A period of fixed cloud motion allows calculation of the linear separation of two segments parallel to the cloud motion. By utilizing two periods with approximately perpendicular cloud motion vectors, a 2-D geometric projection of these positions can be performed, resulting in a prediction of the position for the target segment.

The concrete demonstration of the method utilized combiner-level data from an operational utility-scale plant. In order to implement the method on this data, we employed methods from literature for identifying the cloud motion vector from the distributed combiner measurements and computed delay between the combiners via the phase of their transfer function. It was possible to predict the position of each combiner within the plant, demonstrating the basic functionality of the method. We found that the confidence in the results improved by averaging predicted positions over multiple cloud motion vector pairs, which also allowed for an estimate of the repeatability of the predicted position. While most predicted combiner positions matched relatively closely with their expected design plan positions, several apparent instances of combiner mislabeling were identified. We demonstrated that the methodology allowed a prediction to be made for how the combiner labels could be descrambled in these instances, resulting in a prediction of a "corrected" map of labels for the plant. A partial validation based on inspection of the plant showed that the method's predictions yielded perfect agreement with the true plant layout.

The present study effectively demonstrates the application of this proposed method. The approach may be of interest to operators of photovoltaic plants who have concerns about accuracy of the identification of plant segments (e.g., at string, combiner or inverter levels) and wish to validate these labels without laborious and time consuming hands-on approaches at the physical site. An open source software implementation of the method as used in the demonstration is made available, as described in the code availability section following the appendix.

There are some areas of further research that we hypothesize could improve the application of the method. First, as discussed in Section IV-C, some plants may only have access to data with sampling rates slower than the 10-s data used here. Analysis of the practical accuracy of the method when applied on lower frequency data may make it useful for a wider variety of plants. Additionally, the method demonstrated here relied on some manual interpretation to complete the analysis, e.g., manual calculations were used to identify suitable cloud motion vector time periods and to identify the potentially mislabeled sections of the plant. Developing approaches to automate these workflows may be a fruitful area for future research to improve scalability and allow the method to be quickly applied for validating an entire operational fleet.

APPENDIX
DERIVATION OF (3) AND (4)

Referring to Fig. 2, define P_1 the origin and consider the position vector pointing to the final predicted position, $\overrightarrow{P'_1}$. Recognizing that both vectors $\overrightarrow{e_A}$ and $\overrightarrow{e_B}$ are projections of $\overrightarrow{P'_1}$ onto the cloud motion vector axes, we can observe that the dot product between $\overrightarrow{P'_1}$ and $\overrightarrow{e_A}$ is the square of the magnitude of $|\overrightarrow{e_A}|$

$$\overrightarrow{P'_1} \bullet \overrightarrow{e_A} = |\overrightarrow{e_A}|^2. \tag{7}$$

Writing likewise for $\overrightarrow{e_B}$ yields two equations with the only two unknowns as the components P'_{1x} and P'_{1y}. Writing the two dot product equations in a matrix form gives

$$\begin{bmatrix} e_{Ax} & e_{Ay} \\ e_{Bx} & e_{By} \end{bmatrix} \begin{pmatrix} P'_{1x} \\ P'_{1y} \end{pmatrix} = \begin{pmatrix} |\overrightarrow{e_A}|^2 \\ |\overrightarrow{e_B}|^2 \end{pmatrix}. \tag{8}$$

Applying Cramer's rule to solve for P'_{1x} and P'_{1y} yields the forms shown in (3) and (4).

ACKNOWLEDGMENT

The authors thank Nathan Williams and David Starling for insight toward simplifying the mathematical approach, and Jan Kleissl and Cristian Cortes Aguirre for providing CMVs based on numerical weather prediction for testing.

REFERENCES

[1] IPCC, "Climate change 2022: Impacts, adaptation and vulnerability," in *Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, H.-O Pörtner et al., eds. New York, NY, USA: Cambridge University Press, 2022, doi: 10.1017/9781009325844.

[2] P. Denholm et al., "Examining supply-side options to achieve 100% clean electricity by 2035," Golden, CO: National Renewable Energy Laboratory, Tech. Rep. NREL/TP-6A40-81644, 2022. [Online]. Available: https://www.nrel.gov/docs/fy22osti/81644.pdf

[3] J. Ranalli, "Validation of inverter labeling with plant transfer functions," in *Proc. IEEE 50th Photovolt. Specialists Conf.*, 2023, pp. 1–5.

[4] R. Perez et al., "Spatial and temporal variability of solar energy," *Found. Trends Renewable Energy*, vol. 1, no. 1, pp. 1–44, Jul. 2016. [Online]. Available: https://www.nowpublishers.com/article/Details/REN-006

[5] G. M. Lohmann, "Irradiance variability quantification and small-scale averaging in space and time: A short review," *Atmosphere*, vol. 9, no. 7, Jul. 2018, Art. no. 264. [Online]. Available: https://doaj.org

[6] J. Widén, "A model of spatially integrated solar irradiance variability based on logarithmic station-pair correlations," *Sol. Energy*, vol. 122, pp. 1409–1424, Dec. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X15005940

[7] B. Elsinga, "Chasing the clouds: Irradiance variability and forecasting for photovoltaics," Dissertation, Utrecht University, Nov. 2017. [Online]. Available: http://dspace.library.uu.nl/handle/1874/356774

[8] T. E. Hoff and R. Perez, "Quantifying PV power output variability," *Sol. Energy*, vol. 84, no. 10, pp. 1782–1793, Oct. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X10002380

[9] V. P. A. Lonij, A. E. Brooks, A. D. Cronin, M. Leuthold, and K. Koch, "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors," *Sol. Energy*, vol. 97, pp. 58–66, Nov. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X13003125

[10] L. M. Hinkelman, "Differences between along-wind and cross-wind solar irradiance variability on small spatial scales," *Sol. Energy*, vol. 88, pp. 192–203, Feb. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X12004021

[11] E. Arias-Castro, J. Kleissl, and M. Lave, "A poisson model for anisotropic solar ramp rate correlations," *Sol. Energy*, vol. 101, pp. 192–202, Mar. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X13005549

[12] G. M. Lohmann, A. Hammer, A. H. Monahan, T. Schmidt, and D. Heinemann, "Simulating clear-sky index increment correlations under mixed sky conditions using a fractal cloud model," *Sol. Energy*, vol. 150, pp. 255–264, Jul. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X17303419

[13] M. Lave and J. Kleissl, "Cloud speed impact on solar variability scaling–Application to the wavelet variability model," *Sol. Energy*, vol. 91, pp. 11–21, May 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X13000406

[14] G. C. Wang et al., "Worst expected ramp rates from cloud speed measurements," in *Proc. IEEE 46th Photovoltaic Specialists Conf.*, 2019, pp. 2281–2286, iSSN: 0160-8371.

[15] M. Lipperheide, J. L. Bosch, and J. Kleissl, "Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant," *Sol. Energy*, vol. 112, pp. 232–238, Feb. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X1400557X

[16] J. Ranalli and E. E. M. Peerlings, "Cloud advection model of solar irradiance smoothing by spatial aggregation," *J. Renewable Sustain. Energy*, vol. 13, no. 3, May 2021, Art. no. 033704. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/5.0050428

[17] Z. Peng et al., "3D cloud detection and tracking system for solar forecast using multiple sky imagers," *Sol. Energy*, vol. 118, pp. 496–519, Aug. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X15002972

[18] B. Nouri et al., "Nowcasting of DNI maps for the solar field based on voxel carving and individual 3D cloud objects from all sky images," *AIP Conf. Proc.*, vol. 2033, no. 1, Nov. 2018, Art. no. 190011. [Online]. Available: https://aip.scitation.org/doi/10.1063/1.5067196

[19] A. Hammer, D. Heinemann, E. Lorenz, and B. Lückehe, "Short-term forecasting of solar radiation: A statistical approach using satellite data," *Sol. Energy*, vol. 67, no. 1, pp. 139–150, Jul. 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X00000384

[20] A. Roy, A. Hammer, D. Heinemann, O. Lünsdorf, and J. Lezaca, "Impact of tropical convective conditions on solar irradiance forecasting based on cloud motion vectors," *Environ. Res. Lett.*, vol. 17, no. 10, Oct. 2022, Art. no. 104048. [Online]. Available: https://dx.doi.org/10.1088/1748-9326/ac94e6

[21] A. Gagné, N. Ninad, J. Adeyemo, D. Turcotte, and S. Wong, "Directional solar variability analysis," in *Proc. IEEE Elect. Power Energy Conf.*, 2018, pp. 1–6, iSSN: 2381-2842.

[22] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X17309556

[23] J. Ranalli, "Correlations in spatial variability when accounting for cloud advection," in *Proc. IEEE 49th Photovolt. Specialists Conf.*, 2022, pp. 0032–0037.